Evaluating Longitudinal Impacts of Transit Investments on Transit Demand

by

Albert Lo

A thesis submitted in conformity with the requirements for the degree of Master of Applied Science

Department of Civil and Mineral Engineering
University of Toronto

© Copyright by Albert Lo, 2019
Evaluating Longitudinal Impacts of Transit Investment on Transit Demand

Albert Lo

Master of Applied Science

Department of Civil and Mineral Engineering
University of Toronto

2019

Abstract

Land use and transportation (LUT) policies have lasting impact on the sustainability of cities. Therefore, these should be examined longitudinally to assess the long-term effects on travel behaviour. Specifically, trends of transit accessibility and transit-oriented developments are considered of the Greater Toronto Hamilton Area from 2001 to 2016. Most of the data used comes from the Transportation Tomorrow Survey and augmented by other sources. At a disaggregate level, location choice models are developed to estimate person accessibility. Generally, people living in highly-mixed areas are more sensitive to transit-walk-access wait time. Regarding the population, several regression models are developed to determine the effect of the considered transit variables on the number of household transit trips. It is found that the variables’ effects increase over time in a positive manner. These results show that these policies have consequences in the long-term and should be considered for future policy.
Acknowledgments

This thesis is the culmination of support provided by multiple people. I would first like to thank the Transportation Information Steering Committee (TISC). Without the funds provided, this research would not have been possible. Thank you for continuing to support graduate student research in pursuit of better transportation in the Greater Golden Horseshoe.

Thank you to my supervisor, Professor Habib, for meeting me all those years ago when I was a second-year undergraduate student in industrial engineering. You peaked my interest in the topic of transportation demand and the complex nature of transportation systems. I am thankful for your willingness, patience, and guidance to take me on for my undergraduate thesis and masters.

Thank you to my second reader, Professor Shalaby for providing insightful comments to improve this work. Your advice regarding transit-oriented developments and their effect on transit demand proved useful as well.

As always, thank you to my parents and siblings for being there for me every day.

Finally, thank you to my peers and friends whom have motivated me throughout my journey. There was joy and learning to be had at the ITSlab, conferences, lunches, and dinners.
# Table of Contents

Acknowledgments.......................................................................................................................... iii

Table of Contents........................................................................................................................... iv

List of Tables ................................................................................................................................ vii

List of Figures ................................................................................................................................ viii

List of Appendices ......................................................................................................................... ix

Chapter 1 ..........................................................................................................................................1

  1 Introduction.................................................................................................................................1

    1.1 Chapter Overview................................................................................................................1

    1.2 Motivation............................................................................................................................1

    1.3 Research Questions..............................................................................................................2

    1.4 Dissertation Outline.............................................................................................................2

Chapter 2 ..........................................................................................................................................4

  2 Literature Review........................................................................................................................4

    2.1 Chapter Overview................................................................................................................4

    2.2 Introduction..........................................................................................................................4

    2.3 Transit Investment Definition..............................................................................................4

      2.3.1 Transit Accessibility................................................................................................4

      2.3.2 Transit Oriented Development.................................................................................5

    2.4 Assessing Transit Investment Trends..................................................................................5

    2.5 Long-Term Transit Investment Studies..............................................................................6

    2.6 Chapter Summary................................................................................................................8

Chapter 3 ........................................................................................................................................10

  3 The GTHA Data Context ............................................................................................................10

    3.1 Chapter Overview................................................................................................................10

    3.2 Introduction..........................................................................................................................10
3.3 The Region and Data Sources.........................................................................................10
3.4 Socioeconomics ..............................................................................................................12
  3.4.1 Population by Age Cohort ..................................................................................12
  3.4.2 Population by Region ..........................................................................................13
  3.4.3 Household Size ....................................................................................................14
  3.4.4 Number of Workers ..............................................................................................15
  3.4.5 Self-Containment .................................................................................................16
3.5 Travel Behaviour .............................................................................................................17
  3.5.1 Average Daily Trips ............................................................................................17
  3.5.2 Average Daily Trips by Mode .............................................................................18
  3.5.3 Mode Share by Year ............................................................................................18
  3.5.4 Number of Vehicles per Household .......................................................................19
  3.5.5 Spatial Flows ........................................................................................................20
3.6 Chapter Summary .........................................................................................................20
Chapter 4 ..................................................................................................................................22
4 A Longitudinal Analysis of Accessibility by Transit to Jobs and Schools Using the Utility-
based Accessibility Measure .................................................................................................22
  4.1 Chapter Overview ......................................................................................................22
  4.2 Introduction ................................................................................................................22
  4.3 Data for Empirical Investigation .................................................................................22
  4.4 Methodology of Empirical Investigation ....................................................................24
    4.4.1 Model Formulation ...........................................................................................24
    4.4.2 Dataset for Empirical Investigation ....................................................................27
    4.4.3 Empirical Models ...............................................................................................28
  4.5 Results and Discussion on Empirical Models .............................................................29
  4.6 Chapter Summary .......................................................................................................33
<table>
<thead>
<tr>
<th>Chapter 5</th>
<th>A Macro Model Analysis of the Relationship between Transit Investment and Transit Ridership in the GTHA</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Chapter Overview</td>
<td></td>
</tr>
<tr>
<td>5.2 Introduction</td>
<td></td>
</tr>
<tr>
<td>5.3 Data</td>
<td></td>
</tr>
<tr>
<td>5.4 Methodology</td>
<td></td>
</tr>
<tr>
<td>5.4.1 Pseudo Panel Analysis</td>
<td></td>
</tr>
<tr>
<td>5.4.2 Subject-Specific Models</td>
<td></td>
</tr>
<tr>
<td>5.5 Empirical Models and Discussion</td>
<td></td>
</tr>
<tr>
<td>5.5.1 Pseudo Panel Empirical Models and Discussion</td>
<td></td>
</tr>
<tr>
<td>5.5.2 Subject-Specific Empirical Models and Discussion</td>
<td></td>
</tr>
<tr>
<td>5.6 Chapter Summary</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 6</th>
<th>Conclusion and Future Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Chapter Overview</td>
<td></td>
</tr>
<tr>
<td>6.2 Research Summary</td>
<td></td>
</tr>
<tr>
<td>6.3 Future Research Recommendations</td>
<td></td>
</tr>
</tbody>
</table>

References

Appendix
List of Tables

Table 1. Difference in Regional Spatial Flow for Morning Peak Period, 1996 and 2016 ............. 20
Table 2. Location Choice Model Results.................................................................................... 28
Table 3. Selected Variables’ Sample Average Direct Elasticity..................................................... 30
Table 4. Pseudo Panel Model Results.......................................................................................... 40
Table 5. Elasticity of Independent Variables of Random Effect Model...................................... 42
Table 6. Subject Specific OLS Model Results............................................................................. 43
Table 7. Subject Specific Poisson and Negative Binomial Regression....................................... 45
List of Figures

Figure 1. GTHA Population by Age Cohort and Year (TTS Data) .............................................. 12
Figure 2. GTHA Population by Age Cohort and Year (Census Data) .......................................... 13
Figure 3. Population Distribution by Region and Year ................................................................. 13
Figure 4. Household Size Distribution by Year (TTS Data) ........................................................ 14
Figure 5. Household Size Distribution by Year (Census Data) .................................................... 14
Figure 6. Average Number of Workers per Household by Region and Year .............................. 15
Figure 7. Employment by Region and Year ................................................................................. 15
Figure 8. Proportion of Workers Living and Working in Same Regions as Residence ............... 16
Figure 9. Residential Region of PD1 Workers by Year ................................................................. 17
Figure 10. Average Number of Daily Trips made by GTHA Residents by Year ....................... 17
Figure 11. Average Number of Daily Trips made by GTHA Residents by Mode and Year ...... 18
Figure 12. Mode Share by Year .................................................................................................... 19
Figure 13. Average Number of Vehicles by Region and Year ...................................................... 19
Figure 14. Average Mixed Land Use Entropy of TAZ by Municipality and Year ..................... 23
Figure 15. Average Stop Density of TAZ by Municipality and Year ........................................ 24
Figure 16. CDF of Logsum Accessibility Utility Measure by Year, Students ......................... 32
Figure 17. CDF of Logsum Accessibility Utility Measure by Year, Workers ............................ 33
Figure 18. Accessibility by Transit Map for Students by Year, Centred in Toronto ................. 34
Figure 19. Accessibility by Transit Map for Workers by Year, Centred in Toronto .................. 34
List of Appendices

Appendix A..........................................................................................................................55
Chapter 1

1 Introduction

1.1 Chapter Overview

The main purpose of this chapter is to present the motivation behind this research thesis in Section 1.2. Specific research questions are asked in Section 1.3. Lastly, the following chapters are outlined in Section 1.4.

1.2 Motivation

Public transit and the transport policies that guide its services are believed to be important for sustainability and livability in urban areas. Transit aids in the reduction of carbon emissions and transportation demand of motorways. It provides access to employment and other discretionary activities, especially for those who have limited access to other modes of transport, e.g. private car. To be able to realize these benefits, a major requirement is for public transit to be accessible in the form of routes, stops, lines, and stations.

In addition, it is widely acknowledged that there is close interplay between transportation and land use systems. Introducing major public transit infrastructure presents opportunities for transit-oriented development (TOD) land use policy. TOD is a response to urban sprawl by densifying land use around major transit hubs in a mixed land use manner (Akbari, Mahmoud, Shalaby, & Habib, 2018). TOD also supports an active lifestyle by facilitating multimodal travel i.e. less auto-centric and more transit, pedestrian, and cyclist oriented. Therefore, land use and transport (LUT) policies play important roles in the sustainability of the environment, economy, and community of cities.

A significant characteristic of LUT policy is its long-term effect on the decision people choose to travel. These decisions include vehicle ownership, home, and work location choice. For example, Kasraian et. al. found that a Dutch decentralization land use policy introduced in the 1960s still influenced urbanization patterns decades later (Kasraian, Maat, & van Wee, 2017). Decentralization in the Greater Ranstad Area led to trips typically made between commuter neighbourhoods and the central business district. This effect persisted despite newer densification policies in the 1970s. A longitudinal study done by Tsai et. al. shows that transit
accessibility amongst other variables have a long-run impact on transit demand elasticity (Tsai, Mulley, & Clifton, 2014).

In the GTHA, there have been several major public transit infrastructure developments completed or ongoing, especially in Toronto. These transit investments include the Line 4 Toronto Transit Commission (TTC) subway line, the new BRT routes in York region, the Line 1 TTC subway extension, among others. The federal, provincial, and local governments allocated substantial amounts of money to fund these projects. For example, roughly $1 billion (2002, Canadian dollars) was spent on capital costs for the Sheppard line (Saxe, 2015). Public transit investments are a large financial and political undertaking, not to mention the cost to maintain and operate.

Therefore, it is imperative for policies on transit investments to be evaluated longitudinally. This would contribute to policy makers’ understanding of the long-term implications of different policies and make more informed decisions.

1.3 Research Questions

The objective of this dissertation is to longitudinally examine two specific transit investments: transit accessibility and TOD. The following research questions are proposed.

1. How has transit accessibility and the degree of TOD changed over time?
2. To what extent does transit accessibility and the degree of TOD affect the number of transit trips made by households?

This dissertation attempts to answer what longitudinal effects from LUT policy has had on transit accessibility, TOD, and household transit demand.

1.4 Dissertation Outline

The remainder of the dissertation, separated into chapters, are as followed.

Chapter 2 presents a literature review of longitudinal studies of transit investment. Definitions of transit accessibility and TOD are first examined. Then numerous papers on short-term and long-term effects of transit investments are reviewed and summarized.
Chapter 3 details the conceptual framework of the methodology to answer the research questions.

Chapter 4 sets the data context by displaying socioeconomic and transportation behaviour trends of the GTHA population from 2001 to 2016 as well as touching on the history of transit investment in the region. This chapter acts as an extension to a paper authored by Miller and Shalaby regarding changes in the GTHA population from 1964 to 1996 (Miller & Shalaby, 2003).

Chapter 5 presents empirical work of describing transit accessibility and TOD trends of the GTHA from 2001 to 2016. In addition to a basic count measure, discrete choice models are developed. These models are used to estimate longitudinal utility-based accessibility values which measure person access to locations by transit.

Chapter 6 delves into empirical models that examine the long-term effects of transit accessibility and TOD on the GTHA population. Both pseudopanel and subject-specific regression models are created and results are compared.

Chapter 7 summarizes the key findings from the empirical work and provides recommendations on LUT policies regarding transit investments. This is followed by future work.
Chapter 2

2 Literature Review

2.1 Chapter Overview

This chapter reviews literature relevant to longitudinal study of transit investment. Section 2.2 introduces the topics to be covered. Section 2.3 describes the various definitions and measures of transit accessibility and TOD used in transportation research. Section 2.4 delves into longitudinal trends of transit accessibility and TOD and Section 2.5 is about the longitudinal effect of these transit variables.

2.2 Introduction

The purpose of this chapter is to summarize prior research and findings regarding a multitude of relevant topics. First, transit accessibility and TOD measures are investigated. Then studies which examine transit accessibility at a longitudinal scale are considered. The chapter transitions into investigating the effects of these two transit variables on the transit demand over time. Finally, the topics are all summarized, and the conceptual research framework of this dissertation is set.

2.3 Transit Investment Definition

2.3.1 Transit Accessibility

Before reviewing papers measuring accessibility and its effects over time, it is important to go over the different definitions of accessibility and TOD measures. The accessibility considered in this review and paper is the degree of ease to interact with opportunities (Hansen, 1959). Specifically, transit accessibility is the ease in which an individual is able to reach their destination by transit. This definition does not consider headways and other measure of reliability of transit, though it is an important part of accessibility. This is because it is believed that there is great variability in reliability.

There are a countless number of publications and reviews of transportation accessibility measures including, but not limited to (Handy & Niemeier, 1997), (Geurs & van Wee, 2004), (Miller E. J., 2018) and (Hasnine, Graovac, Camargo, & Habib, 2019). Each paper classifies
accessibility into different categories. The categories set by Hasnine et. al., based on prior papers, include count, gravity, and utility measures. Count measures are typically the number of opportunities within a certain threshold. Gravity measures are usually a function of travel time where a higher value means the opportunity is less attractive. Being simplistic by themselves, these two measures have some disadvantages. Opportunities within a threshold or at the same distance are considered to have the same value. These values do not consider the destination or personal sociodemographic variable either. This may lead to accessibility values that are overestimated. Utility measure provide a more holistic approach considering personal and destination attributes.

2.3.2 Transit Oriented Development

Like accessibility, TOD also has several papers that define it. The TCRP TOD report defines TOD as mixed-use development, centred around transit infrastructure, high density, and active-mode friendly (Cevero, 2004). Virtually, this definition is followed by all papers (Nasri & Zhang, 2014; Singha, Farda, Zuidgeestb, Brussela, & van Maarseveen, 2014). In terms of TOD measures, literature is more varied. Most papers consider a catchment radius of 400 to 800 metres (a quarter to half a mile) (Galeo, Ribeiro, & Martinez, 2014). Some papers use a collection of TOD related variables like population density, employment density, walk score, existence of stations together in a regression or discrete choice model to control for TOD (Akbari, Mahmoud, Shalaby, & Habib, 2018). Others use a single variable like an index (Olaru, Smith, & Taplin, 2011) or a formula (Faghri, 2013; Papa & Bertolini, 2015) based on TOD related attributes.

2.4 Assessing Transit Investment Trends

There have been few papers that explicitly investigating accessibility over time. El-Geneidy and Levinson investigated changes in accessibility in Minneapolis for auto and transit from 1990 to 2000 (El-Geneidy & Levinson, 2007). The accessibility measure is a count of jobs and residents within a fifteen-minute travel time threshold. It concluded there has been an increase and decrease of job accessibility for auto and transit respectively. This methodology of using count measures to calculate job accessibility is similar to other papers; Levinson examined changes in accessibility for the United States (Levinson, 2013), Lionjanga and Venter considers
Johannesburg (Lionjanga & Venter, 2017), and Farber and Fu considers Utah County (Farber & Fu, 2017). Gravity-based measures are utilized by Foth et. al (Foth, Manaugh, & El-Geneidy, 2014). Kim and Song in addition to using a gravity-based accessibility, also used a probabilistic reliability measure (Kim & Song, 2018). The measures were able to capture changes in accessibility and reliability based on empirical results on the Seoul metro from the 1970s to 2010s.

2.5 Long-Term Transit Investment Studies

Several papers are published on transit investment and its effects on travel behaviour and transit demand. The purpose of this section is to investigate the findings and methodology used in those papers.

Much of the studies in literature investigate the effects of a large single transit investment by analyzing data before and after the opening of the transit investment. For example, Golias examines the impact of a new subway transit investment in Athens, introducing 29 additional transit stations (Golias, 2002). Revealed preference data is collected from roadside and onboard surveys. An extreme value mode choice model is estimated. It is found that the transit investment seems to have minimal effect on auto users. This is due to their inelasticity to switch travel mode because of travel cost and time variables. The author concludes that additional policies need to be made in conjunction with public transit investment to support the growth of transit demands. Forsey et al. also uses an extreme value mode choice model to investigate the VIVA bus rapid transit (BRT) system in York region of Ontario (Forsey, Habib, Shalaby, & Miller, 2019). Authors conclude that transit improvements have a small effect on post-secondary student trips.

Xie investigates the effect before and after the opening of three subway lines in Beijing (Xie, 2016). Data comes from the 2007 to 2009 annual household travel survey providing travel diary data. Using a difference in differences method, rail transit significantly increased for commuters who live in affected areas. These affected areas are traffic area zones where the centroidal distance to the nearest subway station decreases in 2008 or 2009. Most of the modal choice to rail is for auto users rather than the bus. The transit investment does not change travel behaviour of unaffected areas.
Regression is used for the analysis of the 2001 national household travel survey of the United States. Spiller et al. examine the effect of the gas tax and public transit investment on auto usage (Spiller, 2014). Their quantification of transit investments is based on the difference in individual predicted commute times of auto and public transit to represent time tradeoff. Improving transit improvement is found to decrease annual vehicle miles traveled (VMT). The gas tax is found to be more significant than transit improvement in predicting VMT.

Other methods for analyzing transit investment include extensive descriptive statistics as done by Vuk and Termida. Vuk finds that transit users using the Copenhagen metro shift their mode choice from bus more so than auto (Vuk, 2005). A new tram extension in Stockholm is found to be quickly adopted for use by middle-income travelers and car owners (Termida, 2016).

Chatterjee investigates the effect of the Fastway BRT service in England using a four-wave genuine panel survey (Chatterkee, 2011). The waves are within months of the opening of the Route 20 service. Travel behaviour and demographic information are collected for 186 individuals. The author models ordered response models with pooled, random effect structure, among others. It is found that there is a gradual increase in bus use over time after the introduction of the service. The author states that this implies the increased awareness of the service. It is also found that coefficients of the level of service variables are underestimated for the pooled model.

Along with an extensive review of pseudo panel methodology, Tsai et al. conduct a pseudo-panel analysis of the Sydney household travel survey (Tsai, Mulley, & Clifton, 2014). The goal of this study is to determine the short-term and long-term public transit demand elasticity. As the household travel survey contains repeated cross-sectional data, the authors transform the data into a pseudo-panel dataset. The grouping criterion for this dataset is the distance from the household to the central business district and birth year of individuals. A pooled and fixed effect model is estimated with independent variables income, bus frequency, among others to predict the number of transit trips per individual. It is found that the elasticity of cost is -0.22 and -0.29 for the short and long term respectively.

Cornut investigates car travel demand and ownership in Paris using a pseudo-panel analysis (Cornut, 2016). Data consists of five cross-sectional travel surveys between 1976 and 2010.
Cornut creates the pseudo dataset by categorizing households into three-year intervals of the birth year of household respondents to create homogenous individuals. The study uses a fixed effect model to determine the generation effect of the household over the years. Important explanatory variables for car travel demand include income, the price of gas, age, among others. It is found that the elasticity for the price of gas is -0.22. The author concludes that elasticity is more critical in high population density areas due to the number of alternative modes available for use.

2.6 Chapter Summary

In terms of the reviewed longitudinal papers assessing transit accessibility, a common observation is the lack of socioeconomic attributes of individuals. Advantages over the count and gravity measure include incorporating destination and socioeconomic attributes of individuals to better estimate accessibility. Disadvantages include an increase in complexity and data requirements. In addition, comparability between different models is problematic because the measure is relative to the model.

In terms of longitudinal effects of transit accessibility and TOD on transit demand, there are several differences and similarities between the papers. Transit investment is often defined as a single major piece of infrastructure though some define transit investment as accessibility over time. All data is revealed preference as they come from surveys (either household travel or panel surveys). Most data is short term than longitudinal, which may reflect upon the difficulty of collecting long-term data. Various methods are used, including mode choice and regression. Pseudo panel analysis is more involved where a pseudo-panel dataset is created by aggregating the data with homogenous individuals. As for the effect of transit investment on travel behaviour, in general, transit investment improves transit demand. The degree of this effect is variable. Most have been minimal to moderate, but the aggressive transit expansion in Beijing seems to be an exception to this.

This thesis adds to existing literature by presenting empirical research of measuring the logsum accessibility utility measure by transit over time for a population. Location choice models are estimated on workers and student datasets incorporating year specific interaction terms to estimate accessibility over time. The datasets cover years from 2001 to 2016 focusing on Greater
Toronto Hamilton Area (GTHA) residents and the regions’ transit systems and GO transit. TOD is considered as a collection of two variables: mixed land use and presence of transit station(s).

Additionally, this thesis adds to the transit investment literature involving longitudinal analysis. Transit investment is defined as new transit stops and stations and the degree of TOD developed from 2001 to 2016. A pseudo-panel analysis and subject-specific regression are conducted considering multiple forms of transit investment variables.
Chapter 3

3 The GTHA Data Context

3.1 Chapter Overview

This part of the dissertation goes over the socioeconomic and transportation behaviour trends of the GTHA from 1996 to 2016. Section 3.2 introduces the purpose and background. The descriptive analysis is split into a description of the GTHA and the data sources in Section 3.3, the socioeconomics of the population in Section 3.4, and the travel behaviour in Section 3.5. The chapter ends with a summary in Section 3.6.

3.2 Introduction

The chapter describes the aggregate trends the GTHA region exhibits in terms of sociodemographic variables and travel behaviour in the past twenty years. The purpose of this descriptive analysis is to determine the long-term trends of the GTHA population. This establishes the data context which aids in informing the modelling specification.

Much of this work is a continuation of work completed by Miller and Shalaby, 2003 for the Neptis Foundation as part of the “GTA Portrait” report (Miller & Shalaby, 2003). This report is meant to present a “comprehensive view of the greater Toronto urban system”. Their work describes population and travel behaviour using data from the MTARTS (Metropolitan Toronto and Region Transportation Study) conducted in 1964 and the 1986, 1991, and 1996 iterations of the Transportation Tomorrow Survey (TTS). There is a gap between 1964 and 1986 as there is no comparable comprehensive surveys of the GTHA at the time period. Survey methodology between the MTARTS and TTS are different but are compared to have a 32-year time span for longitudinal descriptive analysis.

Starting from 1996 to 2016, similar graphics are shown and observations over time are analyzed using the TTS. Trends that differ from the report by Miller and Shalaby are highlighted.

3.3 The Region and Data Sources

The study area of interest is the GTHA. It is in southern Ontario, home to 20% of Canada’s population in 2016, and houses one of Canada’s major economic centres. The region currently
contains transit systems unique to each municipality (e.g. the TTC of Toronto) and one regional transit system (GO transit).

The evolution of land use and policy in the GTHA has been typical to other urban areas, where there is a transition from decentralization to densification policies. To stem urban sprawl, the province of Ontario enacted policies including the Greenbelt Act of 2005 and the Metrolinx Act of 2006 among others (White, 2007). The former legislation prevents urban development by placing a buffer around the Greater Golden Horseshoe (GGH), in which the GTHA is at its centre. The Metrolinx Act established a regional transportation body responsible for planning an integrated transportation system. The agency established its first regional transportation, The Big Move, in 2008 which not only included expansions of public transit, but also establishing “mobility hubs” which are synonymous with TOD (Metrolinx, 2013).

The evolution of public transit saw large expansions between 1964 and 1986, especially with the TTC and GO transit (Miller & Shalaby, Evolution of personal travel in Toronto area and policy implications, 2003). The period after has incremental increases in transit accessibility. However, recent expansions have been developed including the VIVA bus rapid transit system in 2005, expansion of the TTC subway and bus services, GO transit, and among the other municipalities.

To support longitudinal analysis, the TTS is used. It is the household travel survey that is conducted in the GGH every five years since 1986. Cross-sections of data are extracted from 1996 to 2016. Specifically, the data for this analysis is pulled from an online resource called IDRS which is provided by the Data Management Group (DMG) of the University of Toronto. It should be noted that records with “unknown” as the attribute value are included due to the analysis involving making inferences of the entire population. These values are based on the expanded numbers of the TTS.

As a side note, Section 3.4.1 and Section 3.4.2 compares between the TTS and census and the rest of the sections only use TTS.
3.4 Socioeconomics

3.4.1 Population by Age Cohort

Figure 1 shows the distributions of the GTHA population by age cohort from 1996 to 2016 at five-year intervals using TTS data. In general, there has been a population increase in the GTHA as the distribution seems to vertically shift up each year. Prior to 2011, there is a large proportion of the population between the ages of 30 and 50. 2011 and 2016 exhibit a dramatic increase in the 15 to 29 age cohort category.

![Figure 1. GTHA Population by Age Cohort and Year (TTS Data)](image)

Figure 2 shows the same distributions as figure 1 albeit using census data instead of TTS. The purpose in showing this figure is to highlight the similarities and differences between the TTS and census. Most of the population is still the 50 to 54 category and the dramatic increase in the 21 to 25 age category is less pronounced. This difference is most likely due to the data collection procedure between census and TTS. The population is mandated by the Canadian government to complete the census whereas the TTS can only sample from the GGH population. A consistent result until recently has been a lack of 20 to 24 years surveyed in the TTS (Malatest, 2018).

The implications of this result are similar to the one made by Miller and Shalaby. The population of the GTHA is increasing from aging, the net effect of migration, immigration, and death contributes to the increase use of transportation systems. The steady rise in 20 to 40 year olds, whom are of working age, suggests additional travelers during peak hours.
3.4.2 Population by Region

Figure 3 shows the GTHA population distribution by household municipality from the TTS. Clearly, there is a decentralization pattern. Toronto as the core has a reduced proportion of the GTHA population over time which is taken up by the neighbouring regions. However, this effect is less pronounced as times goes by. Interestingly, the region of Hamilton is experiencing a slight decrease in population share as well.
3.4.3 Household Size

Figures 4 and 5 show the household size distributions by year using TTS and census data respectively. Generally, there is an increase in two-person households. Surprisingly, the 2016 TTS has an increase in the proportion of one-person households, after a downward trend from 1996 to 2011. This may reflect on the nature of TTS data collection or expansion of the data. The census data exhibits a steady decrease of one-person households, similar rates for two-person and three-person over the time period, and a higher proportion of 4+ person households. This differs

Figure 4. Household Size Distribution by Year (TTS Data)

Figure 5. Household Size Distribution by Year (Census Data)
from the Neptis report conclusion where the increasing trend of two-person households seems to have diminished; it has been replaced by more households with 4+ members.

3.4.4 Number of Workers

There has been a steady increase in the number of workers per household through all the regions of household as per figure 6. Interestingly, households residing in Toronto and Hamilton have on average less workers than the other municipalities.

![Figure 6 Average Number of Workers per Household by Region and Year](image)

**Figure 6 Average Number of Workers per Household by Region and Year**

![Figure 7. Employment by Region and Year](image)

**Figure 7. Employment by Region and Year**

Figure 7 shows the proportion of GTHA workers whom work in different regions of employment. Most workers work in Toronto and a sizeable number go to PD1. PD1 is the
downtown of Toronto which houses the financial district among other central business district establishments. From 1996 to 2006, there has been growth in employment in York and Peel. Recent years seem to show that regional employment is stagnating across the GTHA.

### 3.4.5 Self-Containment

Self-containment, in this context, means a worker living and working within the same defined region. For example, the 2016 iteration of the TTS shows that 64% of the worker population living in planning district 1 (PD1), also live in PD1 as per figure 8. These values are calculated by dividing the number of workers whom living and work in the same region by the number of workers whom live in said region. Across the different disaggregated regions of the GTHA, there seems to be a decline in people living where they work over time. This implies longer commute times for workers living in the GTHA.

![Figure 8. Proportion of Workers Living and Working in Same Regions as Residence](image)

Figure 8 shows people working in PD1 continually live within or in the surrounding planning districts around their workplace. Over time, more PD1 workers live in Toronto with slow growth in other regions. Hamilton seemingly exhibits very low number of PD1 workers. These trends show that Toronto is densifying greatly with respect to PD1 workers; people working in downtown Toronto are increasingly living within Toronto. However, PD1 is gradually losing self-containment as shown by the other municipalities.
Figure 9. Residential Region of PD1 Workers by Year

3.5 Travel Behaviour

3.5.1 Average Daily Trips

Figure 10 shows a relatively stable and decreasing trend of average number of trips by household and trip type made by GTHA residents. Trips are usually separated into home-based where a trip has home as an origin or destination and non home-based trips does not.

Figure 10. Average Number of Daily Trips made by GTHA Residents by Year
3.5.2 Average Daily Trips by Mode

Much of the trips made in the GTHA have been auto-oriented as shown in Figure 11. However, there is a seemingly decrease in the average daily trips made by auto drive and auto passenger over time. There is a steadily increase in mode share of local transit and regional transit. Rideshare (also known as ride hailing or transportation network companies) started in Toronto between 2011 and 2016 and so not much of a conclusion can be made for these services. Active modes have increased in mode share.

For clarification, trips are considered to be made by local transit if transit is used as the made mode with the exclusion of GO rail. The other category is made up of modes motorcycle, school bus, and unknown. The active mode is made up of cycling and walking.

![Figure 11. Average Number of Daily Trips made by GTHA Residents by Mode and Year](image)

3.5.3 Mode Share by Year

Figure 12 again shows that most trips made in the GTHA have been auto-oriented. There has been a slight increase in active modes and local transit, but overall the trends over the time frame has been consistent.
3.5.4 Number of Vehicles per Household

In terms of number of vehicles, trends have been relatively steady. Figure 13 displays the average number of vehicles per household by year and region. It is clear that there is an increase in cars in the 905 region whereas there is a decrease in Toronto.
3.5.5 Spatial Flows

Table 1. Difference in Regional Spatial Flow for Morning Peak Period, 1996 and 2016

<table>
<thead>
<tr>
<th>Origin/Destination</th>
<th>Toronto</th>
<th>Durham</th>
<th>York</th>
<th>Peel</th>
<th>Halton</th>
<th>Hamilton</th>
<th>Total</th>
<th>Total %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toronto</td>
<td>160,403</td>
<td>4808</td>
<td>16,600</td>
<td>12,394</td>
<td>3,483</td>
<td>616</td>
<td>198,304</td>
<td>21%</td>
</tr>
<tr>
<td>Durham</td>
<td>18,838</td>
<td>52,294</td>
<td>10,372</td>
<td>1,650</td>
<td>235</td>
<td>75</td>
<td>83,464</td>
<td>9%</td>
</tr>
<tr>
<td>York</td>
<td>59,052</td>
<td>3453</td>
<td>178,803</td>
<td>15,788</td>
<td>1,774</td>
<td>430</td>
<td>259,300</td>
<td>28%</td>
</tr>
<tr>
<td>Peel</td>
<td>17,149</td>
<td>492</td>
<td>14,557</td>
<td>201,153</td>
<td>11,890</td>
<td>3118</td>
<td>248,359</td>
<td>26%</td>
</tr>
<tr>
<td>Halton</td>
<td>6,686</td>
<td>149</td>
<td>2,309</td>
<td>25,548</td>
<td>73,324</td>
<td>4,686</td>
<td>112,702</td>
<td>12%</td>
</tr>
<tr>
<td>Hamilton</td>
<td>1,138</td>
<td>160</td>
<td>643</td>
<td>3,447</td>
<td>9,271</td>
<td>21,955</td>
<td>36,614</td>
<td>4%</td>
</tr>
<tr>
<td>Total</td>
<td>263,266</td>
<td>61,356</td>
<td>223,284</td>
<td>259,980</td>
<td>99,977</td>
<td>30,880</td>
<td>938,743</td>
<td></td>
</tr>
<tr>
<td>Total %</td>
<td>28%</td>
<td>7%</td>
<td>24%</td>
<td>28%</td>
<td>11%</td>
<td>3%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1’s values are the difference in the number of trips made from specific origin to destination municipalities of 1996 and 2016 TTS years (2016 trips minus 1996 trips). The number of trips is the morning peak period which is defined as the start time starting from 6:00 to 9:00 AM. The column and row are origin and destination respectively.

Toronto continues to be the main attractor of morning peak period trips. Peel and York are major destination regions as well. There has been major growth in the number of intraregional trips from 1996 to 2016 for York and Peel more so than Toronto. The other municipalities Durham, Halton and Hamilton have had relatively minimal growth. However, the region has experience growing travel demand since 1996 as all differences are positive.

3.6 Chapter Summary

The chapter shows a descriptive analysis of socioeconomics and transportation behaviour of GTHA population. As a region, it has experienced population growth where regions outside the core are experiencing faster growth than Toronto. In terms of age groups, the proportion of 20 to 24 year olds is growing steadily over the years. The descriptive analysis also suggests that workers are increasingly not living in the regions where they work.

In terms of travel behaviour, mode share and trip generation are of interest. Mode share has been steady through 1996 to 2016. Modes related to auto dominate the mode share and transit and active modes are slightly increasing over time. Ridesharing services only exist in the 2016 iteration of the TTS and are a small part of the mode share. In addition, there has been a
significant increase in the number of trips generated to destinations to Toronto, York, and Peel. Most trips are generated within these regions with more growth in York and Peel than Toronto.

The implications of this is twofold. There is more people and the locations in which they decide to live suggest longer commute times for the people living in the GTHA. It seems that the symptoms of urban sprawl are still present within this region.
Chapter 4

4 A Longitudinal Analysis of Accessibility by Transit to Jobs and Schools Using the Utility-based Accessibility Measure

4.1 Chapter Overview

This chapter describes the process of descriptive analysis of transit accessibility and TOD trends. The research design is introduced in Section 5.2. The data, methodology and empirical models are presented in Section 5.3, Section 5.4, and Section 5.5 respectively. Section 5.5 discusses the implications of the models and Section 5.6 summarizes the chapter.

4.2 Introduction

The objective of this chapter is to measure accessibility over time. Observable trends and data context are established via descriptive analysis. Discrete choice models are estimated, and results are discussed. Additionally, distributions and maps are generated for visual representation of accessibility over time. This paper attempts to answer what longitudinal effects the transit policy has had on person accessibility.

4.3 Data for Empirical Investigation

Much of the 2001, 2006, 2011, and 2016 data comes from the TTS. It is a household travel survey conducted in the Greater Golden Horseshoe since 1986 and provides travel and socioeconomic data every five years since then (Malatest, 2018). This data is augmented by the Canadian census and land use data from DMTI Spatial. Additional socioeconomic variables, such as median income, are extracted from the census at the zonal level. Datasets from DMTI Spatial make available enhanced points of interest (EPOI) data which consists of business and recreational establishments across Canada categorized by SIC or standard industrial classification codes (DMTI, 2015). The DMTI Spatial datasets also include shape files of land use types. Unfortunately, this data is only available up to 2014 and so it is assumed that 2014 land use in the GTHA is the same as 2016. Another limitation of the land use files is that they do not consider floor space area; the data is only the land cover area of each land use category.
Zonal level of service (LOS) variables are generated by on the GTAModel V4.0 built by the Travel Modelling Group (TMG) at the University of Toronto. GTAModel V4.0 is an agent-based travel demand model which ultimately generates travel times in EMME (Miller, Vaughan, Reilly, & Yusuf, 2018). As transit accessibility is of interest the LOS variables are the travel times using transit with walk access (TWA). TWA travel time is further separated into in-vehicle travel time (IVTT), wait time, and walk time. Due to data availability, only AM peak travel time values are used.

Figure 14 attempts to show mixed land use trends from 2001 to 2016. Mixed land use is measure as entropy and the formula is in (1). \(i\) is the land use category, \(n\) is the total number of land use categories, and \(P_i\) is the proportion of land use within an area.

\[
\text{entropy} = (\text{stops} > 0) * - \sum_{i}^{n} P_i \log \left( \frac{P_i}{n} \right)
\]

(1)

The inequality portion of the formula is an indicator where it is a value of 1 if there is at least one transit station or stop within the boundaries of the area and 0 if not. The area in this case is defined as the traffic analysis zone (TAZ). The summation portion of the formula is the land use entropy which is normalized to values from zero to one. A higher land use entropy means a higher degree of mixed land use. Commercial, residential, and institutional land use categories are used for this formula as they are associated with TOD land use. It should be noted that if the proportion of a land use category is zero, the entire term in the summation is manually adjusted.

Figure 14. Average Mixed Land Use Entropy of TAZ by Municipality and Year
to zero to avoid undefined values of entropy. The entropy is calculated for each TAZ in each year and municipality and averaged.

Figure 15 shows the average transit stop density for each year and municipality averaged over TAZ. Overall, the graphs seem to show a gradual trend between 2001 to 2016. Average stop density shows a gradual increase whereas the mixed-use entropy decreases. The main driver behind the entropy decrease is a rise in residential land cover area.

![Figure 15. Average Stop Density of TAZ by Municipality and Year](image)

It is important to note that the data is not a fully representative sample of the population due to sampling issues with TTS and its sample frame. In particular, the TTS does not survey university students living away from home and students below eleven years old. There are other general issues with the TTS and its sample frame including overrepresentation and underrepresentation of seniors and young adult populations respectively (Lo, Srikukenthiran, Miller, & Habib, 2019).

### 4.4 Methodology of Empirical Investigation

#### 4.4.1 Model Formulation

In order to be able to estimate the utility accessibility measures, a discrete choice model needs to be developed. The discrete choice model is based on random utility maximization where it is
assumed that decision makers are rational. If a decision maker is presented with a number of choices, the decision maker is assumed to always choose the choice that gives the most satisfaction or utility. These models are developed to attempt to mimic people’s decision making behaviour when presented with a set of alternatives. In this specific context, location choice models are estimated to determine how accessible one’s home TAZ is to one’s choice set of destination TAZs by the mode of transit walk access (TWA). A review paper authored by Schirmer et. al provides a succinct overview of location choice models used in transportation literature (Schirmer, van Eggermond, & Axhausen, 2014).

Regarding choice set formation, there are about 2000 TAZs in the GTHA. The full choice set can be assigned to each person, but may not be realistic and will take much computation time to test different model specifications. To this end, there are several methods to reduce the choice set size by sampling from a feasible choice set. There are several ways to define a feasible choice set including random sampling, threshold sampling, importance sampling, use of anchor points, etc. (Zolfagahri, Sivakumar, & Polak, 2012; Elgar, 2009) In this case, a combination of threshold and importance sampling to create an individual-specific feasible choice set from which to sample from.

From the full choice set, a subset of TAZs are taken for each individual where the total TWA time from the individual’s home TAZ is less than or equal to a threshold value. This threshold value is the 85th percentile of the total TWA time from individuals’ home TAZs to their chosen TAZs. From this subset, TAZs are selected if they have at least one job or school category of the individual (job and school accessibility). For example, at least one person working in the service industry or at least one person going to high school. This is the feasible choice set where nine TAZs are weighted sampled by the number of job or school category and included into the choice set with the chosen alternative.

Having a choice set size of ten is an arbitrary choice. A paper done by Nerella and Bhat suggests that increasing choice set size yields improvements in estimation with diminishing returns (Nerella & Bhat, 2004). It advises at a minimum of sampling one-eighth of the feasible choice set for each person, although the results are based on simulated data. In this paper, choice sets containing ten TAZs of the feasible choice set are formed.
It is acknowledged that this method of choice set formation is inherently incorrect. TAZs are subject to the modifiable areal unit problem, but the LOS variables from the GTA model are zone-based. Especially for non-university students, their choice set is not only dependent on home location but also parents’ dispositions (Elacqua, Schneider, & Buckley, 2006). Nonetheless, this choice set formation is done as a proxy for an individual’s decision making of choosing their choice set to choose from.

The utility equation for individual and alternative is shown in (2) (Train, 2009). In addition to the systematic and error components, there is an additional log term. This is the correction factor which needs to be added due to the importance sampling of the choice set formation (McFadden, 1974). Not doing so would make the model parameter estimates to be inefficient. The correction factor is the natural logarithm of the probability of choosing the choice set given the alternative. The coefficient for this term is strictly one and should not be estimated.

\[ U_{ij} = V_{ij} + \ln(P(C_i|j)) + \varepsilon_{ij} \]  

where

- \( U_{ij} \) is the utility for i individual and j alternative
- \( V_{ij} \) is the systematic component for i individual and j alternative
- \( C_i \) is the choice set of individual i
- \( \varepsilon_{ij} \) is the random error component for i individual and j alternative

If the random error component is assumed to distribute independent and identically distributed (IID) with a Type I extreme value distribution or Gumbel distribution, the following closed form of the probability equation is generated. This model is the conditional multinomial logit (cMNL) model (3), first presented by McFadden (McFadden, 1974). The probability of choosing an alternative \( j \) is dependent on the chosen choice set from the importance sampling.

\[ P_i(j|C_i) = \frac{\exp(V_{ij}+\log(P(C_i|j)))}{\sum_j \exp(V_{ij}+\log(P(C_i|j)))} \]  

(3)
There are several assumptions that are made with this model structure. A major assumption that is made when assuming IID is that all persons are assumed to exhibit the same behaviour. Another assumption as a result of using MNL is the IIA or irrelevant and independent assumption. IIA results in proportional substitution where a change in the probability of an alternative is equally distributed among the rest of the other alternatives. This is the trade off made for model simplicity. Full knowledge of the choices in the choice set is assumed as well.

4.4.2 Dataset for Empirical Investigation

To generate the datasets, attributes are extracted from data sources to associate with individuals and destination TAZs. TAZs attributes are divided by the area to attain densities. All variables are scaled, and unknown values are removed. The choice set is generated for each person. For cases where a person’s chosen choice is unknown, the closest TAZ that is in the feasible choice set is assigned. TWA LOS variables are then assigned to each TAZ in the choice set for each individual. It is important to note that individuals that use modes other than TWA are considered as accessibility of the population is of interest. In addition, LOS variables have values of zero if the destination TAZ in the choice set is the same as the home TAZ of the individual. Finally, destination TAZ attributes are assigned to the choice sets. Numerous software is used to support dataset development including R v3.6.1, Python v3.7.0, and ArcMap v10.2.2.

The dataset is segmented into workers and students. This is done as it is believed that both populations exhibit different travel behaviour. Doing so would also reduce run time of models and hopefully achieve a better model fit by rho square. A person is considered a worker if they have full-time status or is both part-time and not a student. A worker is categorized into general / professional, service, and manufacturing occupation types for the importance sampling. A person is considered a student if they both have a full-time student status and do not work or if they are a part-time student with either no work or part-time work status. Students are categorized into non-university and university where university students are aged eighteen and more.

The cMNL is estimated by forward selection after including a base model which includes year, mixed land use entropy, stop density, among other variables of interest. PandasBiogeme is used to estimate the models using the BHHH (Berndt-Hall-Hall-Hausman) optimization algorithm (Bierlaire, 2018).
4.4.3 Empirical Models

Table 2 shows the estimates of the location choice models developed. Model 1 and 2 were developed with a choice set size of ten for the student and worker dataset respectively. Through model development, it is found that removing individuals where their chosen alternative is greater than the 85th percentile drastically improves the rho square.

A brief description of variables used in the model follows. Mixed land use entropy, retail EPOI, and road density describe the land use of the destination TAZs and represent proxies for TOD. The LOS variables and stop density represent proxies for transit accessibility.

Table 2. Location Choice Model Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>148,842</td>
<td>488,352</td>
</tr>
<tr>
<td>Null log-likelihood</td>
<td>-40,931.29</td>
<td>-24,660.22</td>
</tr>
<tr>
<td>Final log-likelihood</td>
<td>-28,583.36</td>
<td>-12,486.04</td>
</tr>
<tr>
<td>Rho-square</td>
<td>0.30</td>
<td>0.49</td>
</tr>
<tr>
<td>TWA IVTT</td>
<td>-0.042 (-94.1)</td>
<td>-0.0238 (-61.8)</td>
</tr>
<tr>
<td>TWA walk time</td>
<td>-0.0203 (-37.7)</td>
<td>-0.0249 (-56.7)</td>
</tr>
<tr>
<td>Mixed land use entropy</td>
<td>-0.2560 (-11.0)</td>
<td>-0.0723 (-2.19)</td>
</tr>
<tr>
<td>log (retail EPOI density)</td>
<td>0.0428 (8.97)</td>
<td>0.3353 (54.1)</td>
</tr>
<tr>
<td>Road density</td>
<td>-0.1682 (-83.0)</td>
<td></td>
</tr>
</tbody>
</table>

**Interaction terms**

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-university x OD pair</td>
<td>1.1500 (27.3)</td>
<td>-</td>
</tr>
<tr>
<td>Year2006 x Stop density</td>
<td>-0.0154 (-8.1)</td>
<td>0.0317 (17.4)</td>
</tr>
<tr>
<td>year2011 x Stop density</td>
<td>-0.0150 (-8.06)</td>
<td>0.0433 (21.5)</td>
</tr>
<tr>
<td>year2016 x Stop density</td>
<td>-0.0197 (-9.86)</td>
<td>0.0394 (23.9)</td>
</tr>
<tr>
<td>Year 2006 x Home entropy x TWA Wait time</td>
<td>-0.0987 (-9.56)</td>
<td>-0.0187 (-2.24)</td>
</tr>
<tr>
<td>Year 2011 x Home entropy x TWA Wait time</td>
<td>-0.0542 (-4.71)</td>
<td>-0.0196 (-1.90)</td>
</tr>
<tr>
<td>Year 2016 x Home entropy x TWA Wait time</td>
<td>-0.0501 (-4.84)</td>
<td>-0.0192 (-2.29)</td>
</tr>
</tbody>
</table>

The purpose of the interaction variables is to incorporate sociodemographic variables of the individual like home entropy and year. This is so that the values can vary by alternative. Home
entropy is calculated by equation (1) but is home specific rather than TAZ. Circular buffer zones are created for each household point and spatial joined with land use to get home specific land use entropy. The year 2001 is not included as it is taken as the reference.

Model 1 has a large sample size and high rho square against the null model. All variables are highly significant with more than 95% confidence. Signs of the coefficients are expected; the LOS travel time variables are negative. The retail EPOI density is logged in order to scale the variable. To highlight a few variables, the “Non-university × OD pair” interaction variable is positive and high in magnitude. Non-university represents students whom are in high school and lower. OD pair is a binary value where 1 if the home to destination is within a municipality, else 0. This segment of students prefers to go to school TAZs within their same municipality.

Regarding the year interaction variables, the “year × stop density” variables show a slight increase in magnitude over the years, suggesting that students become more sensitive to stop density over time. The sign is negative which simply means that students tend to choose school TAZ that have low stop density. The “year × home entropy × wait time” variables also show a trend of decreased sensitivity over time. For a single year, the coefficient is interpreted that given the same wait time, people living in a more mixed land use area are more sensitive to wait time than those who live in a lower mixed area.

Model 2 again has a large sample size and a very high rho square against the null. Much of difference, comparing to Model 1 is the sign of the year stop density interaction variables. The sign of model 2 is positive meaning that workers tend to choose workplace TAZs with more transit stations and stops. Another comparison is the lack of trends in the year interaction variables; both sets are similar in magnitude.

4.5 Results and Discussion on Empirical Models

To compare behaviour between students and workers, direct elasticities are calculated for each individual and averaged. In this particular case, the alternative considered is the chosen choice. The formula (4) for direct elasticity produces the percent change in the probability of choosing an alternative \( j \), given a 1% change in a variable of interest.
Direct Elasticity \(= x_{ij}(1 - P_i(j))\beta_x \) \hspace{1cm} (4)

where

\(x_{ij}\) is the variable of interest of individual I and alternative j

\(P_i(j)\) is the probability of choosing alternative j for individual i

\(\beta_x\) is the coefficient of variable \(x\)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>TWA IVTT</td>
<td>-0.0742</td>
<td>-0.0148</td>
</tr>
<tr>
<td>TWA walk time</td>
<td>-0.0346</td>
<td>-0.0147</td>
</tr>
<tr>
<td>Mixed land use entropy</td>
<td>-0.0066</td>
<td>-0.0003</td>
</tr>
<tr>
<td>log (retail EPOI density)</td>
<td>0.0098</td>
<td>0.0084</td>
</tr>
<tr>
<td>Year2006 x Stop density</td>
<td>-0.0023</td>
<td>0.0011</td>
</tr>
<tr>
<td>year2011 x Stop density</td>
<td>-0.0022</td>
<td>0.0015</td>
</tr>
<tr>
<td>year2016 x Stop density</td>
<td>-0.0026</td>
<td>0.0014</td>
</tr>
<tr>
<td>Year 2006 x Home entropy x TWA Wait time</td>
<td>-0.0043</td>
<td>-0.0003</td>
</tr>
<tr>
<td>Year 2011 x Home entropy x TWA Wait time</td>
<td>-0.0020</td>
<td>-0.0003</td>
</tr>
<tr>
<td>Year 2016 x Home entropy x TWA Wait time</td>
<td>-0.0018</td>
<td>-0.0003</td>
</tr>
</tbody>
</table>

The results are shown in Table 3. The first observation is that the direct elasticities are small. For example, increasing the log of retail EPOI density by 1% results in only a 0.0084% increase of choosing the chosen alternative of workers. A likely cause is the high probability of the chosen alternative, relative to the other choices in the choice set. This is driven by the small size of the choice set in models 1 and 2. In other words, the probability of choosing the chosen alternative is rather inelastic because it is the best choice.
The second observation is that the direct elasticities of the LOS terms of students are higher than workers. This suggests that students are more sensitive to LOS than workers. This seems to be a non-sensical result as generally workers are obligated to go to the workplace and thus should be more sensitive to longer travel times. Perhaps the reason lies in the threshold sampling in creating the choice set for students and workers. The 85th percentile of total transit time for workers and students is 111.2 minutes and 81.9 minutes respectively. Workers are more willing to go to farther workplaces than students. Non-university students, which make up 67% of the student sample, tend to go to a school close to their home by walking. Going to a further school forces a non-university student to take transit resulting in non-zero wait time and IVTT. Further segmentation may be required to separate university students from the general student population.

To answer the research question of how accessibility has changed over time, the logsum accessibility measure is calculated for each person. These measures are displayed in figures 16 and 17 as cumulative distribution functions (CDF) and in figures 18 and 19 as maps. The formula (5) is as follows.

\[
\text{Accessibility}_\text{utility}_i = \ln(\sum_j V_{ij})
\]

The formula comes from a paper by Ben-Akiva and Lerman (Ben-Akiva, 1985). It is essentially taking the natural logarithm of the denominator of the cMNL probability model. The higher the value, the more accessible the person is to destinations by transit. The more choices are in the choice set, the more systematic terms, and the higher the accessibility utility measure. That is not to say that one should then model on the full choice set, but rather the measure is a relative one.

For student accessibility over time, the figure 16’s CDF shows a slope of zero at the interval between 1.75 and 2 of the accessibility utility measure. This suggests that there are different segmentations of the student sample. This behaviour is driven by the non-university students which is enabled by the large positive coefficient of the “non-university * ODpair”. This also again enabled by the low travel times to student’s chosen alternative. Over time, the region seems to have been a slight increase from 2006 to 2011 and 2016, taking 2001 as the reference.

Conducting an Anderson-Darling k-sample test yields a test statistic of 299.2 and is significant. The statistical test is non-parametric and tests the null hypothesis that all samples do come from
the same population. Having a significant value means that the null hypothesis is rejected and the accessibility distributions by year are statistically different. Nonetheless, the values change little over time.

The CDF of figure 17 shows a more typical normal distribution. The accessibility measures on the high end are driven by the chosen alternatives’ high stop and retail density of workers. The 2006 to 2016 distributions are aligned closely suggesting again that there has been little change in accessibility utility measure over time. The Anderson-Darling k-sample test yields a test statistic of 412.0 which is significant as well.

Figure 18 and 19 show a visual representation of accessibility utility measure change over time by students and workers etc. The value of each TAZ is calculated by taking the sample average of the logsum accessibility for each year. As a side note, the legend is the same for each map. In addition, the maps are focused on Toronto instead of the entire GTHA because it is the municipality with the most data. There are several TAZs where there are no students and workers sampled in the areas marked in light grey. Taking 2001 as the reference, again there seems to be little change over time from 2006 to 2016.
Figure 17. CDF of Logsum Accessibility Utility Measure by Year, Workers

Figure 18 shows that students outside of Toronto seem to have better transit accessibility than those living in. This is largely driven by non-university students whom go to school within or adjacent to their home TAZ, effectively meaning that students need only to walk to their chosen alternative instead of walking to transit.

Overall, these results suggest that the logsum accessibility utility measure does not change over time. This is indicative of the gradual trends in transit accessibility and mixed land use as shown in the data section. The time horizon should be extended in order to capture effects of greater changes of prior years.

4.6 Chapter Summary

This chapter presents location choice models developed with year interaction variables in order to estimate logsum accessibility measures by year. The model is segmented into workers and students to model different travel behaviour and improve the goodness-of-fit value. Results show that for both segments, residents living in a higher mixed land use area are more sensitive to travel time. Students seem to be more sensitive to travel time and cost variables than do the workers. The logsum accessibility measure by year shows minimal change over time which seems to be indicative of the gradual process of stop density and mixed land use. This highlights
Figure 18. Accessibility by Transit Map for Students by Year, Centred in Toronto

Figure 19. Accessibility by Transit Map for Workers by Year, Centred in Toronto
that fact the, transit induces change in urban form (land use pattern) can take a long time to be realized. Regional urban and transportation policies need to be better informed about such slow changes and thereby reflected in project evaluation processes.

Several improvements can be made to enhance model results and interpretation. Different choice set sizes can be tested for its effects on estimates and the logsum accessibility measure. Experimenting with different methods to generate choice set or using the full choice set should be investigated. Further segmentation, especially between university and non-university students would better improve understanding of model coefficients and accessibility. The time frame can be extended to 1986 or even later to examine sensitivity of the logsum accessibility measure through periods of changing transit accessibility and TOD. The population can be expanded to people whom are neither workers or students (e.g. retirees) to understand their accessibility over time. Lastly, the MNL can be improved to a nested logit, mixed logit, etc. to reduce assumptions. An ideal model structure could be a cross-nested structure where the first level is the transit stops to model access to transit and a second level is the destination to model access by transit. This structure would consider the whole trip holistically rather than ignoring the first mile component of a transit trip.
Chapter 5

5 A Macro Model Analysis of the Relationship between Transit Investment and Transit Ridership in the GTHA

5.1 Chapter Overview

The analysis of the GTHA population is presented in this chapter. Section 5.2 details the introduction and research design. Section 5.3 and Section 5.4 describes the data and methodology respectively. This is followed up by the empirical model results and discussion in Section 5.5 and Section 5.6. Finally, Section 5.7 the effects of the variables of interest are summarized.

5.2 Introduction

To identify the effect of transit accessibility and TOD on transit usage, pseudo-panel and subject-specific models are developed using the repeated cross-sectional TTS from 2001 to 2016 (Malatest, 2018). The TTS provides a consistent time series of travel information to determine the longitudinal impact of transit investment. In other words, to what extent does built transit stops and stations, and mixed land use affect transit usage.

5.3 Data

The data used come from multiple sources. Much of the trip and demographic variables are extracted from 2001, 2006, 2011, and 2016 TTS datasets (Malatest, 2018). The TTS surveys five percent of the GGH households, which translates to roughly 150,000 households recruited per TTS year. The trip variables come from a one weekday trip diary per household. These variables include the primary mode of travel, trip purpose, start time, among others. Demographic variables are also asked, such as age, household size, and the number of full-time workers.

The TTS data is supplemented by land use and census data. Land use data come from DMTI spatial shapefiles. These annual shapefiles contain polygons of Canada categorized by land use. These categories include residential, institutional, industrial, etc. Using ArcGIS software, the shapefiles are spatially joined with the 2006 traffic area zone (TAZ) shapefile of the GGH. Land use area of each polygon in the GGH is calculated using ArcGIS as well. It should be noted that
these shapefiles only go up to 2014. The 2016 land use is assumed to be similar to the 2014 land use. Median income before tax and population density is drawn from the census for each corresponding TTS year. This data is extracted at the census tract zonal level.

The datasets are fused with the TTS dataset via location (i.e., census tract, dissemination area, and TAZ). Households that are not within the GTHA are removed.

5.4 Methodology

The following presents two methods in which to develop macro models: pseudo panel and subject-specific regression.

5.4.1 Pseudo Panel Analysis

To investigate the effect of the transit investment variables on transit usage, a pseudo-panel analysis is first conducted. This analysis differs from a true panel analysis, as panel data follows the same individual samples through time. The TTS collects repeated cross-sectional data from different samples drawn and thus, individuals are different for each iteration.

A significant step in the pseudo-panel analysis is the development of a pseudo panel dataset. This involves grouping households based on a criterion of variables which are time-invariant for each time point (Kasraian & al., 2017). The purpose of this is to create homogenous “individuals” which can be followed through time. These individuals can then be modelled using conventional panel analysis. In this analysis, households are grouped with the variables of the birth year of the respondent and sixteen planning districts (PD) that their household resides in. The birth year intervals are as follows: before 1945, 1945 to 1964, 1965 to 1979, and after 1979. These intervals are created roughly based on generation. It should be noted that the planning district is considered to be time-invariant. Therefore, there are 64 different ‘individuals’ (4 birth year intervals x 16 PD).

As the time dimension has two-time points, the total number of data points is 128. In order to group the households into the ‘individuals’, the independent and response variables are averaged by the ‘individuals’. As such, an additional requirement is to have at least 100 households per ‘individual’ in order to be statistically sound. This requirement results in 99 records left in the pseudo panel dataset.
There are several models available for pseudo panel data. For this paper, the following pseudo panel linear regression models are considered: pooled, fixed, and random.

\[
\overline{y}_{it} = \beta x_{it} + \nu_i + \epsilon_{it} \tag{6}
\]

The basic pseudo panel linear regression is shown in (6) where \( i \) represents the individuals and \( t \) represents the time series. \( y \) is the response variable; the systematic component is made up of the coefficient estimates, \( \beta \), of the independent variables \( x \) (Croissant & Millo, 2008). The error component is made up of \( \nu \) and \( \epsilon \). The error component’s first variable is the random individual-specific error, and the second is the idiosyncratic error; it is independent of the coefficients and \( \nu \).

If it is assumed that there is no individual-specific effect, the pooled pseudo panel model can be estimated (Kasraian & al., 2017). In other words, all the individuals are assumed to exhibit the same behaviour. As a standard linear regression model, it is estimated using ordinary least squares (OLS). This assumption may prove to be false as it ignores potential temporal and individual effects.

If the individual-specific error is assumed to be correlated with the \( \beta \) coefficients, the OLS estimates of the pooled model will be inconsistent. The fixed effect model overcomes this by estimating coefficients of dummy variables of the individuals along with the independent variables (Russell & Fraas, 2005). All independent and response variables are centred on the mean which allows deduction of the individual-specific effect. The fixed effect model is also estimated using OLS (Guillerm, 2017).

If the individual-specific error is assumed to be uncorrelated with \( \beta \) coefficients, a random effects model can be used. The individual-specific error is assumed to be independent and identically distributed (Kasraian & al., 2017). Instead of OLS estimation, general least squares is used. To decide whether the fixed effect or random effect model is more appropriate, the Hausman test can be used (Kasraian & al., 2017). The null hypothesis tests whether the individual-specific errors are uncorrelated with the independent variables (i.e., the random error model is preferred).
In this analysis, the individuals are the birth year interval – PD factors and time is the two TTS years, 2011 and 2016. The dependent variable is the number of household transit trips. Forward selection is used; independent variables are tested and added one at a time, ensuring each is significant. Due to the focus on understanding the effect of transit investment on the number of transit trips, transit investment variables are tested first. For each model, collinearity and normality assumptions are checked. R software is used to estimate the models using the plm package (Croissant & Millo, 2008). The estimation method for the random effects is based on work done by Swamy and Arora (1972).

5.4.2 Subject-Specific Models

Subject specific macro model regression differ from pseudo panel models in that there is no need to transform the disaggregated data into a panel dataset. Instead, the raw data is used as is. Advantages include many samples and use of more disaggregated data compared to the aggregated approach of panel models. Disadvantages include the assumption of homogeneity. For example, controlling for time points means one assumes that each subject is homogenous in each year. Another disadvantage is the time for each run, especially when using maximum likelihood. The subject of this analysis is the household.

In order to develop valid causal inferences for an entire population, weights factors should be incorporated in the maximum likelihood procedure. It is found that the default methods for regression in R (packages GLM and zeroinfl) allows inclusion of a weight vector. However, this weight vector is a frequency weight, which effectively expands the dataset by making multiple copies of households. This is not appropriate for sample or probability weights. To handle this, the survey package in R is used (Lumley, 2004). The estimator is the Horvitz-Thompson estimator. It is a pseudo likelihood, so it is somewhat inefficient. However, the estimator is unbiased and produces the correct standard errors.

The multiple model specifications are described in chapter 5.5.2.

5.5 Empirical Models and Discussion

The following describes the pseudo panel and subject-specific analysis. Pseudo panel models were first developed as an experiment on a dataset of only Toronto households. These models
had relatively poor performance compared to the subject-specific models. The subject-specific models are developed on the full GTHA household dataset. For completeness sake, the pseudo panel models are shown here.

5.5.1 Pseudo Panel Empirical Models and Discussion

The pooled model results in an adjusted R-squared value of 0.608 and several variables (Table 4). The response variable is the natural logarithm of the number of transit trips made by each homogenous birth year interval – PD individual. This response variable is shared with the fixed effect and random effect models. The natural logarithm is taken to reduce the heteroskedasticity and the skewness of the raw response variable. Variables include the 5-year difference TTC express bus transit investment variable along with several sociodemographics. Signs of variables are reasonable. Assumptions of normality and no collinearity are met. Normality is investigated via viewing normal quantile-quantile plots and residuals versus fitted plots. The linear

Table 4. Pseudo Panel Model Results

<table>
<thead>
<tr>
<th>Pseudo Panel Model</th>
<th>Pooled</th>
<th>Fixed Effect</th>
<th>Random Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Observations</td>
<td>99</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.637</td>
<td>0.742</td>
<td>0.624</td>
</tr>
<tr>
<td>Adjusted R-Squared</td>
<td>0.613</td>
<td>0.722</td>
<td>0.612</td>
</tr>
<tr>
<td>Variable</td>
<td>Coefficient (t-value)</td>
<td>Coefficient (t-value)</td>
<td>Coefficient (t-value)</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.5018 (5.326)</td>
<td>-</td>
<td>0.6703 (21.017)</td>
</tr>
<tr>
<td>Transit Investment: TTC Express Bus, 5-year Interval</td>
<td>0.0581 (2.282)</td>
<td>0.0469 (2.212)</td>
<td>0.0414 (1.997)</td>
</tr>
<tr>
<td>Household Size</td>
<td>0.1191 (7.818)</td>
<td>0.1305 (9.062)</td>
<td>0.1259 (10.142)</td>
</tr>
<tr>
<td>Centroidal distance to nearest TTC subway station</td>
<td>-0.0137 (-4.159)</td>
<td>-0.0117 (-4.222)</td>
<td>-0.0149 (-3.683)</td>
</tr>
<tr>
<td>Number of transit passes</td>
<td>0.1813 (4.916)</td>
<td>0.3738 (8.741)</td>
<td>-</td>
</tr>
<tr>
<td>Average Age</td>
<td>0.0026 (3.175)</td>
<td>0.005 (6.582)</td>
<td>-</td>
</tr>
<tr>
<td>Median Income</td>
<td>-0.0013 (-2.445)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Residential Area</td>
<td>-</td>
<td>-0.044 (-2.920)</td>
<td>-</td>
</tr>
</tbody>
</table>

**TTS-year Effects**

<table>
<thead>
<tr>
<th></th>
<th>Pooled</th>
<th>Fixed Effect</th>
<th>Random Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>-</td>
<td>0.179</td>
<td>-</td>
</tr>
<tr>
<td>2016</td>
<td>-</td>
<td>0.096</td>
<td>-</td>
</tr>
</tbody>
</table>
correlation between independent variables are less than 0.5, and all variance inflation factors (VIF) are below 4.

Heteroskedasticity of the pooled model is examined with the Breusch-Pagan test. The null hypothesis is that the model is homoskedastic which is not rejected. Determining whether models are heteroskedastic is important because estimated standard errors would be incorrect. This does not affect the beta coefficients: they remain as efficient estimators. Standard errors are used to calculate statistics in the hypothesis test, including the Chow test, the Breusch-Pagan Lagrange multiplier test, and the Hausman test. These tests are be used to determine the preferred model.

To determine whether the pooled or random effect model is preferred, the Breusch-Pagan Lagrange multiplier hypothesis test is used. The null hypothesis is that the individual-specific error term is correlated with the independent variables. The null hypothesis is rejected, which suggests that there is the presence of random effects in the data and that the pooled model is the incorrect model specification.

The Chow test is a $F$-test that checks whether the data can be pooled. Like the Breusch-Pagan Lagrange test, the null hypothesis test is rejected. The pooled model is not the correct model for this data.

The fixed effect model includes the TTSyear effect (Table 4). The time effect is estimated rather than the individual-specific effect because transit investment should be different across different time points. The independent variables included in the model are similar to the pooled model except for the residential area variable. The adjusted R squared is high with a value of 0.722, and signs are again reasonable. Assumptions are checked and are met. The model is homoskedastic.

The interpretation of the coefficients of the fixed effect model is as follows. A one unit increase in the independent variable means the beta coefficient increase in the natural logarithm of the number of transit trips produced by the individual per TTS year.

The random model displays an R-squared of 0.623 (Table 4). The individual-specific and idiosyncratic error terms are normally distributed with a mean of 0 and a variance of 0.0020 and 0.0014, respectively. Signs are sensible, and the assumptions are satisfied. The intercept is relatively high compared to the beta coefficients. The Hausman test is used to compare the fixed
effect and random model to determine which model is preferred. The test yields a Chi-squared statistic of 2.2414 with 3 degrees of freedom, meaning the null hypothesis is not rejected. The random error is the preferred model to represent the panel data. The coefficients of the random effect model can be interpreted as the change in the natural logarithm of the number of transit trips when the independent variables change by one unit over between individuals and time.

**Table 5. Elasticity of Independent Variables of Random Effect Model**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transit Investment: TTC Express Bus, 5-year Interval</td>
<td>1.042</td>
</tr>
<tr>
<td>Household Size</td>
<td>1.134</td>
</tr>
<tr>
<td>Centroidal distance to nearest TTC subway station</td>
<td>0.985</td>
</tr>
</tbody>
</table>

Several findings are discussed. The significant determinants of the number of transit trips appear to be household size, centroidal distance to the nearest TTC subway station, and the TTC express bus 5-year-interval transit investment variables of the random effect model. The elasticity of log to linear regression models is the Euler’s number raised to the beta coefficients. For the random effect model, an increase of one kilometre in transit investment of express bus sees a 1.042 multiplicative increase in the number of transit trip of the individuals, as per Table 5. Out of the eight transit investment variables created only the TTC express bus, five year-interval is significant. Based on the low t-value, it seems to be a weak predictor relative to the other independent variables. Perhaps this reflects upon the incremental nature of transit investment in Toronto. Nonetheless, it appears that transit investment has a positive effect on the number of transit trips.

### 5.5.2 Subject-Specific Empirical Models and Discussion

Table 5 presents the first of five subject-specific empirical models. The model is an ordinary least-squares (OLS) where the error term of the regression model is assumed to be a normal distribution and is independent and identically distributed for each household. Because of the discrete nature of the response variable, the number of household transit trips, the natural logarithm transformation is applied.

The OLS model achieves an adjusted R-squared value of 0.709 which means that the model explains 70.9% of the variance. Signs of the independent variables make sense. For example, as
the number of vehicles in the household increase, the number of transit trips decreases. The t-statistic of household transit fare expense is significantly higher than the other independent variables in the model. This particular variable is a great predictor because it always has a value

Table 6. Subject Specific OLS Model Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>(t-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Observations</td>
<td>589,349</td>
<td></td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.709</td>
<td></td>
</tr>
<tr>
<td>Adjusted R-Squared</td>
<td>0.709</td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>Coefficient</td>
<td>(t-value)</td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.027 (-2.581)</td>
<td></td>
</tr>
<tr>
<td>Number of transit stops within 1 km of household</td>
<td>0.002 (6.858)</td>
<td></td>
</tr>
<tr>
<td>Mixed land use entropy within 800m of household</td>
<td>0.552 (33.96)</td>
<td></td>
</tr>
<tr>
<td>Household transit fare expense</td>
<td>0.221 (250.20)</td>
<td></td>
</tr>
<tr>
<td>Number of vehicles</td>
<td>-0.069 (-11.68)</td>
<td></td>
</tr>
<tr>
<td>Number of workers</td>
<td>0.091 (28.76)</td>
<td></td>
</tr>
<tr>
<td>Dummy variable, year 2006</td>
<td>-0.076 (-9.19)</td>
<td></td>
</tr>
<tr>
<td>Dummy Variable, year 2011</td>
<td>-0.077 (-8.96)</td>
<td></td>
</tr>
<tr>
<td>Dummy Variable, year 2016</td>
<td>-0.095 (-12.33)</td>
<td></td>
</tr>
<tr>
<td>Interaction Variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of transit stops within 1 km of household : year2006</td>
<td>0.006 (13.23)</td>
<td></td>
</tr>
<tr>
<td>Number of transit stops within 1 km of household : year2011</td>
<td>0.006 (13.97)</td>
<td></td>
</tr>
<tr>
<td>Number of transit stops within 1 km of household : year2016</td>
<td>0.006 (15.11)</td>
<td></td>
</tr>
<tr>
<td>Mixed land use entropy within 800m : year2006</td>
<td>-0.084 (-3.94)</td>
<td></td>
</tr>
<tr>
<td>Mixed land use entropy within 800m : year2011</td>
<td>-0.137 (-6.30)</td>
<td></td>
</tr>
<tr>
<td>Mixed land use entropy within 800m : year2016</td>
<td>-0.168 (-8.37)</td>
<td></td>
</tr>
</tbody>
</table>
of zero for households that do not make transit trips. As the number of transit trips increase, so does transit fare expense.

In terms of the transit accessibility and TOD, transit accessibility has quite a small effect on the number of transit trips. For every increase in a transit stop within 1 km of the household, the number of transit trips increases by a factor of 0.2%. This trend continues in the interaction variables. The mixed land use entropy variable is more interesting as noted by the relatively larger t-statistic. If the area of each land use category in an 800m radius around a household is equal, the number of transit trips increases by a factor of 74%. However, this effect decreases over time as the interaction variables for each year increasingly becomes more negative.

The results of table 5 are interesting as it suggests that transit accessibility had little to no effect. To correctly handle the discrete nature of the response variable, count regression models are developed. These models are similar to OLS, but the error term is assumed to distribute as a poisson or negative binomial distribution. The assumption of IID and thus homogeneity is still assumed in these models. These models are developed using the Horowitz-Thompson likelihood.

Table 6 displays these poisson and negative binomial regression models. As these models are developed using likelihood, the rho squared value is used as a pseudo measure for goodness of fit. The response variable is the raw number of household transit trips. Both the poisson regression and negative binomial exhibit correct signs. The theta parameter is an additional parameter of the negative binomial distribution that is estimated. The interpretation is similar to the OLS models where

To decide between which distribution, the dispersion parameter is used. The poisson distribution assumes a dispersion parameter is one; the variance of the distribution is equal to the mean. It is found that the data is over-dispersed (dispersion parameter of 1.14) and so the negative binomial regression is preferred.Interestingly, the effect of the log of the transit accessibility variable does increase over time. The effect of mixed land use decreases over time. For an increase of 1 unit of the main effect of transit accessibility is an increase of 42% of the count of household transit trips.
### Table 7. Subject Specific Poisson and Negative Binomial Regression

<table>
<thead>
<tr>
<th>Variable</th>
<th>Poisson</th>
<th>Negative Binomial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Observations</td>
<td>589,349</td>
<td>589,349</td>
</tr>
<tr>
<td>Rho-squared</td>
<td>0.325</td>
<td>0.299</td>
</tr>
<tr>
<td>AIC</td>
<td>1,177,188</td>
<td>1,014,578</td>
</tr>
<tr>
<td>Variable</td>
<td>Coefficient (t-value)</td>
<td>Coefficient (standard error)</td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.654 (-25.02)</td>
<td>-2.736 (0.018)</td>
</tr>
<tr>
<td>log (Number of transit stops within 1 km of household + 1)</td>
<td>0.308 (12.86)</td>
<td>0.352 (0.007)</td>
</tr>
<tr>
<td>Mixed land use entropy within 800m of household</td>
<td>0.901 (22.23)</td>
<td>0.783 (0.021)</td>
</tr>
<tr>
<td>Household transit fare expense</td>
<td>0.083 (58.21)</td>
<td>0.210 (0.001)</td>
</tr>
<tr>
<td>Number of students</td>
<td>0.081 (6.17)</td>
<td>0.114 (0.002)</td>
</tr>
<tr>
<td>Number of workers</td>
<td>0.122 (9.84)</td>
<td>0.081 (0.002)</td>
</tr>
<tr>
<td>Average age of household members</td>
<td>-0.010 (-32.15)</td>
<td>-</td>
</tr>
<tr>
<td>Dummy variable, year 2006</td>
<td>-0.396 (-5.65)</td>
<td>-0.613 (0.028)</td>
</tr>
<tr>
<td>Dummy Variable, year 2011</td>
<td>-0.427 (-5.94)</td>
<td>-0.789 (0.031)</td>
</tr>
<tr>
<td>Dummy Variable, year 2016</td>
<td>-0.499 (-5.65)</td>
<td>-1.008 (0.027)</td>
</tr>
<tr>
<td>Interaction Variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log (Number of transit stops within 1 km of household + 1): year2006</td>
<td>0.121 (4.68)</td>
<td>0.172 (0.010)</td>
</tr>
<tr>
<td>log (Number of transit stops within 1 km of household + 1): year2011</td>
<td>0.168 (5.95)</td>
<td>0.227 (0.0112)</td>
</tr>
<tr>
<td>log (Number of transit stops within 1 km of household + 1): year2016</td>
<td>0.180 (4.79)</td>
<td>0.303 (0.011)</td>
</tr>
<tr>
<td>Mixed land use entropy within 800m : year2011</td>
<td>-0.283 (-4.33)</td>
<td>-0.142 (0.037)</td>
</tr>
<tr>
<td>Mixed land use entropy within 800m : year2016</td>
<td>-0.279 (-3.75)</td>
<td>-0.196 (0.033)</td>
</tr>
<tr>
<td>Theta parameter</td>
<td>-</td>
<td>1.43 (0.011)</td>
</tr>
</tbody>
</table>

Appendix A shows results of the zero-inflated count regression models. These models differ from the count regression models in table 6 as it combines two types of count processes: binomial logit and the poisson / negative binomial regression. The binomial logit model determines the binary outcome of whether the response variable is zero or a positive non-zero value. The poisson / negative binomial regression component then models the non-zero counts.
The purpose of modelling these two processes is to attempt to account for ‘excess’ zeroes. It is believed that there are two kinds of zeroes: true zeros and excess zeroes. True zeroes are where samples can only have zero count as their response. Excess zeroes are where the response is truly zero. For example, zero transit trips made by households because perhaps the household is located where no transit is available (true zeroes) or household members usually use another mode even if transit is available (excess zeroes).

‘Eta’ and ‘Logitp’ refers to the count regression and binomial logit component respectively. The signs of the binomial logit portion at a first glance may not make sense. However, interpreting the main effect of transit accessibility suggests that the higher the log count of transit stops 1 km around a household, the less likely the count will be zero. These models results were not as good as the previous count regression models, especially the zero-inflated negative binomial regression as noted by the low rho squared value. Forward selection for this model often yielded insignificant independent variables. Therefore, the final count regression chosen is the negative binomial count regression model in table 6.

5.6 Chapter Summary

The chapter presents an investigation of the direct relationship between transit investment and the demands for transit usage in Toronto and the GTHA. The demand for transit is measured by total transit ridership per household. Transit investment is measured as the number of transit stops within 1 km of the household. Mixed land use as a proxy for TOD is measured as entropy of land use categories within 800 m of the household. Data from a repeated cross-sectional household travel survey are used. Household transit demand information is drawn from these data and is first treated as a pseudo-panel dataset, and a pseudo-panel log-normal regression with pooled, fixed, and random effects are used. The empirical model reveals a low elasticity of demands for transit service in the city of Toronto. Low variability in the transit measurement may explain the poor explanatory power.

The subject-specific negative binomial regression model performs best for the entire GTHA region. This model is estimated using the Horowitz-Thompson estimator to properly account for the expansion factor of each household. To determine the transit investment effect over time, the main effect is included with interaction terms that interact with year. The model suggests that the
effects of transit investment and mixed land use increases and decrease over time. Several other socioeconomic independent variables are included as well.

In terms of future work, relaxing the assumption of homogeneity would be potentially useful. This could be done by using estimation method like weighted least squares instead of OLS and difference covariance structures for the likelihood estimation.

An additional model to attempt is an ARIMA (Autoregressive Integrated Moving Average) model. This idea came about from the prior pseudo panel models. The fixed effect model handles in part the assumption of homogeneity and serial correlation between individuals by removing the between effects. The fixed effect models only model the within effects of each individual for all years.

The issue with using raw cross-sectional data (as is the TTS) is that there are no cohorts; none of the households sampled at one time point is sampled again at another time point. To control for serial correlation between households within the same year across years, a paper done by Lebo and Weber details the method of ‘double filtering’ (Lebo & Weber, 2014). To do so, an ARIMA model can be developed in order to get residual values which are taken from the response and independent variables. However, multiple time points are needed to develop an actual time series model.
Chapter 6

6 Conclusion and Future Work

6.1 Chapter Overview

Section 6.2 summarizes the conclusions of the analysis and the contributions of this research. The thesis ends in Section 6.3 with recommendations on future work to extend analysis of transit accessibility and TOD trends and their effects on transit demand.

6.2 Research Summary

The purpose of this thesis is to determine the effects of transit investments variables, transit accessibility, and TOD on transit demand. These transit investments are thought to be influenced by long term policies and so the aim is to determine the transit investments’ effect over time. This is achieved by utilizing the TTS cross sectional data from 2001 to 2016 to develop descriptive analysis and models.

A review of the literature is first conducted to understand prior work of transit investment longitudinal analysis. It is found that there are some papers that analyze the longitudinal trends of transit accessibility using count measures. A research contribution of this thesis is to assess longitudinal trends of transit accessibility using the logsum accessibility measure with the aid of location discrete choice models. In terms of research done of the effects of longitudinal transit accessibility and TOD, the result have been variable but mostly show a positive relationship between increased transit investment and transit demand.

To establish data context, various socioeconomic variables are plotted from 1996 to 2016. In general, findings suggest a growing population especially among the 20 to 29 age group. Workers in the GTHA are gradually increasingly living within the same municipalities as their workplace. Mode shares across the GTHA have been consistent with the majority being auto-oriented. However, there have been gains for transit and active modes in recent years. These findings imply a greater use of the GTHA transportation network and longer commute times during peak period.
Chapter 4 delves into the location choice models which help in developing transit accessibility measures sensitive to personal attributes. It is found that people living in more mixed land use areas are more sensitive to travel time. Surprisingly, students tend to be more sensitive to travel time and cost than other workers. The actual logsum accessibility by transit measure shows small changes over time which may be because of the gradual change in stop density.

Finally, regression models are generated for the entire GTHA population based on the expansion factors provided by the TTS data. The final model is a negative regression model where the response variable is the number of household transit trips. The model suggests that the effect of transit accessibility and mixed land use entropy does increase and decrease over time. As the number of transit stops and stations increase in a 1 km radius around a household, the number of transit trips increases by a factor 42%.

6.3 Future Research Recommendations

This analysis assumes that transit investments like increased number of transit stations and stops and mixed land use does correlate positively with increase transit demand. As future research, it would be interesting to determine other factors that influence transit demand such as the 2008 recession or introducing Uber. Perhaps the 2008 recession effect encouraged more mode switching to cheaper modes. Ridesharing or ridehailing services effect may show over time that certain transit services compete with their services. Literature so far has been inconclusive with some papers stating that the service is a complement (2) whereas others do not (3).

Additional future work besides improving the actual models would also include doing similar analysis for other regions. This is to determine if behaviour is the same as exhibited in the GTHA population.

A large obstacle is the data requirement, especially for long-term analysis as data needs to be collected over time. Analyzing past behaviour may be nigh impossible if data is not available during the time period of interest. It is recommended to collect data whenever possible in order to have a time series of datasets to perform longitudinal analysis. Another consideration is current trends. Data on ride hailing has only been collecting in the 2016 iteration of the TTS. More time is needed for current phenomena to development to understand the actual effects of these events over time.
References


Appendix

Appendix A: Subject Specific Zero Inflation Poisson and Negative Binomial Regression

<table>
<thead>
<tr>
<th>Variable</th>
<th>ZI Poisson Coefficient (standard error)</th>
<th>ZI Negative Binomial Coefficient (standard error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Observations</td>
<td>589,349</td>
<td>589,349</td>
</tr>
<tr>
<td>Rho-squared</td>
<td>0.251</td>
<td>0.068</td>
</tr>
<tr>
<td>AIC</td>
<td>1,099,958</td>
<td>1,307,210</td>
</tr>
<tr>
<td>ETA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.584 (0.012)</td>
<td>0.758 (0.009)</td>
</tr>
<tr>
<td>log (Number of transit stops within 1 km of household + 1)</td>
<td>0.018 (0.006)</td>
<td>0.005 (0.003)</td>
</tr>
<tr>
<td>Number of transit passes in household</td>
<td>0.253 (0.001)</td>
<td>-</td>
</tr>
<tr>
<td>Mixed land use entropy within 800m of household</td>
<td>0.230 (0.013)</td>
<td>0.1620.007</td>
</tr>
<tr>
<td>Dummy variable, year 2006</td>
<td>-0.134 (0.017)</td>
<td>-0.119 (0.017)</td>
</tr>
<tr>
<td>Dummy Variable, year 2011</td>
<td>-0.299 (0.015)</td>
<td>-0.115 (0.017)</td>
</tr>
<tr>
<td>Dummy Variable, year 2016</td>
<td>-0.244 (0.010)</td>
<td>-</td>
</tr>
<tr>
<td>Interaction Variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log (Number of transit stops within 1 km of household + 1): year2006</td>
<td>0.040 (0.008)</td>
<td>0.045 (0.006)</td>
</tr>
<tr>
<td>log (Number of transit stops within 1 km of household + 1): year2011</td>
<td>0.074 (0.007)</td>
<td>0.051 (0.006)</td>
</tr>
<tr>
<td>Mixed land use entropy within 800m : year2006</td>
<td>-0.077 (0.017)</td>
<td>-</td>
</tr>
<tr>
<td>Mixed land use entropy within 800m : year2011</td>
<td>-0.059 (0.017)</td>
<td>-</td>
</tr>
<tr>
<td>Mixed land use entropy within 800m : year2016</td>
<td>-0.045 (0.018)</td>
<td>-</td>
</tr>
<tr>
<td>LOGITP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>2.576 (0.040)</td>
<td>2.207 (0.040)</td>
</tr>
<tr>
<td>log (Number of transit stops within 1 km of household + 1)</td>
<td>-0.399 (0.015)</td>
<td>-0.466 (0.017)</td>
</tr>
<tr>
<td>Number of transit passes in household</td>
<td>-2.156 (0.020)</td>
<td>-0.743 (0.057)</td>
</tr>
<tr>
<td>Mixed land use entropy within 800m of household</td>
<td>-0.690 (0.044)</td>
<td>-0.743 (0.057)</td>
</tr>
<tr>
<td>Dummy variable, year 2006</td>
<td>0.375 (0.085)</td>
<td>0.306 (0.074)</td>
</tr>
<tr>
<td>Term</td>
<td>Coefficient</td>
<td>Standard Error</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>-------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Dummy variable, year 2011</td>
<td></td>
<td>0.204 (0.075)</td>
</tr>
<tr>
<td>Dummy variable, year 2016</td>
<td>1.084</td>
<td>0.126</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.360 (0.066)</td>
</tr>
<tr>
<td>Interaction Variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \log(\text{Number of transit stops within 1 km of household + 1}) : \text{year2006} )</td>
<td>-0.136</td>
<td>0.031</td>
</tr>
<tr>
<td>( \log(\text{Number of transit stops within 1 km of household + 1}) : \text{year2011} )</td>
<td>-0.112</td>
<td>0.029</td>
</tr>
<tr>
<td>( \log(\text{Number of transit stops within 1 km of household + 1}) : \text{year2016} )</td>
<td>-0.223</td>
<td>0.044</td>
</tr>
<tr>
<td>Mixed land use entropy within 800m : \text{year2006}</td>
<td></td>
<td>0.272 (0.081)</td>
</tr>
<tr>
<td>Mixed land use entropy within 800m : \text{year2011}</td>
<td></td>
<td>0.399 (0.079)</td>
</tr>
<tr>
<td>Mixed land use entropy within 800m : \text{year2016}</td>
<td></td>
<td>0.438 (0.071)</td>
</tr>
<tr>
<td>THETA parameter</td>
<td></td>
<td>27.363 (2.95)</td>
</tr>
</tbody>
</table>