UTTRI

Research Report

Canadian Ridership Trends Research Project

Final Report

Eric J. Miller, Amer Shalaby, Ehab Diab, Dena Kasraian

June 2018

UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING
Transportation Research Institute

Canadian Ridership Trends Research Project

Final Report

Submitted to

Canadian Urban Transit Association (CUTA)

Prepared by:

Eric J. Miller, Ph.D.

Amer Shalaby, Ph.D., P.Eng.

Ehab Diab, Ph.D.

Dena Kasraian, Ph.D.

7 June 2018

Table of Contents

List of Appendices

Part I

Report Introduction

1. PROJECT OVERVIEW

Recently, there have been growing concerns about the negative impacts of rising automobile use and road congestion on personal mobility, safety, air quality and climate change. To address these issues, special attention has been given to improving and expanding transit services in order to attract new riders in pursuit of a number of environmental and societal goals. This special emphasis on public transit has been reflected in many cases by additional capital investments in public transit systems across Canada. Despite these efforts, transit ridership across Canada has been slowing down and declining over the past few years. For example, ridership statistics in 2015 showed levelling-off and declining trends in ridership in many Canadian cities including Halifax, Montreal, Ottawa, Toronto, Saskatoon, Calgary and Vancouver (Curry, 2017). Similar trends have been observed in many cities across the US. This highlights the need of a better understanding of the underlying factors affecting transit usage in each region.

Understanding, measuring, and modelling the factors affecting transit ridership has traditionally been done by analyzing transit level of service factors in addition to socioeconomic and land use factors. Emerging evidence from New York and other regions[1], however, suggests a strong impact of emerging new technologies and transport alternatives (e.g., Uber-like services and bike-sharing systems) on the use of public transit. Therefore, it is now more important than ever to provide a comprehensive and clear understanding of the reasons behind changes in ridership, while acknowledging the likely contribution of different emerging factors. The "Canadian Ridership Trends Research" project intends to perform this task and offers policy-relevant recommendations for improving transit service ridership across Canada.

This document presents the final report for the "Canadian Ridership Trends Research" project. The project's overarching objective is "*to conduct an in-depth study on current and future conventional ridership trends through research and consultation with transit systems*" which is *"to provide an understanding of the correlation between causal factors and ridership in Canada and provide explanation(s) of ridership decline at a transit system, Census Service Area (CSA), and national level."* This report brings together the deliverables for the Canadian Ridership Trends Research project organized into five parts.

Part II provides a comprehensive literature review of recent ridership prediction models and the factors affecting ridership at the aggregate level. The reviewed literature consists of both the academic literature (28 papers) and reports from transport authorities and research centres (14 documents) since 2000. The reviewed literature covers different levels of spatial aggregation (i.e., within-city or city level and multi-city studies) and temporal scope (including cross-sectional and longitudinal studies). This part provides an overview of the applied methods of ridership prediction. Furthermore, it synthesizes the findings of the literature on the significance

[1] Fitzsimmons, E. (2017), *Subway Ridership Declines in New York. Is Uber to Blame?* https://www.nytimes.com/2017/02/23/nyregion/new-york-city-subway-ridership.html

of various transit ridership factors including the built environment, socioeconomic, transit service and other external and contextual variables.

Part III presents the results of a survey of transit agencies regarding their ridership prediction practices. This survey was carried out in February and March of 2018 to understand the current state of ridership prediction practice and the key factors affecting ridership. Thirty six CUTA (Canadian Urban Transit Association) member agencies completed the web survey which amounted to a 35% response rate. This part summarizes the acquired information on ridership prediction methodologies used in the industry, ridership data sources and quality, the level of satisfaction with data and methods (both qualitative and quantitative), and suggestions for improvements.

Part IV outlines the preparatory steps taken to describe and analyze the nationwide trends in ridership, its influential factors, and their relation over time. It starts with an overview of the various collected longitudinal datasets of ridership as well as the influential internal and external factors. Then, it demonstrates the trends in annual transit ridership, measured in millions of linked trips, from 1991 to 2016 at various levels. Finally, it presents the findings of a number of exploratory analyses in terms of the relationship between ridership and influential indicators nationwide.

Part V provides an empirical investigation of variables explaining variations in transit ridership among transit systems and over time, using data from CUTA member agencies and a comprehensive set of indicators related to a) built environment attributes, b) socioeconomic factors, c) transit service factors and d) other external/contextual factors. The goal is to improve our understanding of the association of these various factors with past ridership trends, which should consequently improve our ability to forecast future trends. This part starts with the methodology used in the analysis, provides summary statistics of a number of key factors related to ridership trends and it investigates several case studies. It presents the results and interpretations of the implemented models before ending with some concluding remarks and policy implications.

Part VI presents a description of three policy analytical tools that have been developed in MS-Excel to estimate and predict ridership as a function of various influencing factors. The tools are based on the models of the study. The main objective is to provide an understanding of the relative impact of each factor on ridership at the transit agency level as well as the national level, while keeping all other variables constant at their mean values.

Part II

Literature Review

Table of Contents

List of Tables

1. INTRODUCTION

The main purpose of this chapter is to provide a comprehensive literature review of recent ridership prediction models and the factors related to ridership at the aggregate level. The literature review covers various bodies of literature including academic journals, conference proceedings, government reports, and publications by research centres.

The chapter begins with a discussion of the applied literature review methodology. Next, it presents the recent empirical research regarding ridership prediction, with special emphasis placed on understanding the used models and the variables included in these models. Subsequently, the chapter focuses on providing an overview of external factors –including land use, fuel price, socioeconomic factors, and access to transit– as well as internal factors such as transit fare pricing, transit service coverage, average headway, transit service intensity, and transit orientation pattern. Finally, it provides concluding remarks on the reviewed literature.

2. LITERATURE REVIEW METHODOLOGY

This section describes the literature review's methodology and sources. It first describes the review of academic literature on ridership estimation models and factors affecting ridership at the aggregate level and then it provides a brief overview of the reports by CUTA, APTA, TRB (Transit Cooperative Research Program (TCRP)), and publications of various transportation research centres.

2.1 Review of the academic literature

A systematic literature review method was utilized to identify and analyze all relevant publications in the academic literature on the factors affecting transit ridership and the methods used in ridership prediction modelling. Since the scope of this project is to understand ridership trends at the transit system, CSA, and national levels, we focused on the studies conducted at the aggregate level.

A systematic search strategy consisting of two phases was conducted. The first phase included a search of the Web of Knowledge, Scopus and TRID online article databases as of February 2018. TRID is a comprehensive database that includes more than one million records of transportation research worldwide (TRID, 2013). The search consisted of the following terms within the ''title'' search field: "(Transit OR Bus) AND Ridership OR Demand", OR "(Transit OR Bus) AND Ridership AND (Predict OR Forecast OR Model)". Only results yielding full articles and papers conducted within the past 20 years were retained for closer examination. Additionally, the search was also restricted to include only publications in English or French related to transportation, urban studies, social sciences and engineering. Finally, only the studies based on empirical model-driven analyses were included. The second phase of the search strategy began once the database search had identified the relevant articles based on a predetermined set of inclusion and exclusion criteria (Table 2-1). The reference lists of all articles were examined to identify new articles. Finally, the identified articles from both phases were fully reviewed, focusing on the factors related to ridership changes and modelling approaches.

The first phase of the search yielded 326 papers in total, of which 316 were excluded due to irrelevance and application of exclusion criteria. The second phase of the search strategy began once the database search results had been reduced to 23 relevant articles based on the predetermined set of exclusion criteria. Then, the reference lists of all articles were examined which yielded an additional 5 articles. Finally, articles that passed this process were reviewed and synthesized (see Appendix A). It should be noted that due to the low number of published studies that analyze transit systems of multiple regions, a few conference proceedings were included in the final database. The studies selected for the review focused on ridership modelling at the aggregate level of stop/station, neighbourhood or system.

Appendix A presents each study's sample size, investigated factors, significant factors, modelling approach and key findings. The appendix is also broken down into two sections. The first section, "City level and multi-city studies" studies, includes those which used a city or transport system as their unit of analysis or have been done across cities. The second section, "Within-city'" studies, includes those that use aggregate data at the stop, neighbourhood or route levels.

Table 2-1: Inclusion and exclusion criteria for the literature review

Inclusion criteria	Exclusion criteria
• Full published articles • Peer reviewed • English and French records • Using empirical model-driven analyses • Focusing on fixed-route transit service • Investigating the factors affecting transit usage and including models that incorporate two or more explanatory factors • Aggregated level studies • Published as of February 2018 • Published over the past 20 years	• Abstracts and short articles • Not peer reviewed • All languages other than English and French • Only a summary statistics study • Focusing on flex transit systems, on-demand services, specialized bus services (school service) • Investigating disaggregated users' travel behaviour and mode choice • Focusing on trip distribution and assignment methods, simulation techniques and mathematical optimizations methods • Focusing on testing new modelling techniques, visualizations approaches (e.g., heat maps) and indices (e.g., walkability index, urban form index)

2.2 Transport authorities and research centre reports

Reports by transport authorities include important information on the agencies' guidelines, policies and approaches, and they are often used to communicate these aspects to the public. This chapter explores and reviews the research findings of 14 recent reports publicised by CUTA, American Public Transportation Association (APTA), Transit Cooperative Research Program (TCRP), and other transportation research centres (e.g., Pew Research Center, Victoria Transport Policy Institute). These reports were identified by either the project's Steering Committee or through a brief research on the TRID portal. Some of these reports focus on the impact of emerging new technologies and transport alternatives (e.g., Uber-like services) on the use of public transit. These reports were included due to the scarcity of peer-reviewed literature on emerging transport alternatives. Appendix B presents each report's sample size, investigated factors, significant factors, modelling approach (if applicable) and key findings and recommendations.

3. REVIEW OF THE LITERATURE

As mentioned in the previous section, our review of the academic literature focuses only on empirical model-driven studies, excluding studies of other types. A considerable number of studies have used descriptive analyses and summary statistics to understand ridership changes and the factors underlying such changes (Brown & Thompson, 2008; Shalaby, Woo, & Currie, 2010; Thompson & Brown, 2012; Thompson, Brown, Sharma, & Scheib, 2006). However, these studies are usually criticized for the weakness of their descriptive approaches which can be highly subjective and inconclusive due to the lacking measures of statistical significance of individual factors. Other studies used data from customer stratification and travel behavioural surveys, and they estimated econometric models to gain a better understanding of the attitudes, behaviour and perception of travellers, and the factors that could increase their willingness to use public transit (Abdel-Aty, 2001; Chava, Newman, & Tiwari, 2018; Rojo, Gonzalo-Orden, dell'Olio, & Ibeas, 2012). The main advantage of these studies is their ability to provide empirical evidence on the decision making process of travellers and their propensity to switch modes or discontinue some trips. However, these studies have normally examined disaggregate travel behaviour at a single time step, as opposed to changes over time. Longitudinal analysis of disaggregate travel behaviour has been rare because temporal data at the disaggregate level are very hard and expensive to obtain. As a result, some researchers argue that aggregate analyses provide a feasible and adequate alternative to understand the determinants of transit ridership (Arana, Cabezudo, & Peñalba, 2014; Miller & Savage, 2017; Taylor, Miller, Iseki, & Fink, 2009), which is the focus of this literature review. More specifically, we focus on empirical model-driven aggregate level studies that explore the determinants of transit ridership.

This section summarizes the general findings in three bodies of literature, including research papers in academic journals and reports published by transit authorities and research centres.

3.1 Academic Literature

A. Within-city studies vs. city level and multi-city studies

Within-city studies

A sizable part of the literature has focused on investigating the factors affecting ridership at the stop/station, route, and neighbourhood levels within a given city. Some of those studies attempted to estimate changes in ridership as a function of various factors (Brakewood, Macfarlane, & Watkins, 2015; Campbell & Brakewood, 2017; Cervero, Murakami, & Miller, 2010; Chakour & Eluru, 2016; Gutiérrez, Cardozo, & García-Palomares, 2011; Miller & Savage, 2017; Wang & Woo, 2017). Usually, the scale and scope of each study was based on specific and limited objectives. For example, several studies focused on understanding the impacts of real-time information systems or bike-sharing systems on transit service ridership within a specific location or city (Brakewood et al., 2015; Campbell & Brakewood, 2017; Tang & Thakuriah, 2012). Similarly, other studies were concerned with assessing the effects of weather conditions at the route and system levels in a specific city (Arana et al., 2014; Singhal, Kamga, & Yazici, 2014; Tao, Corcoran, Rowe, & Hickman, 2018).

Bernal, Welch, and Sriraj (2016) examined the impacts of slow zones (due to track conditions or construction projects) on rail service ridership along one route in Chicago, namely the “El” Blue Line. Other studies focused on understanding the impact of the built environment and residential socioeconomic factors on station and/or route ridership (Chakour & Eluru, 2016; Jun, Choi, Jeong, Kwon, & Kim, 2015; Wang & Woo, 2017). Chakour and Eluru (2016) investigated the influence of stop level infrastructure and the built environment on bus ridership at the stop level in Montreal. Gutiérrez et al. (2011) also looked at transit ridership at the station level, using the Madrid Metro network as a case study. Normally, the *within-city* studies are considered more accurate and more effective because they are designed for a particular city in order to answer a particular research question. However, such studies have a few limitations which are mainly related to their narrow scope and the focus on a single study area. As a result, the recommendations and policy implications of these studies cannot be directly transferred and generalized to other cities or regions in a straightforward manner.

City level and multi-city studies

In contrast to within-city studies, fewer studies can be found in the literature that estimated changes in ridership across different cities and regions or used a city-wide transit system as the unit of analysis. These *city level or multi-city* studies are less common for several reasons including data limitations and modelling complexity. However, these studies can help with the identification of major trends and the generalization of results, overcoming the problem of external validity related to the use of limited-scale case studies. Therefore, the findings from these studies are considered applicable to other study areas. In our systematic literature review, we were able to identify 11 studies that used data from more than one city and transit agency or used citywide data as the unit of analysis (Boisjoly et al., 2018; Currie & Delbosc, 2011; Durning & Townsend, 2015; Guerra & Cervero, 2011; Kain & Liu, 1999; Kuby, Barranda, & Upchurch, 2004; Lane, 2010; Lee & Lee, 2013; Taylor et al., 2009; Thompson & Brown, 2006). Other studies that used more than one city or transit agency exist in the literature; however, they were solely based on descriptive analyses rather than comprehensive model-driven analyses to understand the correlation between transit ridership and other factors (Brown & Thompson, 2008; Thompson & Brown, 2012; Thompson et al., 2006).

Four out of the 11 studies used data for specific routes or stations from different cities in order to understand the common determinants of transit ridership (Currie & Delbosc, 2011; Durning & Townsend, 2015; Guerra & Cervero, 2011; Kuby et al., 2004). More specifically, Currie and Delbosc (2011) investigated which aspects of Bus Rapid Transit (BRT) design have a higher impact on transit ridership, by analyzing 77 BRT and non-BRT bus routes in four Australian cities (Melbourne, Brisbane, Adelaide, Sydney). Durning and Townsend (2015) presented a direct ridership model for Canadian rail systems. In their study, they analyzed ridership at 342 rail stations in Canada's five largest cities. Kuby et al. (2004) assessed the factors influencing light-rail station boardings at 268 stations in 9 US cities. Guerra and Cervero (2011) combined investment and station-level data from 50 fixed-guideway transit projects on 23 transit systems in the US to investigate the influence of job and population densities on transit ridership.

One study by Chiang, Russell, and Urban (2011) focused on forecasting ridership for the Metropolitan Tulsa Transit Authority in the US. Six studies focused on understanding the factors effecting the differences in ridership across different cities and regions at the system or metropolitan level (Boisjoly et al., 2018; Kain & Liu, 1999; Lane, 2010; Taylor et al., 2009; Thompson & Brown, 2006). Five of these studies focused on the transit agency level or city (or urbanized area-UZA) level, while only one study by Thompson and Brown (2006) considered the metropolitan level. More specifically, Boisjoly et al. (2018) explored the determinants of transit ridership at the transit agency level. Taylor et al. (2009) utilized the urbanized areas (UZA), arguing that the transit agency level can be problematic for several reasons. First, people live, work, and travel in UZAs instead of transit operator service areas. Second, it is likely that large UZAs are served by more than one transit agency with overlapping boundaries that are hard to define. Third, in the US, federal subsidies are calculated based on UZAs instead of transit service areas. Nevertheless, Thompson and Brown (2006) utilized the metropolitan scale, while indicating that the superiority of one scale over another cannot be accurately determined. Therefore, the appropriate unit of analysis depends on the nature of the study, the availability of suitable data, and the authors' perspectives.

B. Research Methodology

Within-city studies

The majority of the studies conducted at the stop/station (see Appendix A), route and neighbourhood levels used multivariate regression analysis to distinguish the degree of influence of different internal and external variables (Arana et al., 2014; Bernal et al., 2016; Campbell & Brakewood, 2017; Cervero et al., 2010; Gutiérrez et al., 2011; Jun et al., 2015; Lee, Eom, You, Min, & Yang, 2015; Sung & Oh, 2011; Wang & Woo, 2017). For instance, Singhal et al. (2014) and Arana et al. (2014) utilized multivariate regression models to investigate the impact of weather conditions on transit service ridership at the station level. Cervero et al. (2010) investigated the impacts of bus stop attributes and surrounding area on the average number of daily boardings in Los Angeles County, California.

Besides the multivariate regression models, some researchers used time-series analysis such as Auto Regressive Integrated Moving-Average (ARIMA) and seasonal Auto Regressive Integrated Moving-Average (SARIMA) methods (Chen, Varley, & Chen, 2011; Tao et al., 2018). For example, Tao et al. (2018) computed several time-series regression (i.e., ARIMA and SARIMA) models to capture the concurrent and lagged effects of weather conditions on bus ridership (derived from a three month smart card data) in Brisbane, Australia. The authors did not compare their model results to those of multivariate regression models, but they indicated that time-series modelling better accounts for self-dependency and temporal autocorrelation.

Other researchers used fixed-effects and mixed-effects linear regression models to account for the systematic differences between stations, routes or days (Brakewood et al., 2015; Miller & Savage, 2017; Tang & Thakuriah, 2012). For example, Miller and Savage (2017) estimated fixed effects regression models to understand the impacts of four fare increases implemented in 2004, 2006, 2009 and 2013 on ridership at rail stations in Chicago. Brakewood et al. (2015) used a fixed effects model to assess the effect of real-time information on public transit ridership in New York City. They analyzed bus ridership data collected over a three-year period, while controlling for changes in transit services, fares, local socioeconomic conditions, weather, and other factors. Tang and Thakuriah (2012) analyzed longitudinal data on route-level monthly ridership in Chicago from January 2002 through December 2010 to evaluate the likely effects of real-time information on public transit ridership using a linear mixed-effects model.

City level and multi-city studies

Five studies used multivariate regression models to distinguish the degree of influence of various internal and external variables on transit ridership (Chiang et al., 2011; Currie & Delbosc, 2011; Durning & Townsend, 2015; Kuby et al., 2004; Lane, 2010; Thompson & Brown, 2006). To give a few examples, Durning and Townsend (2015) estimated a regression model of the rail station boardings in five Canadian cities (Montreal, Toronto, Calgary, Edmonton, and Vancouver) in

2012 as a function of 44 explanatory variables of various socioeconomic, built environment, and system-related factors. Lane (2010) used several regression models to analyze the relationship between gasoline prices and transit ridership from January 2002 to April 2008 in nine major US cities. The models used monthly ridership as the dependent variable to assess the degree to which variability in transit ridership is attributable to gasoline costs, while controlling for service changes, seasonality, and inherent trending.

Six studies used other model types or a combination of models. Boisjoly et al. (2018) estimated longitudinal multilevel mixed-effect regression to explore the determinants of transit ridership from 2002 to 2015 for 25 transit authorities in Canada and the United States. Taylor et al. (2009) conducted a cross-sectional study of 265 urban areas in the US to explain the transit ridership in 2000. In their study, they used two-stage simultaneous equation regression models to account for the relationship between transit supply and demand. Lee and Lee (2013) also used two-stage least squares regression analysis to examine the impacts of gas prices on transit ridership in 67 urbanized areas in the US. In their study, they used longitudinal data collected between 2002 and 2010. They argued that their findings should be more generalizable than previous cross-sectional or time-series studies found in the literature (e.g., Taylor et al. (2009) study).

Kain and Liu (1999) used cross-sectional regression and time series ridership models to assess the impacts of different factors on ridership in Houston and San Diego in the US between 1980 and 1990. Chiang et al. (2011) analyzed the monthly ridership of the small case study of Tulsa Transit Authority (i.e., 20 routes) to identify the relevant factors that influence transit use. They used a combination of regression analysis, neural networks, and ARIMA models. They indicated that a simple combination of these forecasting methods yields greater forecast accuracy than the individual models separately. However, this methodology is yet to be tested on larger case studies and multiple cities. Similarly, Guerra and Cervero (2011) used a combination of regression analysis with fixed effect and random effects estimators and two stage least square model to investigate the influence of job and population densities on transit ridership between 2000 and 2008.

A common problem indicated by the authors of the above *city level and multi-city* studies is the high level of correlation among predictor variables within each city, which limited their investigations (Cervero et al., 2010; Currie & Delbosc, 2011; Lane, 2010; Taylor et al., 2009; Thompson & Brown, 2006). The multicollinearity problem, or the high degree of correlation among explanatory independent variables, usually occurs among various spatial variables, transit service variables, and between spatial and socioeconomic variables. For example, Currie and Delbosc (2011) had to estimate several models with and without certain variables to deal with multicollinearity, and indicated that certain service level variables had a dominating influence on the model results. Other researchers estimated correlation matrices of variables and used experience to eliminate certain variables from the models, which limited their ability in identifying the relative impact of each variable on total ridership volumes.

3.2 Reports of transport authorities and research centres

The reviewed reports can be categorised into three main groups. The first group generally summarizes the existing literature on ridership determinants, presents existing (local) ridership trends and provides recommendations for effective (local) transit policies. Some studies in this group focus on a specific aspect of transit demand (e.g. price elasticities and cross elasticities (Litman, 2017)) while others focus on the local context (City of Edmonton, 2016; York Region Transit, 2017).

The second group presents emerging ridership trends and issues using primary data sources (surveys and expert interviews). Results are typically presented in the form of graphs, tables and descriptive statistics. CUTA (2007) provides a profile of Canadian transit ridership for different classes of cities, regions, and the whole country. More recent reports on the socio-economic and travel behaviour characteristics of TNC users fall within this group as well (Clewlow & Mishra, 2017; Schaller, 2017; Smith, 2016; TCRP, 2016).

The third group uses empirical data and model-driven analyses to examine transit ridership trends. Three reports performed aggregate analyses at the national level (MSA and transit system levels) and used multi-variate regression and/or correlation analysis to measure the relative association of various internal and external factors with transit ridership (Alam, Nixon, & Zhang, 2015; Kohn, 2000; Taylor et al., 2002). Manville, Taylor, and Blumenberg (2018) focused on the determinants of transit ridership reduction in the Southern California region. A noteworthy work is that of Feigon and Murphy (2018) who used original TNC trip origins and destinations in five different US regions and performed exploratory analyses to test relationships among demographic variables, TNC use and transit availability at six different combinations of time of day and day of week, across the study regions.

4. TRANSIT RIDERSHIP FACTORS

The studies and reports discussed above have investigated a wide range of factors that may have an impact on ridership at the aggregate level. These factors can be classified into four categories: built environment characteristics, transit service attributes, socio-economic characteristics and other general factors. Table 4-1 presents the list of these factors. From another point of view, these factors can be broadly divided into two categories: external and internal. External factors are generally not within the control of transit agencies and their managers. These factors include the built environment, socio-economic and other general factors such as the availability of bike-sharing and ride-sharing (e.g., Uber, Lyft) services. The built environment and socio-economic factors (e.g., building density, household income) are used at the aggregate level as proxies for a large numbers of individual factors that are related to transit use. However, their aggregate nature imposes several challenges for modellers and researchers when dramatic variations in their values exist. Internal factors, or transit service factors, are normally within the control of transit agencies and authorities. Among others, they include service levels and transit fares (see Table 4-1). The following section examines the relative influence and possible interactions of the most common internal and external factors.

Table 4-1: Transit ridership factors

A. Built environment factors	B. Socioeconomic factors	C. Transit service factors	D. Other factors (external/contextual)
1. Population / Population density 2. Urban land area 3. Land use mix 4. Green space 5. Local opportunities: businesses 6. Local opportunities: recreation 7. Freeway network length and exits 8. Highway network length and exits 9. Street network length and	1. Age 2. Gender 3. Student population 4. Senior population 5. Workforce 6. Unemployment rate 7. Household size 8. Car ownership rates 9. Ownership of driver's license 10. Household composition (e.g. couples with (out) children, singles, etc.) 11. Household disposable income 12. Household's expenditure on transport 13. Average rent/shelter cost	1. Service frequency 2. Service reliability (e.g., headway adherence, on-time performance) 3. Service network coverage 4. Network design/type 5. Service span/hours 6. Vehicle revenue hours 7. Vehicle revenue miles 8. Fare 9. Fare/income	1. Weather (temperature, snow on ground and precipitation) 2. virtual connectivity (telecommuting, online shopping) 3. Air quality (Air Quality Index and Air Quality Health Index) 4. Price of car ownership (fuel/energy, insurance, maintenance) 5. Congestion (average level/cost)

A. Built environment factors	B. Socioeconomic factors	C. Transit service factors	D. Other factors (external/contextual)
number of intersections 10. Railway lines and stations 11. Private dwellings by type (e.g. single-detached, apartment) 12. Dwelling characteristics (e.g., period of construction, condition) 13. Dwelling tenure 14. Property value 15. Work location (in/outside census subdivision of residence) 16. Distance to downtown 17. Employment in downtown	14. Employment status and type (part-time/full time) 15. Employment sector (e.g. agriculture, utilities, construction) 16. Employment/population ratio 17. Education (highest certificate) 18. Immigration status (citizen or not) 19. Immigration period	10. Availability of integrated-fare payment systems 11. Composition of fleet and modes (bus, subway or LRT) 12. Density of dedicated bus lanes and transit preferential treatment 13. Availability of real-time information 14. Transit service accessibility 15. Average network load 16. Transit funding User satisfaction level	6. Active transportation support systems (availability and promotion, bike-sharing scheme) 7. Vehicles for hire/ride-sharing availability (Uber, Lyft, Taxis)

4.1 Built environment factors

- *Population, population density and employment density*

Two of the most important determinants of transit ridership used by the majority of studies at the aggregate level are population, population density and employment density (Boisjoly et al., 2018; Chakraborty & Mishra, 2013; Durning & Townsend, 2015; Jun et al., 2015; Kain & Liu, 1999; Kuby et al., 2004; Miller & Savage, 2017; Taylor et al., 2009). Among others, Kuby et al. (2004) and Jun et al. (2015) reported a strong positive impact for these two factors on transit service ridership. Some researchers have used housing density characteristics as a proxy for population density factors (Chakraborty & Mishra, 2013).

The impacts of population and employment density, however, differ across cities and regions. For instance, less transit usage is expected in low-density areas and decentralized employment locations due to other factors related to auto-centric travel behaviour and land-use patterns. Researchers usually indicate that population density factors are highly correlated with other factors, such as car ownership. More specifically, residents of suburban areas (or low-density areas) are likely to have higher access to cars, while residents of central city areas (high-density areas) are likely to have lower car access.

- *Land use type, mix and diversity*

Urban sprawl and suburbanisation are postulated to contribute to transit ridership reduction. Suburbanisation goes hand in hand with lower densities and higher automobile use which are the nemeses of transit usage. For example, Taylor et al. (2009) indicated a positive impact of the average area of urbanization on transit ridership. In addition, the geographical context of cities also play a role: areas constrained by natural phenomena like mountains or bodies of water are more likely to follow a compact development pattern which makes them suitable for transit services. On the other hand, cities on flat open land tracts are likely to consume more land and develop into areas less accessible to transit (Kain & Liu, 1999). Several researchers used dummy variables in their model to control for the fixed impact of locations on transit usage (Currie & Delbosc, 2011; Tao et al., 2018; Wang & Woo, 2017).

A number of researchers incorporated land use types in their models (e.g., housing, commercial, health care, recreational, governmental, institutional land uses). However, they did not find a significant or consistent (in terms of direction) impact of different land use types on transit usage (Chakraborty & Mishra, 2013; Sung & Oh, 2011). In contrast, land use mix and diversity usually are found to have a significant and positive impact on transit usage (Gutiérrez et al., 2011; Jun et al., 2015; Sung & Oh, 2011). However, Taylor and Fink (2013), among others, have cautioned that identifying the associations between transit usage and land use and density are usually complicated due to high levels of collinearity among spatial variables as well as between spatial and socioeconomic variables in some cities.

- *Street network and design*

Several researchers have shown positive impacts of intersection density factors on transit ridership. This is expected since more intersections mean shorter blocks and more walkable neighbourhoods (Sung & Oh, 2011; Taylor et al., 2009). On the other hand, some researchers found that density of highways, freeways and streets in urban areas have negative associations with transit usage. However, it should be noted that street network and design variables were not used by most researchers, which can be attributed to their correlation with other spatial factors.

4.2 Socioeconomic factors

- *Population and employment characteristics*

Demographic trends such as population and employment growth have been shown to significantly influence transit ridership (Kain & Liu, 1999). Population characteristics such as the share of college students, population in poverty, average income, proportion of recent immigrants and ethnic composition are acknowledged as the most significant external socioeconomic factors by (Chakraborty & Mishra, 2013; Taylor et al., 2009; Wang & Woo, 2017). For instance, Wang and Woo (2017) found that poverty rates and low income populations in suburban areas, compared to areas in downtown and inner-city, influenced positively transit ridership. Their findings also indicated that transit usage increased with increases in the proportion of renters and minorities.

Taylor et al. (2009) examined the effects of several population characteristics and reported the share of college students, recent immigrants, population in poverty to have strong associations with transit usage in 265 urban areas in the US. Other researchers did not find a significant association between income and transit usage in some areas, which was explained by the high level of income at these locations (Durning & Townsend, 2015).

Growth in total employment was one of the factors with the highest correlation with ridership growth during the 1990s in 227 transit agencies in the US (Taylor et al., 2002). Unemployment rate is shown to be negatively correlated with ridership (an exception is (Taylor et al., 2002)). Chiang et al. (2011) reported a negative association between of the number of individuals receiving food stamps (a proxy for unemployment) and transit usage in Tulsa, Oklahoma. Tang and Thakuriah (2012) using longitudinal data from Chicago indicated a mixed effect of unemployment on bus ridership. They showed that transit ridership reached its minimum when Chicago unemployment rate was 8.51% and grew when the unemployment rate increased or decreased, controlling for other factors. Miller and Savage (2017) discovered inconsistent impacts of age and gender on transit usage. They also found a higher decline in transit ridership in lower-income neighbourhoods compared to high-income neighbourhoods due to transit fare increases.

- *Car ownership*

Manville et al. (2018) examined a host of factors across time to determine the causes of transit ridership decline in the Southern California region. The results showed that the increase in motor vehicle access especially among low-income households with limited mobility was the main contributor to falling transit ridership. Currie and Delbosc (2011) demonstrated similar results in terms of the negative impact of car ownership on transit usage in four Australian cities. Boisjoly et al. (2018) explored the determinants of transit ridership from 2002 to 2015 for 25 transit authorities in Canada and found negative impacts of car ownership on transit ridership.

4.3 Transit service factors

While the built environment and socioeconomic factors substantially influence transit ridership, transit service factors also play an important role. Several studies draw contrasting conclusions on the relative explanatory strength of the built environment-related, socioeconomic and transit service factors. For example, Taylor et al. (2009) conducted a cross-sectional study of 265 urban areas in the US to explain transit ridership. They found that while fares and service frequency are significant, the majority of the variation can be explained by factors outside of the control of transit management, such as population density, personal income, regional location, highway system and demographic characteristics. In contrast, Currie and Delbosc (2011) analyzed 77 BRT and non-BRT bus routes in four Australian cities (Melbourne, Brisbane, Adelaide, Sydney) and found that service level dominates predictions of boardings per route. BRT infrastructure treatments (such as right of way) within the context of high service levels have a significant impact on ridership.

- *Transit supply*

Most of the reviewed studies demonstrated a strong and positive association between transit supply measured by indicators such as *revenue vehicle hours* and *revenue vehicle miles/kilometres* and transit ridership (Boisjoly et al., 2018; Currie & Delbosc, 2011; Kain & Liu, 1999; Lane, 2010; Taylor et al., 2009; Thompson & Brown, 2006). This is expected since transit ridership is a function of transit service supply to some extent. In addition, transit service supply is also largely a function of transit demand. The rationale is that in practice, transit agencies often adjust their service over time to match the change in ridership level. Therefore, service supply factor is highly correlated with city size and population. Larger cities (with larger populations) are expected to have higher level of service and transit ridership, while smaller cities (with smaller populations) struggle with providing the critical mass for viable transit systems and tend to have lower levels of transit services.

- *Transit fares*

Transit fare is shown to be negatively related with transit ridership (Boisjoly et al., 2018; Brakewood et al., 2015; Chen et al., 2011; Chiang et al., 2011; Kain & Liu, 1999; Miller & Savage, 2017; Tang & Thakuriah, 2012). For example, Kain and Liu (1999) attribute the large increases in transit use in Houston and San Diego in the 1980s to sizeable service increases and fare reductions enabled by large subsidies from federal, state and local governments. Interestingly, while most of the reviewed studies indicate a negative impact of fares on transit usage, some studies suggest that the demand for urban transit services could be inelastic. For instance, changes in average fares were found to be loosely correlated with ridership in the case of US agencies with increased ridership during the 1990s (Taylor et al., 2002). Similarly, Kohn (2000) showed that during the same period in Canada, increases in revenues and average fares were simultaneous. Thus, commuters continued to use urban transit services despite the rise in fares. The author noted that fare increases could be considered marginal in the face of costs of operating automobiles and downtown parking. Litman (2017) concluded that when the starting point of a fare increase is relatively high, transit elasticities seem to somewhat increase due to fare increase. Chen et al. (2011) showed that transit ridership exhibits an asymmetric behaviour in response to transit fare. Asymmetric behaviour implies that “the magnitude of change in behavior may be different depending on the direction of change” (Litman, 2005, p.404). Their finding supports the observation that the effect of fare increase on the reduction of ridership is greater than the effect of its decrease on the increase of ridership (Litman, 2017).

- *Network design/type*

Network design and type has a strong impact on transit ridership. A good example of that is the implementation of Houston’s “system Reimagining Plan”(NACTO, 2018). In summer 2015, Houston’s Metro Transit replaced the bus radial system with a grid system, while increasing the number of high-frequency routes and expanding the weekend service. The new grid network offered simpler and more direct routes to various destinations. Passengers were no longer required to go through Downtown. Several reports show a substantial increase in local bus ridership, particularly during weekends, after the implementation of the new bus network. Shalaby et al. (2010) explored factors that are driving ridership in Toronto and Melbourne. They revealed that while both cities have fairly comparable bus fleet size and rolling stock of rail vehicles, higher transit ridership in Toronto is likely the result of a combination of transit service and network characteristic as well as socio-economic factors.

4.4 Other external/contextual factors

- *Gas price*

Gas price is shown to be an influential factor on transit ridership; however, results concerning the extent of its influence are mixed. Gas price was found to be the only significant external factor on travel demand by bus in a national study across the US (Alam et al., 2015). Similarly, Boisjoly et al. (2018), Bernal et al. (2016) and Chiang et al. (2011) reported a strong impact of gas prices on transit usage. In contrast, Chen et al. (2011) showed that the effect of gas price is significant yet small in magnitude. The effect was also found to extend over a year and to be more tangible as the prices rise rather than fall. Lane (2010) also found a small but statistically significant amount of ridership fluctuation due to changes in gas prices. The author also demonstrated that there is a lagging response of ridership to changes in gasoline prices at the monthly level.

- *Weather conditions*

Weather conditions were shown to have an important impact on transit usage. Several studies found that wind, rain and snow could result in a decrease in transit usage, while temperature having a mixed effect on transit ridership (Arana et al., 2014; Singhal et al., 2014; Tao et al., 2018). Tao et al. (2018) showed that hourly bus ridership on weekends was more affected by changing weather conditions than weekdays. Their findings also indicated that weather impacts on bus ridership varied not only between weekdays and weekends, but also across trip destinations. Singhal et al. (2014) assessed the impact of weather conditions on ridership in New York and showed similar results in terms of weather condition effects on weekend travellers. They also discovered that ridership is most affected during the PM time period, followed by the midday period and it was least affected during the AM period. Weather conditions are also found to have a higher negative impact on ridership at street-level stations than underground stations.

- *Shared-use modes*

The impact of shared-use modes has been the focus of a number of recent studies. Shared-use modes are the result of a growing number of alternative transportation options enabled by technological and organizational developments. They include car-sharing (car2go, Zipcar), bike-sharing, and a variety of ride-hailing services (such as Uber and Lyft) provided by the so-called Transportation Network Companies (TNCs).

A few studies indicate that the pervasive use of ride-hailing services has occurred at the expense of transit use. According to the results of an online travel and residential survey by Clewlow and Mishra (2017) with a large representative sample of urban and suburban populations, the transit use of ride-hailing users has witnessed a net decline (6% reduction for Americans in major cities). This finding is corroborated by Schaller (2017) who analyzed the trends from 2012 to 2016 in New York and concluded that TNCs have diverted transit –especially bus– users.

There are only a handful of studies based on actual TNC trip data due to the unavailability of such data. A recent analysis of TNC trip origins and destinations in five different US regions indicates that most TNC trips are short and concentrated in downtown neighbourhoods (with the exception of airports). Temporally, these trips mostly take place during evening hours and weekends (Feigon & Murphy, 2018).

It is argued that the complementary or substitutive role of TNCs is highly dependent on the type of transit service. Two studies have identified the main victim so far as the bus service (Clewlow & Mishra, 2017; Schaller, 2017). For instance Clewlow and Mishra (2017) showed that ride-hailing has hit the bus service most severely (a 6% reduction) followed by light rail (a 3% reduction). However, it has contributed to commuter rail services (a 3% increase).

A number of studies have not found a significant relationship between TNCs and transit ridership (Boisjoly et al., 2018; Manville et al., 2018). Boisjoly et al. (2018) note that while factors such as the presence of ride-hailing services (Uber) and bicycle sharing are statistically insignificant in their model, they are associated with higher levels of transit ridership.

In contrast, a TCRP (2016) report claims that TNC trips appear more likely to replace automobile trips, thus complementing public transit. Their findings originate from a capacity and demand analysis of ride sourcing (from Uber Application Programming Interface (API) data) and public transit data, interviews with public agency officials and private mobility operators and a survey of 4500 shared-use mobility consumers in seven major US cities. Similarly, Smith (2016) shows that frequent ride-hailing users are less likely to own a vehicle and more likely to use a range of transportation options (56% regularly take public transportation). The findings of Feigon and Murphy (2018) indicate that the majority of TNC trips are concentrated in downtown neighbourhoods which are highly accessible by transit. However, there is an absence of a clear relationship between the level of peak-hour TNC use and the decrease or increase in transit ridership in the study regions in the long-term. This could be explained by the observation that TNC trips primarily occur during evening hours and weekends (Feigon & Murphy, 2018).

Regardless of their impact on transit ridership, there seems to be a unanimous agreement that TNCs' significance (in terms of users/VKT) will grow and that public entities should seize opportunities to further engage with them (TCRP, 2016). However, the recommended priorities and degrees of engagement differ among transit agencies in large, midsized and smaller urban areas (Feigon & Murphy, 2018).

4.5 Comparison between internal and external factors

The results on the significance and the extent of influence of transit ridership determinants are mixed. Many argue that transit ridership is influenced primarily by external factors (Taylor et al., 2009). In contrast, Alam et al. (2015) and Currie and Delbosc (2011) showed that internal factors were the most significant predictors of travel demand and concluded that transit ridership is controllable by managers and operators irrespective of external factors.

Among the external factors, access to private car is found to be an extremely influential factor, to the degree that it has been referred to as the "single largest factor affecting transit use" (City of Edmonton, 2016). Manville et al. (2018) among others contend that socioeconomic and location factors influence transit use mainly through influencing people's access to private cars.

However, the importance of external factors does not undermine the role of transit policy. Taylor et al. (2009) analyze numerous internal and external determinants of transit ridership at the national level. They conclude that while the overall level of transit use is highly dependent on the nature of the urbanized area –shaped by its metropolitan economy, regional geography, population characteristics and elements of automobile/highway systems– transit policies, reflected in service frequency and fare levels, can make a significant contribution.

A primary driver of transit ridership is the reduction of the price of transit. Here, price is a general term which includes both monetary costs as well as non-market costs such as travel time and discomfort, or in other words the *users' perceived marginal cost* (Litman, 2017). Litman (2017) summarizes previous research on price elasticities and cross-elasticities and concludes that transit fares, service quality and parking pricing are likely to be the most influential determinants of transit ridership.

5. LITERATURE REVIEW SUMMARY OF KEY FINDINGS

The review of literature and reports concerned with transit ridership reveals a diverse spectrum of methods and data sources resulting in mixed findings. Overall, the relationships between ridership and influential external and internal factors are very complex. These factors are numerous, interrelated, and their degree of influence may change from time to time and according to context. The diversity of findings is very likely due to the differences in the type, scale and number of transit systems studied, the applied methodology, indicators and the contextual conditions. Nevertheless, ridership trends and their determinant factors should be constantly monitored and assessed against the background of ongoing socio-economic, behavioural and technological trends, in order to disentangle their dynamic relationship. The following points represent the key findings in terms of the characteristics of the reviewed literature as well as the reported significant factors:

- Recent transit ridership trends: a number of recent studies and reports have shown a trend of stabilizing or declining transit ridership after a period of growth during the 1990s and 2000s in the US and Canada (Curry, 2017; Manville et al., 2018)
- Level of aggregation: the reviewed studies can be classified into two groups:
  - The majority of studies are *within-city* studies. They investigate ridership determinants at the stop/station, route and neighbourhood levels within a single city. These studies are more suitable for distinguishing context-specific trends and developing empirical models of transit use. However, their results are less likely to be generalizable in a straightforward manner.
  - *City level and multi-city* studies are smaller in number. These studies investigate a number of transit systems aggregated at the city level, across several cities or nationwide. They produce more robust results which can be applicable to other systems and metropolitan regions (Taylor & Fink, 2013). However, they are demanding in terms of data requirements and can be insensitive to local variations.
- Investigated variables: although many significant factors featured in various studies, several interrelated variables were repeatedly found to have significant association with transit usage:
  - Population and employment density (among the built environment variables)
  - Car ownership, income and unemployment rate (among the socioeconomic variables)
  - Transit fare and level of supply (among the transit service factors)
  - Gas price (among the other external factors)
- The role of shared-use modes: findings concerning the influence of shared-use modes on transit usage are inconclusive, mainly due to the scarcity of primary data. However, it seems that the increase in usage of shared-use modes has not occurred at the cost of at least rail-based transit use.

- The role of the local context and type of investigated transit system: the context of the study in terms of geographical location, transit policy history and horizon of the analysis (e.g. the 1990s vs the 2010s) is important. In addition, results can be dependent upon the type of the investigated transit system (e.g. bus or rail).
- Methodology: the majority of studies are cross-sectional and they use regression analysis to identify the factors having significant association with transit usage. A small number of studies applied more sophisticated models using smaller datasets and/or longitudinal data. There is a need for more comprehensive approaches to transit ridership data modelling which can distinguish between long- and short-term changes and account for the reverse causality between transit demand and supply and the asymmetrical response of transit ridership to changes in certain determinants. These issues are further elaborated in the next section.

Issues in need of attention and further investigation

Based on the literature review, we identify a number of issues which call for further investigation. First, there is an increasing need for longitudinal studies. Changes in transit demand as a result of external and internal factors are gradual. The majority of existing studies are cross-sectional and fail to capture the long-term effects of changes in internal and external variables. It is shown that the effects of determinants can vary over time. For instance, long-run elasticities are consistently found to be more significant than short-run ones (Chen et al., 2011). It is important to distinguish between short-run and long-run effects to be able to design effective transit policies accordingly.

Second, an important yet often overlooked issue is the simultaneity between transit supply and consumption, or the issue of endogeneity. Taylor et al. (2009) is one of the few studies that examined this simultaneous relationship. Transit supply is assumed to determine its ridership. However, a reverse causality exists where increased ridership affects supply. This is due to the fact that in practice transit operators adjust the level of service supply to match the changes in demand. This adjustment will in turn further influence transit use. Thus, a special attention should be given to this relationship, while estimating transit ridership models (Taylor & Fink, 2013).

The third issue has to do with multicollinearity. The use of numerous independent variables for each group of variables (built environment or socio-economic characteristics) can reduce the problem of omitted variables. However, various variables within a group correlate with each other (e.g. urbanized area and population). Furthermore, some built environment and socio-economic variables correlate with each other as well (e.g. building typology and income). As a result, several model specifications and hypotheses should be tested to decide the appropriate combination of indicators.

The fourth issue is concerned with the asymmetrical behaviour of transit ridership to increase and decrease in some determinants. For instance, it is shown that the rate of decline in ridership due to a fare increase by a certain amount is different from the rate of ridership growth due to a fare reduction by the same amount (Chen et al., 2011). The same pattern was observed with respect to gas prices. Investigation of such asymmetric behaviour is possible with longitudinal data which emphasizes the need for long-term analysis.

In addition, there is a need for more comparative studies to determine to what extent the varying results can be attributed to differences in the types of investigated transit systems, the applied methodology and the context of the study. To analyse national ridership trends, a system level approach which aggregates data at the transit system level appears to be the most robust, generalizable and feasible method. This method can benefit from the availability of data across various regions and over relatively long periods. In fact, this approach can help us understand the current trends in transit ridership across Canada. Several cities, including Halifax, Montreal, Ottawa, Toronto, Saskatoon, Calgary and Vancouver, are observing levelling-off and even declines in transit ridership (Curry, 2017). This has occurred despite the increase in capital and operating funds over the past few years. Therefore, a better understanding of the factors affecting transit usage and the drivers behind the observed trends in Canada is essential. This will be done in order to better inform future land use and transportation policies with an overarching goal of improving transit service ridership.

Part III

Survey of Canadian Ridership Prediction Practice

Table of Contents

List of Tables

List of Figures

1. INTRODUCTION

This chapter presents the results of a survey of transit agencies regarding their ridership prediction practices. The purpose of this survey is to obtain information on the current state of practice in fixed-route transit ridership prediction in Canada. It was designed to elicit information on ridership prediction methodologies used in the industry, level of satisfaction with these methods, and suggestions for improvements. It also aimed at developing a better understanding of the extent to which Canadian transit agencies use models to predict future ridership changes, and the ridership factors incorporated in these models. The survey included five main sections. The first section gathered information on the ridership prediction typology and general practice. The fourth section elicited information on the explanatory factors and associated data inputs considered for each ridership prediction method, and it also measured the agency's satisfaction with these methods. Finally, the last section of the survey inquired about the agency's requirements for robust ridership prediction modelling. The term "prediction" was used in the survey to refer to the estimation or forecasting of ridership at a future time period. The survey was forwarded to all CUTA members. English and French versions of the survey were made available for all participants. The rest of this chapter discusses the survey design, its respondents and findings per survey section. The full survey instrument can be found in Appendix C.

2. SURVEY RESULTS

2.1 Survey respondents

A survey of transit agencies in Canada was carried out in February and March 2018 to understand the current state of ridership prediction practice and the key factors affecting ridership. Using CUTA's membership list, the survey was sent by e-mail to all 103 transit agencies who are CUTA members in February 2018. This task was conducted by CUTA to ensure a higher response rate. Follow-up emails were sent approximately two weeks after the original survey distribution to encourage responses. Furthermore, some transit agencies were contacted directly by CUTA by e-mail to encourage completion of the survey. In total, 36 transit agencies in Canada completed the survey, with a response rate of 35%. Table 2-1 presents the list of responding agencies, while Figure 2-1 shows the distribution of the responding agencies in Canada.

As seen in Table 2-1, 9 of the 10 largest transit agencies in Canada completed the survey, including: Toronto Transit Commission (TTC), Société de transport de Montréal (STM), OC Transpo (Ottawa), Calgary Transit, Edmonton Transit, Mississauga Transit (Miway), TransLink (Vancouver), Winnipeg Transit and Réseau de transport de la Capitale (Québec city). Since not all questions applied to all respondents, the number of responses varies among questions. Therefore, the number of respondents per question will be highlighted as much as possible.

Table 2-1: List of responding agencies

Count	Transit agency	Within the 10 largest agencies	Count	Transit agency	Within the 10 largest agencies
1	Airdrie Transit	No	19	North Bay Transit	No
2	Barrie Transit	No	20	OC Transpo	Yes
3	Brampton Transit	No	21	Peterborough Transit	No
4	Calgary Transit	Yes	22	Réseau de transport de la Capitale	Yes
5	City of Hamilton	No	23	Ride Norfolk	No
6	City of Moncton	No	24	Sanit JohnTransit Commission	No
7	City of Red Deer Transit	No	25	Sarnia Transit	No
8	Durham Region Transit	No	26	Spruce Grove Transit	No
9	Edmonton Transit	Yes	27	St. Catharines Transit Commission	No
10	Fort Sask Transit	No	28	STM	Yes
11	Grand River Transit	No	29	Stratford Transit	No
12	Grande Prairie Transit	No	30	Strathcona County Transit	No
13	Greater Sudbury Transit	No	31	Thunder Bay	No
14	Guelph Transit	No	32	Toronto Transit Commission	Yes
15	Halifax Transit	No	33	Winnipeg Transit	Yes
16	Kingston Transit	No	34	York Region Transit	No
17	Lethbridge Transit	No	35	TransLink	Yes
18	Mississauga Transit	Yes	36	Prince Albert Transit	No

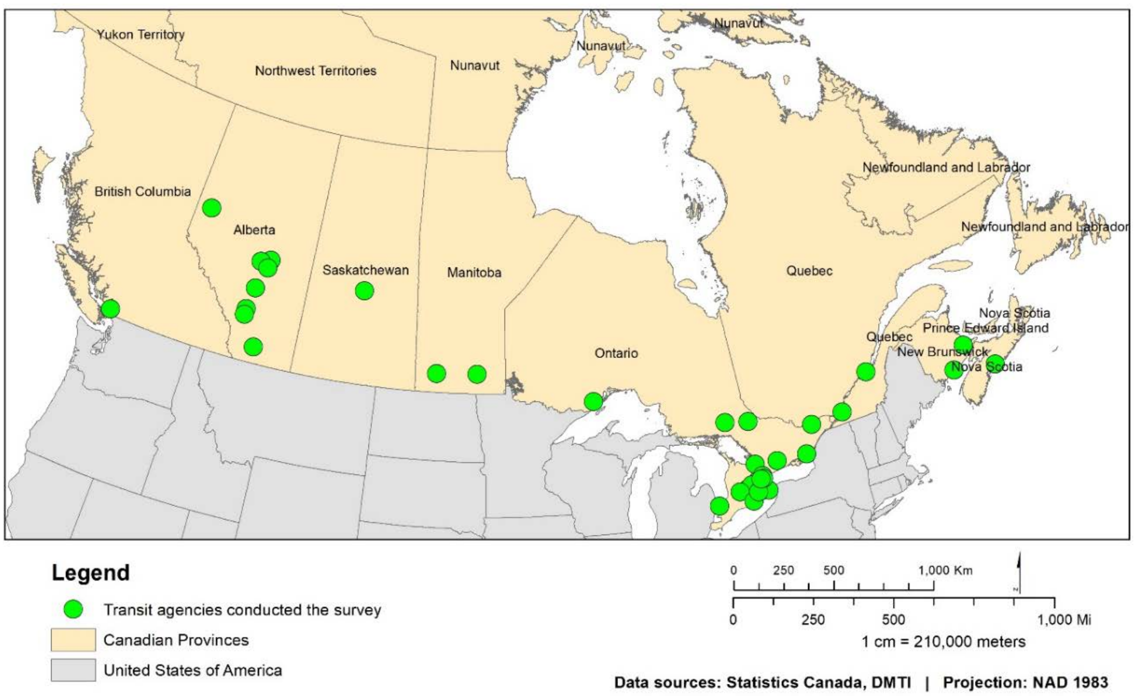


Figure 2-1: Surveyed transit agencies

2.2 Ridership prediction typology

The first section of the survey included general questions about the use of ridership prediction methods. More specifically, it asked participants to report the cases in which they predict ridership. It also inquired if transit agencies have formal guidelines that define the cases where ridership prediction is required and the specifications of prediction models. Participants were then asked whether they predict ridership in terms of linked or unlinked trips.

As seen in Figure 2-2, the majority of surveyed transit agencies predict ridership for the addition of a new route or transit corridor, complete or partial network redesign, changes in fare, and short-term (next fiscal year) and long-term (next 5 or 10 years) system planning. Fewer transit agencies reported ridership prediction for minor scheduling and route adjustments, major scheduling or route adjustments, and service improvements (e.g. introduction of transit signal priority (TSP) system, reserved bus lanes). Respondents were provided with a clear definition of minor and major adjustments in the survey. The former refers to changes affecting less than or equal to 25% of a route schedule or structure, while major adjustments refer to changes affecting more than 25% of a route schedule or structure. Interestingly, responding transit agencies were much less likely to generate ridership predictions for the introduction of a new mode and specific project evaluation. This could be because, normally, these tasks are outsourced to external consultants (this point will be discussed later). The previous results suggest that there may be unwritten thresholds in terms of the scale of service change that would trigger an internal ridership forecast.

Figure 2-3 shows that 78% of the transit agencies (28/36) have no formal guidelines for ridership prediction, and the need and type of prediction methods are largely decided on a case by case basis. In contrast, 19% of transit agencies (7/36) indicated that they follow both formal guidelines and informal practices. Most of the respondents' comments showed that formal guidelines are used for ridership prediction for larger transit service projects. These formal guidelines are related to the use of certain method types (e.g., regional transportation models). In contrast, for smaller projects and route service adjustments, ridership predictions are applied on a case by case basis, or without conducting a ridership prediction.

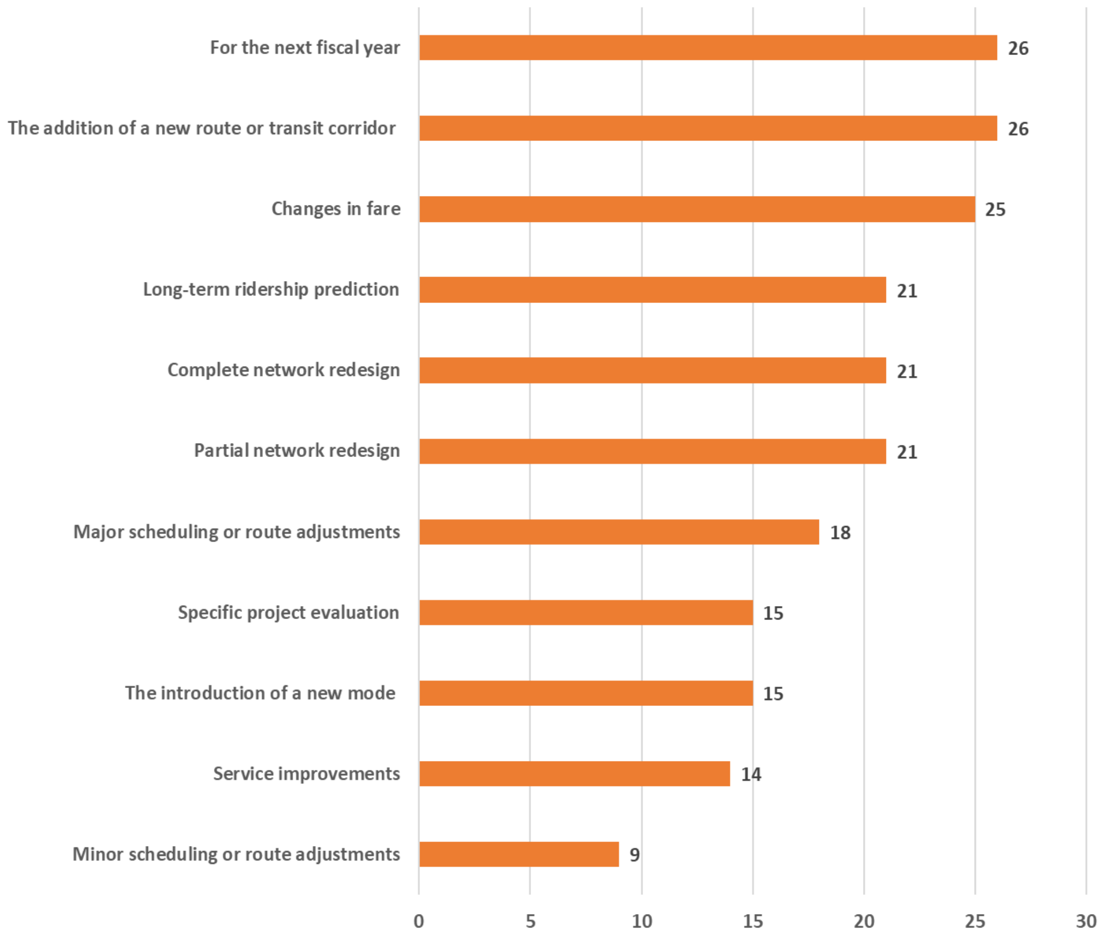


Figure 2-2: Reasons for predicting ridership

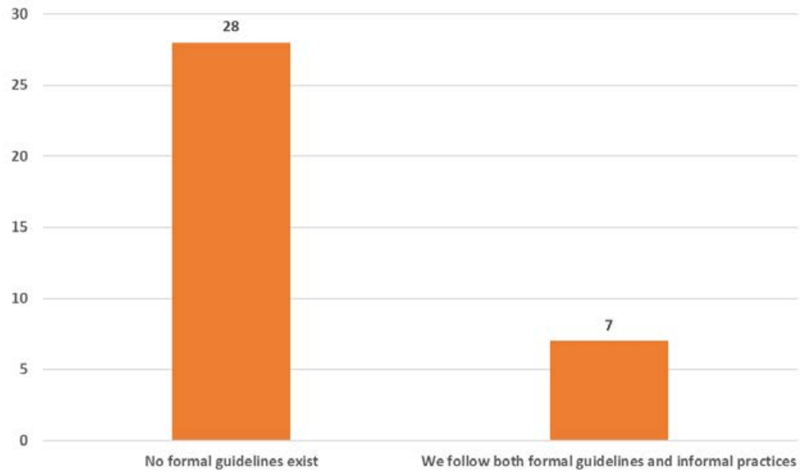


Figure 2-3: Formal guidelines and informal practices for ridership prediction.

About 56% of the transit agencies (20/36) estimate and predict ridership in terms of both linked and unlinked trips (Figure 2-4). In the survey, precise definitions of linked and unlinked trips were provided as an explanatory note. More specifically, linked trips were defined as trips from origin to destination under one transit agency, where individual trips involving transfers are only counted once. Unlinked trips were defined as the number of times passengers board public transportation vehicles. Here, passengers are counted each time they board vehicles no matter how many vehicles they use to travel from their origins to destinations. Data about unlinked ridership is normally easier to collect using APCs and fareboxes. In some places, such as the US, transit agencies are much more likely to predict ridership in terms of unlinked trips than linked trips. However, using unlinked trips in predicting ridership normally yields to the overestimation of ridership, when the transit system requires passengers to transfer more often. This issue makes it harder to compare the impact of transportation projects that aim at increasing ridership across different transit agencies.

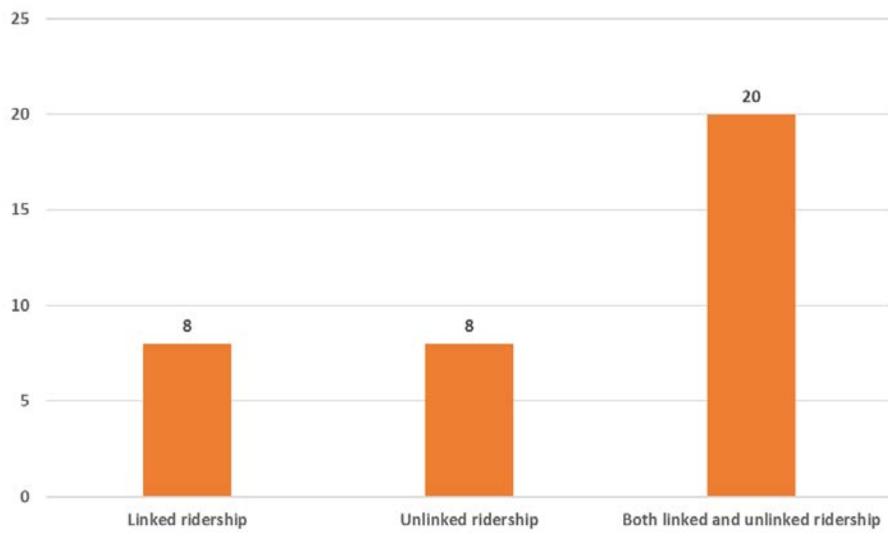


Figure 2-4: Predicting ridership in terms of linked trips and unlinked trips

2.3 Ridership prediction methodology

The third section of the survey included more detailed questions about the types of ridership prediction methods used by the respondent's agency. First, it required respondents to name the types of prediction methods used. These methods were broken down into qualitative/judgement-based and quantitative methods. The first type, qualitative/judgement-based methods, includes the use of professional judgment; rules of thumb or similar-route cases and analysis; trend line analysis; and elasticity based methods. Quantitative methods incorporate the use of more than one explanatory factor in a model. They include the four-step travel demand forecasting models and econometric models. The specific definitions of professional judgment, rules of thumb, and trend line analysis methods were provided in the note section of the survey. Second, this survey section required participants to report whether the methodology for ridership prediction varies according to the scale/scope of the model (e.g., stop, route or network level), the mode of service (e.g., bus, light rail transit, subway services) and horizon (short-term and long-term predictions). Finally, the participants were asked to report the software(s) they use for ridership prediction, if any.

The most frequently used methods for ridership predictions are qualitative/judgement-based in nature (Figure 2-12). About 80% of transit agencies (28/35) use professional judgment for ridership prediction. As defined in the survey, this method relies on the judgment and experience of the analyst. For example, an estimator might use professional judgment to adjust a ridership estimate developed by means of another technique depending on his/her subjective expectation. Around 60% of transit agencies (21/35) use rules of thumb or similar-route cases and analysis to predict ridership. Rules of thumb or similar-route analysis refers to predicting ridership on a given route based on the observed experiences on other routes with similar service areas and frequencies. Finally, about 60% (22/35) and 46% (16/35) of transit agencies use trend line analysis and elasticity based methods to predict ridership, respectively.

In contrast to qualitative methods, only few transit agencies in Canada use quantitative methods for ridership prediction. More specifically, only 20% (7/35) and 17% (6/35) of the transit agencies utilize econometric models (e.g., regression equations) and four-step travel demand forecasting models (trip generation, trip distribution, mode choice and transit assignment) to predict ridership, respectively. It should be noted that most of transit agencies, who are using econometric models, are also using the four-step travel demand forecasting models. The total number of transit agencies that use four-step travel demand forecasting or econometric models are eight transit agencies. All of them, expect for one agency, are among the largest 10 transit agencies in Canada in terms of linked ridership in 2016. Detailed discussion of the model inputs will be presented in the following section.

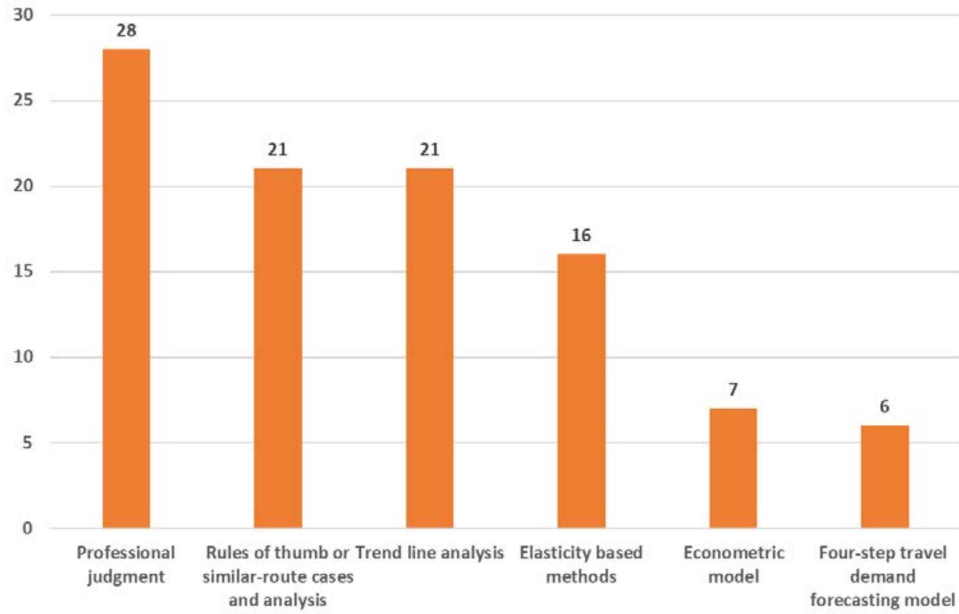


Figure 2-5: Prediction methods used by transit agencies

More than half of the agencies use more than one method of forecasting ridership, depending on the scale and scope of the change (Table 2-3). Interestingly, out of the nine transit agencies that reported operating multimodal systems, only five of them use more than one method for ridership depending on the mode of service (e.g., bus, light rail transit, subway services). The majority of responding agencies (69%, 24/35) do not use different forecasting methods for long-term and short-term predictions (Table 2-3). For transit agencies that use different methods according to the horizon (e.g., short-term vs. long-term), a follow-up question was asked to understand how they define short- and long-term predictions. The results of this question are summarized in Table 2-4 and Table 2-5. The tables show a disagreement between transit agencies regarding the horizon of the short- and long-terms predictions.

Table 2-2: Changes in the prediction methods

		Number of agencies	% of agencies*
According to the scale/scope of the change	Yes	20	57%
	No	15	43%
According to the mode of service	Yes	5	14%
	No	4	11%
	Not applicable	26	74%
According to horizon (short-term vs. long-term)	Yes	11	31%
	No	24	69%

** This percentage is calculated based on the total number of valid responses for the question (n=35)*

Table 2-3: Short-term period definition

Number of years	Number of agencies	% of agencies*
<=1 year	4	44%
<=2 years	1	11%
<=3 years	2	22%
<=4 years	1	11%
<=5 years	1	11%

** This percentage is calculated based on the total number of valid responses for the question (n=9)*

Table 2-4: Long-term period definition

Number of years	Number of agencies	% of agencies*
>1 year	3	30%
>2 years	1	10%
>4 years	1	10%
>5 years	3	30%
>10 years	2	20%

** This percentage is calculated based on the total number of valid responses for the question (n=10)*

Table 2-6 highlights the primary uses for each prediction method. As seen in the table, the majority of the transit agencies use professional judgment, rules of thumb/similar route cases and analysis, and trend line analysis all the time as well as for short-term predictions for route, service, and schedule changes. Elasticity based methods are mainly used to predict the impacts of fare changes. The four-step travel models and econometric models are often used for major projects and for long-term predictions. It is noteworthy that, as mentioned earlier, most of the transit agencies using the quantitative methods are large transit agencies, with possibly more resources to conduct these types of predictions. With only a few exceptions, most of the transit agencies do not use any specialized software for ridership prediction (Table 2-7).

Table 2-5: The primary uses for each prediction method

Method	N.	Method	N.
- Professional Judgment		**- Rules of thumb/similar route cases and analysis**	
Minor scheduling/route adjustments	7	All the time (for major or minor changes)	3
All the time (for major or minor changes)	6	New routes/route adjustments	2
New routes/route adjustments	3	Short-term/minor scheduling/route adjustments	2
Short-term/minor scheduling/route adjustments	2	Long-term changes	1
All the time/with the Elasticity method	1	Major/minor scheduling/route adjustments	1
Major changes	1	Minor bus network adjustments	1
Short-term changes	1	Minor scheduling/route adjustments	1
Special cases	1	Short-term changes	1
To estimate yearly ridership	1	Special cases	1

Method	N.	Method	N.
		To estimate yearly ridership	1
		Yearly changes	1
- Trend line analysis		**- Elasticity based methods**	
Special cases	3	Fare changes	9
Long-term changes	3	Fare changes and short-/medium- term changes	1
All the time (for major or minor changes)	2	Fare/route level changes	1
Short-term changes	2	Fare/service level changes (e.g., frequency)	1
Fare changes	1	In combination with other qualitative methods	1
Long-term and monthly forecasting/validation	1	Major changes	1
Mid-range network changes	1	Budget forecasts/investment plan forecasts	1
Minor scheduling/project related changes	1		
Minor scheduling/route adjustments	1		
Short- to medium- term changes to network	1		
Yearly changes	1		
- Four-step travel demand models		**- Econometric models**	
Long-term changes	3	Regression analysis/annual ridership prediction	2
Major changes	2	For budgeting/fare setting	2
Not yet	1	Long-term changes	1
Special cases (RT projects, network changes)	1	Major changes	1
		Short- and long- term ridership	1

2.4 Data inputs and satisfaction with prediction methods

The fourth section of the survey focused on understanding the inputs of the used prediction methods and the participant's level of satisfaction with them. This section included two subsections. The first was concerned with the use of qualitative/judgement-based methods, if the participants indicated using them in the previous section. This subsection required participants to report the inputs into and the level of satisfaction with such methods. It also asked to rate the importance of certain issues related to these prediction methods. Lastly, it inquired about how participants calibrate and validate the results of their qualitative methods.

The second subsection was concerned with the use of quantitative methods. In this subsection, participants were requested to identify the inputs of each prediction model individually. The survey allowed them to add the inputs for a maximum of six models. It asked the participants to start with the models which have the highest number of explanatory factors. A list of 63 factors was provided to respondents, from which they could select the relevant factors. They were also required to add a brief description of how the selected factors were calculated or defined. Then, the respondents were asked to add any remarks about the models, such as the level of aggregation, base year and forecast years. The remarks could also be related to how the models were integrated and used with other models. This subsection also required participants to report their level of satisfaction with the quantitative methods and to rate the importance of certain issues related to these methods. Finally, it inquired about how participants calibrate and validate the results of their quantitative methods.

A. Qualitative methods

As demonstrated in Figure 2-12, 16 transit agencies indicated using elasticity-based methods to predict ridership. Therefore, a follow-up question regarding the inputs of this method was provided. Table 2-8 shows the results of this question. As shown in the table, fare types and changes in addition to current ridership levels are the most dominant inputs for this method. Only few agencies use other information regarding parking prices, financial changes and service frequency in their ridership predictions.

Table 2-7: Inputs to the elasticity based method

	Number of agencies	**% of agencies***
Fare type and change/ridership	8	53%
Ridership/books and manuals	2	13%
Fare change/service frequency/service hours/ridership	2	13%
Fare change/parking rates	1	7%
Ridership/financial data	1	7%
Service levels/ridership	1	7%

** This percentage is calculated based on the total number of valid responses for the question (n=15).*

Figure 2-13 shows the respondents' satisfaction with the qualitative methods used. It also shows the level of importance of each aspect. This graph was based on the opinion of 33 respondents. As shown in the figure, the average satisfaction level with all aspects varies between 3.2 and 3.7 points, out of 5 points, except for their satisfaction with the suitability of qualitative methods for long-term ridership predictions, which has a score of 2.8 points. In contrast, they rated the importance of long-term ridership prediction among the top most important aspects, which shows a gap in practice.

Transit agencies were most satisfied with the availability of data for their ridership predictions, accuracy of the prediction results, and the time and effort required to generate those predictions, with an overall satisfaction of 3.7 points. In addition, it should be noted that they were less satisfied with the flexibility of the qualitative methods to be used in a wider variety of cases. Regarding the importance scale, respondents rated the availability and accuracy of input data at the appropriate level as the most important aspects. This is followed by the accuracy of the prediction results. This shows the importance of data issues and accuracy of methods for transit agencies.

Different approaches have been reported regarding how transit agencies validate their methods and assess the accuracy of their predictions (Table 2-9). Most of these methods are based on comparing the predicted ridership with the actual ridership. About 33% of transit agencies

(11/33) compared predicted ridership with actual ridership once changes to transit service were implemented. This approach helped them evaluate the accuracy of the original prediction. Nine percent of transit agencies (3/33) compared predicted ridership with actual ridership, while using professional judgment to understand the differences between both. In addition, several respondents indicated they used professional judgments and their broad understanding of the system (15%, 5/33) or they accepted the results based on comparisons with historical data (9%, 3/33). Conversely, 21% of transits agencies (7/33) indicated that they do not normally calibrate and validate the results of their predictions.

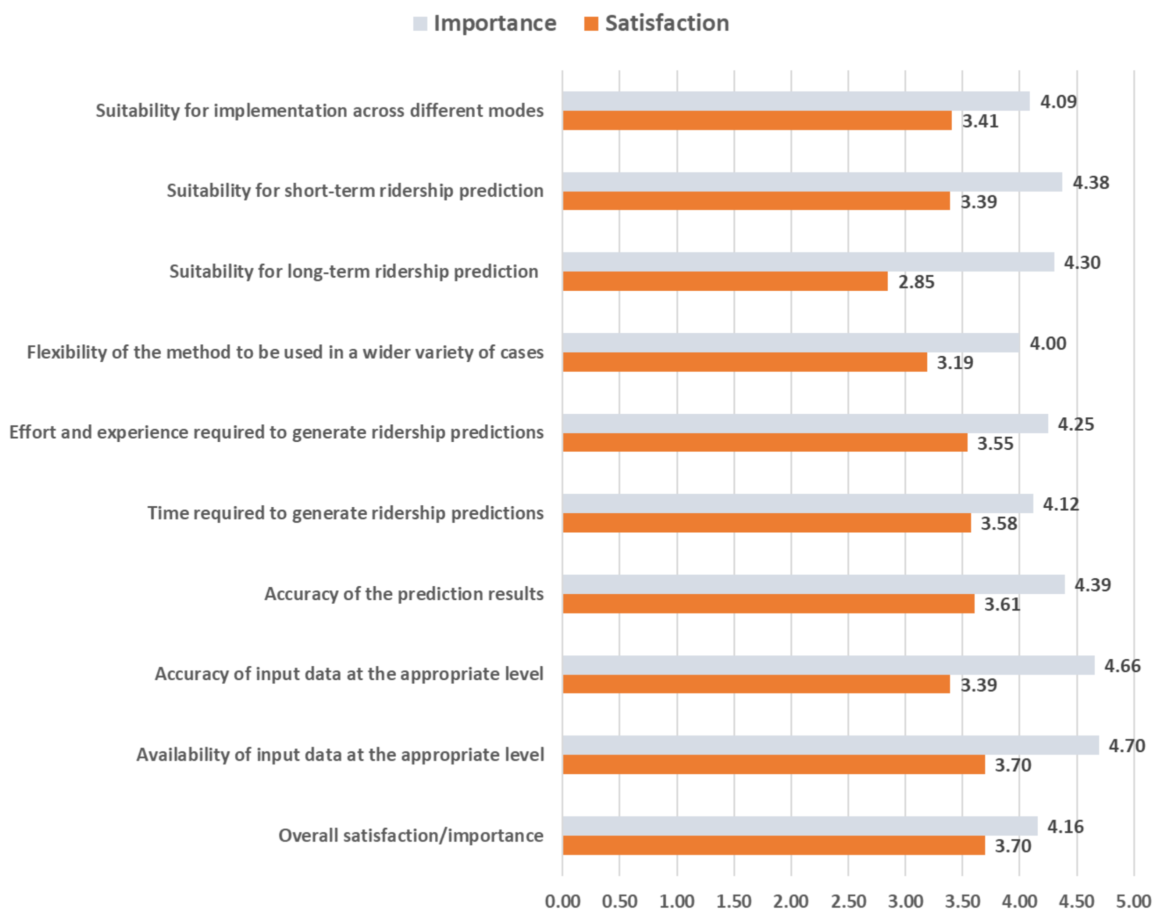


Figure 2-6: Satisfaction with qualitative methods and the importance of difference aspects

Table 2-8: Methods used in assessing the reliability and accuracy of the prediction results

Row Labels	**Number of agencies**	**% of agencies***
Comparison of the predicted and actual ridership	11	33%
Not applicable (e.g., we normally don't calibrate and validate our methods)	7	21%
Professional judgment/board understanding of the system	5	15%

Acceptance based on historical calculations	3	9%
Comparison of the predicted and actual ridership/Professional judgement	3	9%
Best practice research/bench marking	1	3%
Comparison with others lines	1	3%
Comparison with revenues	1	3%
Iterative process to calibrate the prediction results	1	3%

** This percentage is calculated based on the total number of valid responses for the question (n=33).*

B. Quantitative methods

As mentioned earlier, only seven transit agencies use econometric models or/and four-step travel demand forecasting models to generate ridership predictions. Only three out of the eight transit agencies indicated using two models (instead of one model) for ridership prediction. This gives us a total of eleven models. Six of these models are econometric models in nature, while the other five are four-step travel demand forecasting models. The following section discusses the data inputs of the nine models.

Table 2-10 shows the factors that were incorporated in the models with their frequency of usage. The table also includes a brief description of these factors for cases in which such information was provided. For existing ridership factors, transit agencies normally use/produce existing ridership at the network, route, route segment, and stop levels. The most common built environment factors that have been incorporated in the models are population density and street network factors. Other used built environment factors include: freeway and highway network lengths, railway lines and stations, number of local opportunities, distance to downtown, and employment in downtown. In the “other” category, physical barriers (e.g., streams) were mentioned. Regarding the socioeconomic factors, the most commonly used factors are population age, student population, workforce statistics, and household disposable income. The other socioeconomic factors include gender, senior population, unemployment rate, household size, car and driver's license ownership, household composition and expenditure on transport, mode of transport, and employment status and type. The most common sources of these factors are census information and origin-destination household surveys.

Service frequency and transit fares are the most common transit factors used in the models. Other used transit related factors include service network coverage, service span/hours, vehicle revenue hours and miles, composition of fleet and modes, density of transit preferential treatment, and transit service accessibility. In the “other” category, transit speed and travel time factors were mentioned. Regarding the other external/contextual factors, respondents indicated the use of weather conditions, congestion and price of car ownership factors in predicting transit ridership. The only factor that is not currently common in the models, but has been recommended by several transit agencies is “gas prices”.

Figure 2-14 shows the respondents' satisfaction with the quantitative methods used as well as the importance of different aspects. It is noteworthy that only seven transit agencies reported the values used in the figure. As shown, the average satisfaction level with all aspects of the quantitative methods varies between 3.0 and 3.7 points. This is slightly lower than the respondents' level of satisfaction with the qualitative methods, which may highlight a challenge in encouraging transit agencies to utilize these methods. Regarding the degree of importance, respondents rated most of the issues highly. The lowest importance scores were given to the flexibility of the quantitative method to be used in a wider variety of cases and the time required to generate ridership predictions. This could be understandable since quantitative methods are used for specific proposes, while taking longer times to implement in order to provide more precise results. All the transit agencies calibrate the four-step travel demand models using aggregated household survey data and transit and traffic counts. For ridership predictions, they compare predicted ridership with actual historical ridership data.

Table 2-9: Data inputs of the models

Factor	#	Description*
A. Existing ridership		
Network ridership	9	Annual/ monthly ridership reported to CUTA
Route/route segment/stop ridership	5	Ridership count from AFC system/ O-D survey data
Similar route ridership	2	Ridership count from AFC system/ O-D survey data
B. Built environment factors		
Population density	5	Total service area population from Census
Urban land area	2	
Local opportunities: Business	2	
Local opportunities: Recreation	2	
Freeway network length and exits	3	
Highway network length and exits	3	
Street network length/number of intersections	4	
Railway lines and stations	3	GTFS data
Distance to downtown	2	
Employment in downtown	2	Employment data for every TAZ
Dwelling type	1	
Green space	1	
Work location (in/outside Census Subdivision (CSD) of residence)	1	
Other	1	Physical barrier (e.g., streams); other factors according to the model
C. Socioeconomic factors		
Age	4	Census data/ O-D survey data
Gender	3	Census data/ O-D survey data
Student population	4	U-Pass program data/ O-D survey data
Senior population	2	Census data/ O-D survey data
Workforce	5	Labor force statistics/government reports
Unemployment rate	2	Labor force statistics/government reports
Household size	3	Census data/ O-D survey data
Car ownership	3	O-D survey data

Factor	#	Description*
Ownership of driver's license	1	O-D survey data
Household composition	1	O-D survey data
Household disposable income	4	O-D survey data/median income in relation to CPI change
Household's expenditure on transport	2	
Employment status and type	1	Census data/ O-D survey data
Employment/population ratio	1	Census data
Mode of transport	2	O-D survey data/calculated by the model
D. Transit service factors		
Service frequency	5	GTFS data/revenue hours
Service network coverage	3	
Service span/hours	2	Revenue hours
Vehicle revenue hours	3	
Vehicle revenue miles	2	
Fare	5	
Composition of fleet and modes	4	
Density of transit preferential treatment	3	Bus lanes are modelled as distinct facility types
Transit service accessibility	1	
Transit funding	1	
Other	3	Transit speed and travel time/according to the model
E. Other factors (external/contextual)		
Weather	1	
Price of car ownership	2	
Congestion (average level/cost)	2	
Availability of bike-sharing system	1	
Vehicles for hire/ride-sharing availability	1	
Price of gasoline	1	
Other	1	According to the model

** GTFS data: General Transit Feed Specification data; TAZ: Traffic analysis zone; U-Pass program: Universal Transit Pass program for student; CPI: consumer price index; and CSD: Census subdivision*

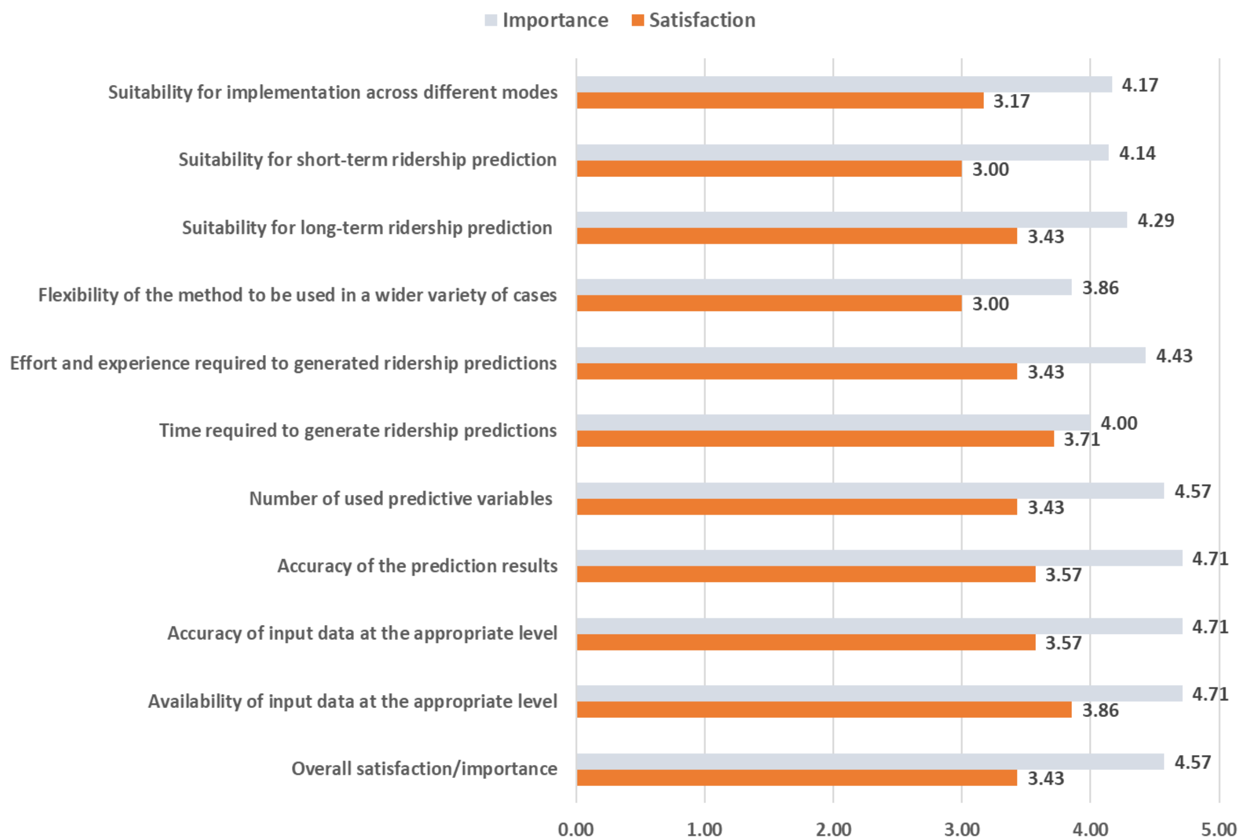


Figure 2-7: Satisfaction with quantitative methods and the importance of different aspects

2.5 Concluding questions

The last section of the survey collected information on the agency's major initiatives to increase transit ridership, their requirements for a robust ridership prediction model and the learned lessons that could benefit other transit agencies. It also inquired if the participants can share documents/guidelines about their ridership prediction processes and recent studies within the past five years investigating the factors related to transit ridership changes. Finally, it asked the participants if they were willing to answer further questions (in a telephone interview) about the prediction methodology they use and their requirement for a robust prediction model.

Table 2-11 shows a list of major initiatives undertaken by transit agencies to increase ridership, while Table 2-12 shows the prediction methods used to predict the impact of those initiatives. Table 2-11 indicates about 24% of the transit agencies that have answered this question did not undertake any recent initiative to increase transit ridership. The main reasons highlighted for this were budget constraints and the need to balance resources with service requirements to maintain ridership in a climate where revenue from ridership is declining. Regarding the transit agencies that undertook some initiatives to increase ridership, 16% of transit agencies are working on expanding and building new LRT lines as a means to attract more riders, particularly choice riders. In addition, about 16% of transit agencies have introduced Universal Post Secondary Student Pass (UPass) to increase ridership. Another 12% of agencies are conducting major scheduling and network adjustments in terms of adding more direct and express runs, and higher frequency routes on main corridors to increase ridership.

Other initiatives undertaken by transit agencies to increase ridership include raising the service frequency/span of the service, drafting a transit/transportation master plan, presenting new bus priority measures and real time arrival information, introducing new branding and marketing strategies, and working on improving transit service quality and reliability. As discussed earlier, many transit agencies used qualitative methods (34%, 5/15) including trends line analysis, professional judgement, rules of thumb or similar-route cases and analysis to predict the impact of the previous initiatives (Table 2-12). Other transit agencies used modelling (34%, 5/15). In contrast, two agencies (13%, 2/15) relied on reports from external consultants, while three agencies (20%, 3/15) did not conduct any ridership prediction to estimate the impact of their initiatives.

Table 2-10: Recent initiatives by transit agencies to increase transit ridership

	Number of agencies	% of agencies*
No recent initiative to increase the transit ridership	6	24%
Expanding/building new LRT lines	4	16%
Lower fares/UPass (Universal Post Secondary Student Pass)	4	16%
Major scheduling and network adjustments	3	12%
Increase the service frequency/time span	2	8%
Transit/transportation master Plan	2	8%
Bus priority measures/ real time arrival time	1	4%
New branding and marketing	1	4%
Improve system quality (i.e., reliability)	1	4%

** This percentage is calculated based on the total number of valid responses for the question (n=25).*

Table 2-11: Used prediction methods for recent initiatives to increase transit ridership

	Number of agencies	% of agencies*
Trend line analysis and professional judgement	4	27%
No prediction	3	20%
Regional transportation model	3	20%
External consultants estimations	2	13%
Modelling and professional judgements	1	7%
Rules of thumb or similar-route cases and analysis	1	7%
Econometric model	1	7%

** This percentage is calculated based on the total number of valid responses for the question (n=15).*

More than half of all survey transit agencies shared some thoughts about a robust ridership prediction model. Table 2-13 shows the model's major features identified by the respondents. As seen in the table, the respondents identified an accurate model based on accurate data inputs as the most important feature of the model. Then, respondents identified the ease of use and understanding as the second most important feature. In other words, respondents see a considerable value in a model that is time and cost effective. Respondents also indicated that the model should consider all the variables and causal factors leading to understanding ridership.

Other identified features for a robust model include being scalable from short-term to long-term predictions. Furthermore, the model should be easy to calibrate and modify, stable (not sensitive to minor adjustments), flexible (to accommodate different scenarios) and implementable at the micro level (route and intersection levels). In addition, it should have neither too many nor too few variables. The previous thoughts provide very detailed and comprehensive ideas about the required model, which cannot be entirely applicable. However, these comments show that some issues are more important than others. Most importantly, transit agencies would like to have an accurate model that is based on actual and accurate data, which is easy to use and understand. The other aspects of the model vary in importance from one agency to another.

Table 2-12: Major features identified for the model

	Number of agencies
Accurate model based on accurate inputs	8
Easy to use/easy to understand/time and cost effective	7
Considers all the variables and causal factors leading to understanding ridership	3
Scalable from short-term to long-term predictions	2
Easy to calibrate/modify	2
Flexible to accommodate different scenarios	2
Have neither too many nor too few variables	2
Stable model/not sensitive to minor adjustments	1
Data source used for validation should be different than the data used to develop the model	1
Reflects best practices and using similar techniques as peer agencies	1
At the macro level (route and intersection levels)	1

More than half of all surveyed transit agencies shared lessons learned from the process of developing and using ridership predictions methods. Table 2-14 groups the lessons learned into some broad categories. As shown in the table, the most common lesson learned is related to the need for reviewing prediction results and accepting them according to the estimator's experience and full understanding of the system (and its current ridership trends and behaviours). Prediction methods and models are only tools and, thereby, consensus and agreement about their quality of results is all what matters. Some respondents stressed that some of the model's results are not necessarily accurate and reviewing them in comparison with other corridors and cases is needed.

The second most common lesson learned is to pay a close attention to the quality of data inputs that are used in ridership predictions, as one of the respondents indicated "Garbage in, garbage out." Every data source may have its limitations and issues. Therefore, a better understanding of these issues is important in the early stages of the model development. The third most common lesson learned is related to investing in new technologies. Thus, upgrading to automated passenger counter (APC) and automated fare collection (AFC) is recommended as these technologies will have a positive effect on the accuracy of input data. Other lessons learned are related to the need of constantly re-testing prediction methods and reflecting on them in order to improve them. Each agency is unique and not every methodology will work for everyone; the importance of having more than one method ready to use; and simple models that do not use too many variables are among the lessons learned by transit agencies.

Finally, five respondents indicated that they can share documents/guidelines about their ridership prediction processes. Therefore, these respondents were contacted by CUTA to collect these reports for additional insights about the current state of ridership practice and current factors affecting ridership.

Table 2-13: Lessons shared by respondents

	Number of agencies
Review/accept the results based on your experience and broad understanding	6
Data quality is crucial to ridership prediction process/Data quality changes according to source	4
Invest in new technology	2
Constantly re-testing the practices/methods and reflect on them	2
Each agency is unique and not every methodology will work for everyone.	2
It is important to have more than one tool ready to use	2
Simplify models—do not use too many variables to predict ridership	2
Determine a correlation of revenue service hours and ridership	1
Do not expect immediate impacts of projects: residents take time to adjust to new travel options	1
It is important to document the prediction process and to acknowledge the risk of loss of expertise	1
Ask many questions about models reliability and accuracy	1

3. SURVEY KEY FINDINGS

This chapter presents the results of a survey conducted to understand the current state of practice in transit ridership prediction in Canada. By completing a web-based survey, information about ridership prediction methods and level of satisfaction with these methods were elicited from 36 transit agencies across Canada. The survey also gathered information to understand the extent to which Canadian transit agencies use models to predict future ridership changes, and the ridership factors incorporated in these models. The main findings of the survey can be summarized into the following points:

- The majority of transit agencies do not have formal guidelines for ridership prediction. The need for and the type of prediction methods are largely decided on a case by case basis. In addition, most of the surveyed transit agencies indicated that the use of ridership prediction methods differ according to the scope of change and scale of change, mode of service and horizon. Therefore, it is not surprising that most transit agencies have not developed a single formal methodology for ridership prediction.
- It is common to see that transit agencies use professional judgment, rules of thumb/similar-route cases and analysis, and trend line analysis for short-term predictions of route, service, and schedule changes. Elasticity based methods are mainly used to predict the impacts of fare changes. Four-step travel models and econometric models are often used for major projects and for long-term predictions.
- In our survey, only eight out of the 36 surveyed transit agencies indicated using different types of quantitative methods for ridership predictions. Most of these transit agencies (7 out of 8) are among the largest 10 transit agencies in Canada. This may reflect the impact of transit agency scale on ridership prediction practice. The other (smaller) transit agencies possibly have less resources and work force to undertake the prediction task.
- Many transit agencies in Canada saw the emergence of automated data collection systems over the past two decades as an opportunity for having more accurate estimates of ridership volumes. This resulted in a higher percentage of agencies that are satisfied with the reliability and quality of current ridership data.
- Most transit agencies do not have guidelines regarding the optimal amount of data (i.e., sample size) required for ridership prediction. However, the emergence of new technologies of automated data collection systems seems to be addressing this issue by providing data at a detailed route-segment and stop levels in large quantities.
- There is disagreement between transit agencies concerning a number of basic issues, such as how they define the horizon of the short- and long-term predictions. This disagreement will make it difficult to have a common guideline for ridership predictions and for sharing and comparing results.
- The majority of transit agencies assess the reliability and accuracy of the prediction results through a comparison of the actual ridership with the predicted values, once

changes to transit services are implemented. A broad understanding of the system and professional judgement are also important methods that are commonly used to evaluate the accuracy of ridership prediction.

- Regarding the factors that were incorporated in the quantitative models, the most common built environment factors include population density and street network factors. Regarding the socioeconomic factors, population age, student population, workforce statistics, and household disposable income are the most common factors. In addition, service frequency and transit fares are the most common transit factors used in the models. Regarding the other external/contextual factors, respondents indicated the use of weather conditions, congestion and price of car ownership factors in predicting transit ridership. The only factor that is not currently common in the models, but has been recommended by several transit agencies, is "gas prices".
- A good percentage of the transit agencies did not undertake any recent initiatives to increase transit ridership. The main highlighted reasons were budget constraints and the need to balance resources with service requirements to maintain ridership in a climate where revenue from ridership is declining. Nevertheless, the most common initiatives undertaken by transit agencies to increase ridership include expanding and building new LRT lines, introducing Universal Post Secondary Student Pass (UPass) and conducting major scheduling and network adjustments.
- Regarding the requirements of robust ridership prediction models, transit agencies generally identified an accurate model that is based on accurate data inputs as the most important feature of the model. They also identified the ease of use and understanding as the second most important feature. In other words, the respondents see a considerable value in a model that is time and cost effective. The model should also consider all the variables and causal factors leading to understanding ridership.
- The most common lesson learned by transit agencies is related to the need for reviewing prediction results and accepting them according to the estimator's experience and full understanding of the system characteristics (and its current ridership trends and behaviours). Prediction methods and models are only tools and, therefore, consensus and agreement about their quality of results is all that matters.

Part IV

Data Overview and Ridership Trends

Table of Contents

List of Tables

List of Figures

1. INTRODUCTION

This chapter presents an overview of the data used in the analysis and the preparatory steps taken to analyze the factors affecting ridership trends. It also discusses the nationwide trends in ridership, the key drivers, and their relationship over time. The chapter starts with an overview of the various collected and obtained longitudinal datasets. Then, it demonstrates the trends of annual transit ridership, measured in millions of linked trips, from 1991 to 2016. Finally, it presents the findings of a number of exploratory analyses in terms of the relationship between ridership and influential indicators nationwide.

2. OBTAINED AND COLLECTED DATA

An extensive spatio-temporal dataset was created using various sources of data on four major sets of indicators: a) built environment, b) socioeconomic factors, c) transit service factors and d) other external/contextual factors. The transit service factors were obtained from the Canadian Urban Transit Association (CUTA). Appendix D provides an overview of the collected datasets, their sources and availability for various years and geographies (spatial units).

Census subdivisions (CSDs) falling within the service area of each transit agency were distinguished by consulting the transit system maps from their website. Census data corresponding to these CSDs were then extracted and aggregated at the transit agency level from 1991 to 2016, controlling for changes in CSD boundaries and census classifications. Each transit agency's service area was used, based on its corresponding CSD, as a “cookie cutter” to extract each transit agency's spatial data such as the length of railways and roadways in the service area.

For the indicators unavailable at the CSD level, data from higher levels of aggregation (city or province level) were used. Since census data are available only at five-year intervals, the data between each two intervals were calculated by linear interpolation.

CUTA provided the data on transit service factors annually since 1991 at the transit agency level. Inspection of the database revealed a number of missing values and discrepancies (e.g. in the calculation of service area population) due to no/inconsistent information reported by the member transit agencies. Furthermore, not all variables were available for the 1991-2016 period. For example, “Number of fixed routes by mode” and “Federal Operation Contribution” fields are not available prior to 2002. Several transit agencies in the years between 2011 and 2016 have not reported the number of fixed routes. Most of these transit agencies are located in British Columbia. Fares by passenger (e.g., Adult Fare Cash) fields are missing for some transit agencies or contain zero values. In addition, fare structure data for systems that are distance-based are not available prior to 2009. Overall, there are many missing values in different fields (e.g., Service Area Size, Total Regular Service Passenger-Kilometres, Total Operating Costs, Total Operating

Revenue, etc.). For the 1993 data, there are a number of fields that are not available (e.g., fare structure data, Total direct operating expenses, Total operating expenses, Total revenue).

It should be noted that several transit agencies have merged over time. One example is Ajax Transit which operated from 1991 to 2001 serving the Town of Ajax in Ontario (with service area population of 70,000 in 2000). In 2001, a new transit agency was formed to serve the Town of Ajax and City of Pickering in the Durham Region in Ontario. This agency, named “Ajax-Pickering Transit” (with service area population of 159,960 in 2001), operated between 2001 and 2006. Later in 2006, all transit agencies in Durham Region were merged to create Durham Region Transit (with service area population of 501,910 in 2006). Another example is the case of Aurora Transit and Newmarket Transit. Aurora Transit was operating from 1992 to 2000 (with service area population of 39,000 in 2000). Newmarket Transit was operating from 1992 to 2000 (with service area population of 65,000 in 2000). Both transit agencies were merged with other transit agencies in 2001 to create York Region Transit (with service area population of 687,500 in 2001). Table 2-1 shows the number of transit agencies by province per year with available data. Remarkably, the number of transit agencies in British Columbia experienced a drastic increase in 1998.

Table 2-1: Number of transit agencies by province by year.

Year	AB	BC	MB	NB	NL	NS	NT	ON	PE	QC	SK	YK
1991	10	2	2	2	1	2		39		3	2	1
1992	8	2	2	3	1	2		39		3	2	1
1993	9	2	2	3	1	2		39		4	3	1
1994	9	2	2	3	2	2		39		4	3	1
1995	9	2	2	3	2	2		39		4	3	1
1996	9	2	2	3	2	2		39		4	3	1
1997	9	2	2	3	2	2		39		4	3	1
1998	9	25	2	3	2	2		40		5	3	2
1999	9	26	2	3	2	2		42		5	3	2
2000	9	26	2	3	2	2		41		5	3	2
2001	9	26	2	3	2	1		36		5	3	2
2002	9	26	2	3	2	2	1	36		6	3	1
2003	9	26	2	3	2	2	1	37		6	4	1
2004	9	26	2	3	2	2	1	38	1	6	4	1
2005	9	26	2	3	2	2	1	38	1	6	4	1
2006	11	26	2	3	2	2	1	34	1	6	4	1
2007	10	26	2	3	2	2	1	35	1	6	4	1
2008	10	26	2	3	2	2	1	36	1	6	3	1
2009	10	26	2	3	2	2	1	36	1	6	3	1
2010	10	26	2	4	2	2	1	37	1	6	4	1
2011	11	27	2	4	2	2	1	38	1	6	4	1
2012	13	27	2	4	2	2	1	39	1	6	4	1
2013	13	27	2	4	2	2	1	39	1	6	4	1
2014	13	27	2	4	2	2	1	40	1	6	4	1
2015	14	27	2	4	2	2	1	40	1	6	4	1

Year	AB	BC	MB	NB	NL	NS	NT	ON	PE	QC	SK	YK
2016	13	27	2	4	2	2	1	40	1	6	4	1

Table 2-2 shows the number of transit agencies with a complete dataset. “Complete dataset” means that the value of the total number of linked trips over the span of several years for a specific transit agency exists. For example, the total number of linked trips is available for 51 transit agencies to run a model that includes all years between 1992 and 2016. However, other data fields might be missing. Year 1992 was considered in the table instead of Year 1991 due to the existence of many missing values for year 1991.

Table 2-2: Number of transit agencies with complete dataset for the years between 1992 and 2016.

	Number of discontinued transit agencies/with no recent data over the past 5 years	With complete yearly records between 1992 and 2016	With complete yearly records between 1999 and 2016	With complete yearly records between 2002 and 2016	With complete yearly records between 2006 and 2016	With complete yearly records between 2011(or later) and 2016
Number of agencies	18	51	73	88	94	103

Based on the above observations and the availability of other collected data (Appendix D), the longest period of analysis, which can benefit from the availability of the majority of the data fields is the 2002 to 2016 period. The collection of other data was based on the 103 transit agencies present at 2016.

3. RIDERSHIP TRENDS

This section presents the development of annual transit ridership measured in millions of linked trips, from 1991 to 2016 at various levels. At the national level, a rather steady rise in ridership is observed since the mid 1990s but that trend levels off after 2014 (Figure 3-1).

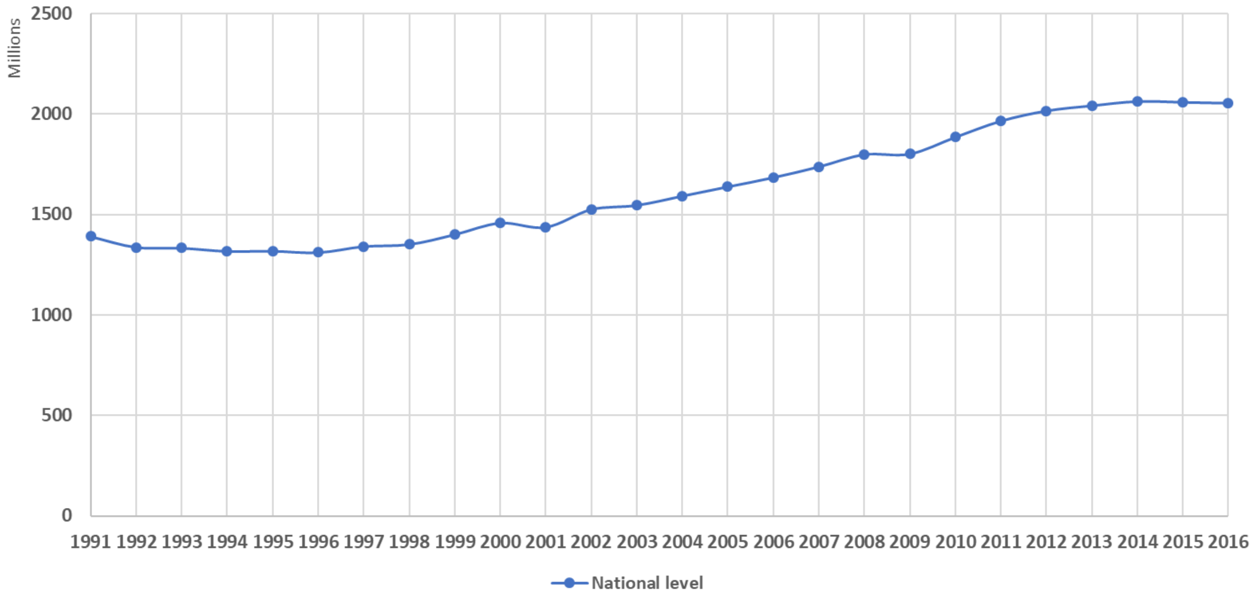


Figure 3-1: Canadian ridership trend from 1991 to 2016 in million of linked trips.

At the province and transit agency levels, there are drastic variations between ridership levels across provinces and transit agencies. Ontario has the highest ridership level, followed by Quebec, British Columbia and Alberta (Figure 3-2). The ridership levels of these four provinces range from 100 to a 1000 annual million linked trips. Manitoba, Saskatchewan and Nova Scotia's ridership range between 10 to 60 million linked trips and the rest of the Canadian provinces have annual ridership levels of below 10 million linked trips. The Yukon Territory, the Northwest Territories and Prince Edward Island have less than one million annual linked trips (Figure 3-3).

Figure 3-4 shows ridership trends by transit agency, indicating that the five highest ridership levels belong to those in the largest urban regions, i.e., Toronto, Montreal, Vancouver, Calgary and Ottawa, respectively. More specifically, Toronto enjoys the largest transit ridership, with more than 530 million linked trips in 2016, followed by Montreal (~ 420 million linked trips in 2016), Vancouver (~230 million linked trips in 2016), Calgary (~ 102 million linked trips in 2016), and Ottawa (~96 million linked trips in 2016). As seen in the figure, for all transit agencies, there was a considerable increase in ridership over time. This trend started to level off noticeably after 2014.

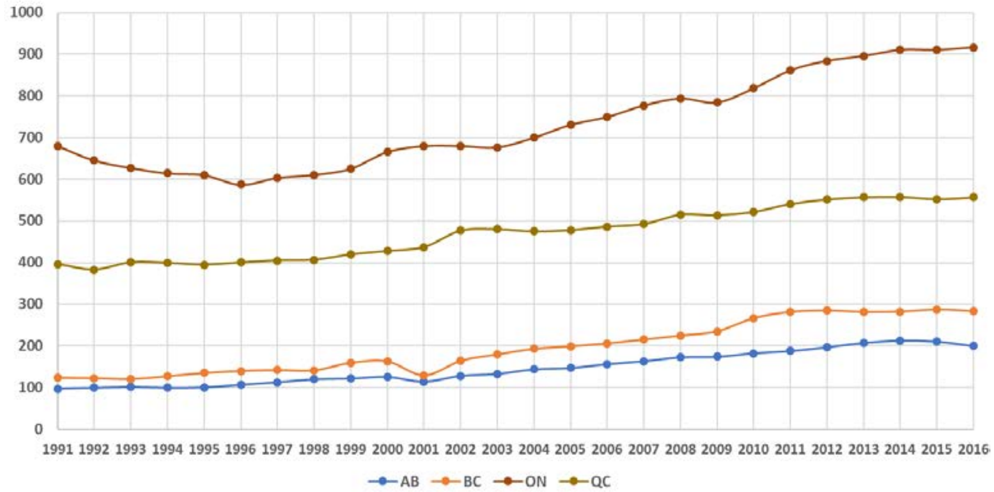


Figure 3-2: Ridership trends of the four provinces with the highest ridership levels from 1991 to 2016.

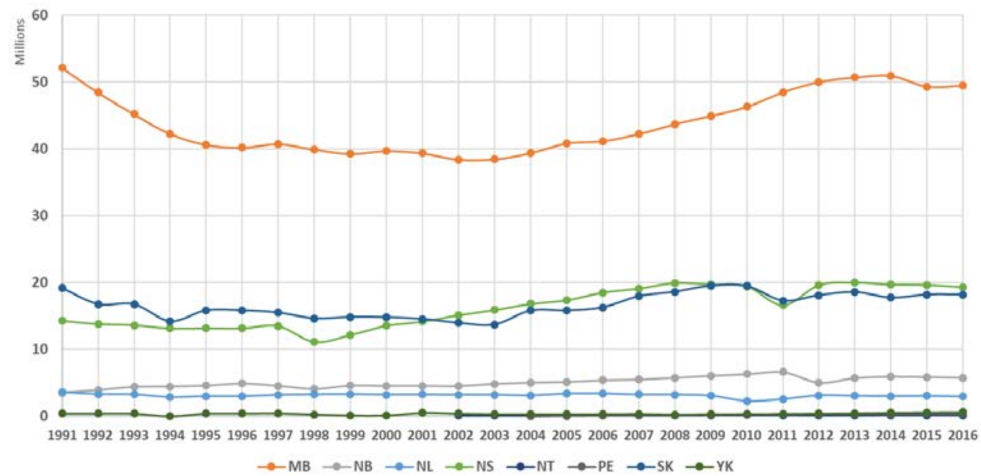


Figure 3-3: Ridership trends of the rest of the provinces from 1991 to 2016.

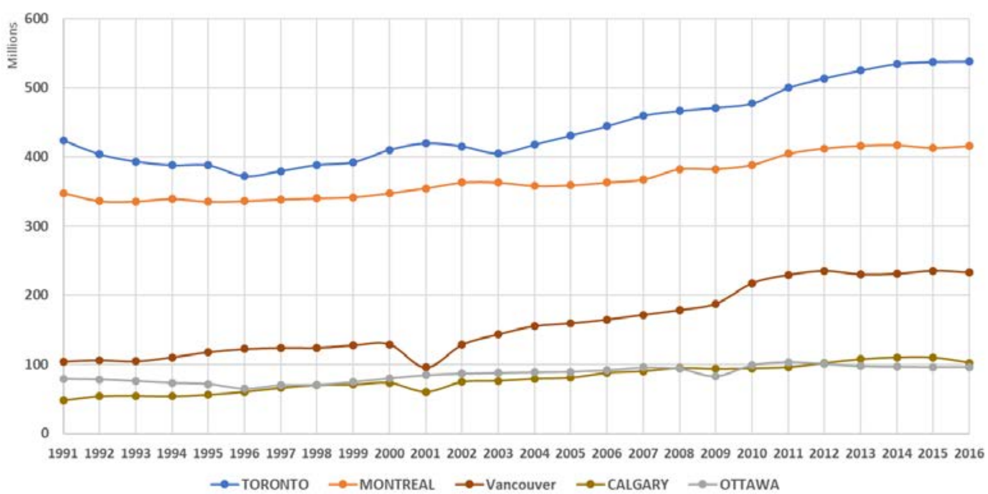


Figure 3-4: Ridership trends of the five largest transit agencies in Canada from 1991 to 2016.

4. RELATIONSHIPS BETWEEN RIDERSHIP AND A NUMBER OF INDICATORS

This section provides exploratory analyses of the relationship between transit ridership for the 103 CUTA transit agencies present in 2016 and a number of influential variables resulting from the literature review (see Part II of this document). The choice of variables was further influenced by the availability of annual data for the entire 1991-2016 period. The only two variables selected to be included in the analysis that goes from 2002 to 2016 and 1997 to 2016 was the highway and major road length, and the percentage of households having a vehicle (owned or leased), respectively .

The national transit ridership trends were plotted against transit demand, indicated by service area population, and transit supply, measured as vehicle revenue kilometres and vehicle revenue hours. Moreover, the relationship of ridership with each of five external factors was explored: gasoline prices, median household income, postsecondary students' population, percentage of households having a vehicle (owned or leased), and highways and major roads length. In addition, the relationship between ridership and one internal factor, namely adult cash fare price, was explored. Finally, a correlation matrix was generated to better understand the relationship between ridership and influential variables.

4.1 Trend line exploratory analysis

The exploratory analysis confirms the expectations based on the literature review for the most part. The growth in ridership follows a trend that is closely matched by the trends of service area population and transit supply (vehicle revenue hours and kilometres). As seen in Figure 4-1, a growing number of linked trips that matches the increase in service area population (based on census data) can be seen. As seen in Figure 4-2 and 4-3, there is a growth in ridership that is visually associated with the increase in transit vehicle revenue kilometres and vehicle revenue hours, respectively.

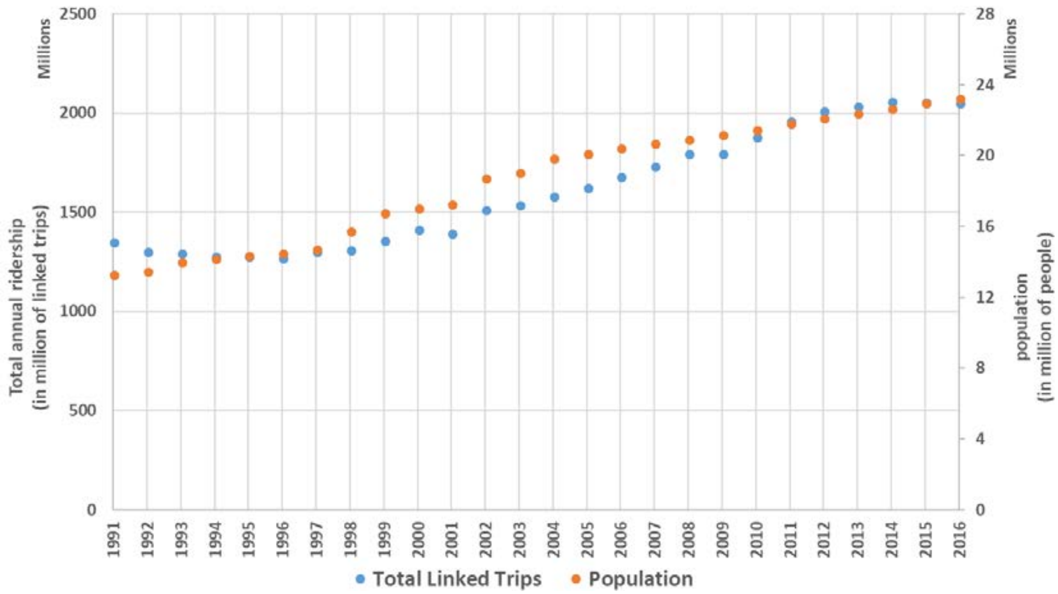


Figure 4-1: Trends of ridership and service area population from 1991 to 2016 at the national level.

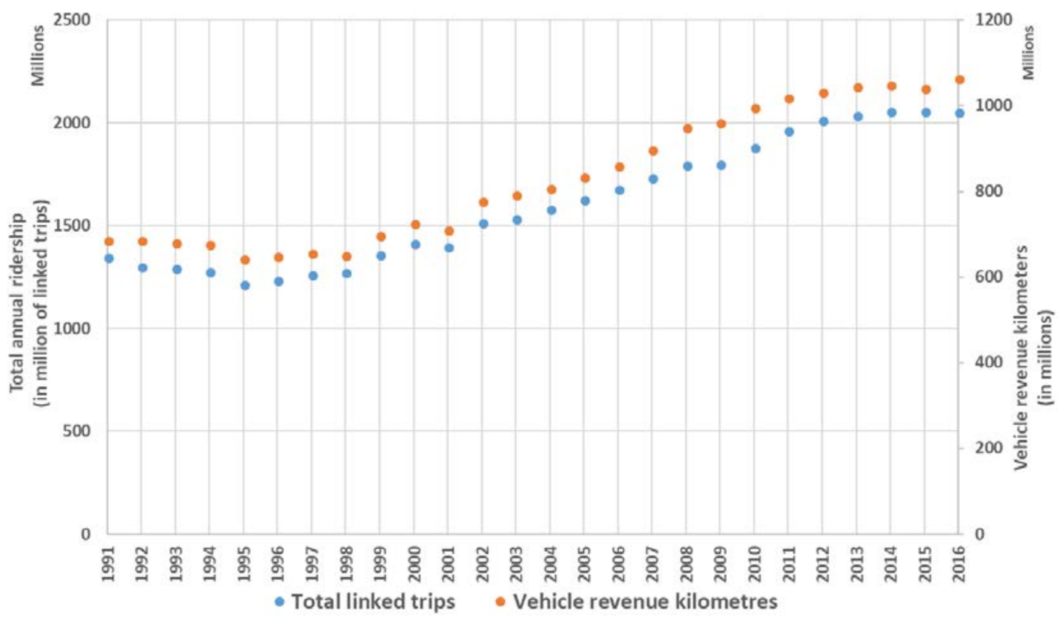


Figure 4-2: Trends of ridership and vehicle revenue kilometres from 1991 to 2016 at the national level.

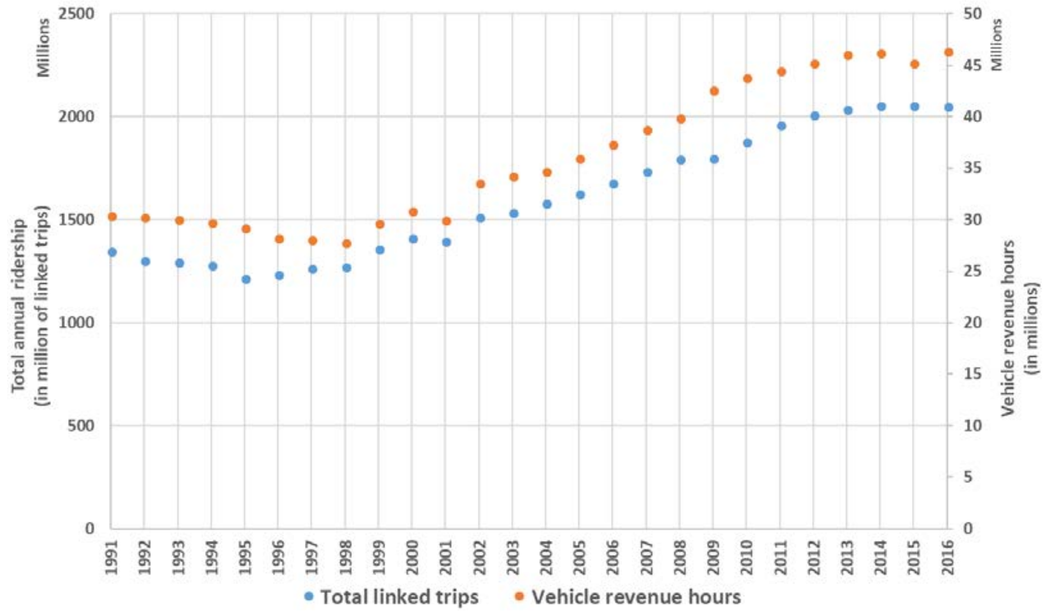


Figure 4-3: Trends of ridership and vehicle revenue hours from 1991 to 2016 at the national level.

Figure 4-4 shows the relationship between ridership and gasoline prices from 1991 to 2016 (in 2016 dollars). The gas prices have been fluctuating over the years. However, over the past couple of years (in 2015-2016), it saw a sharp decrease. In contrast, the total number of linked trips has seen a levelling-off trend during the same period. Figure 4-5 shows the relationship between ridership and adult cash fare prices (in 2016 dollars) from 1991 to 2016 at the national level. As seen in the figure, although fare prices increased over the years at a declining rate, there was a considerable jump in prices over the two years of 2015 and 2016. As mentioned earlier, this was associated visually with a levelling-off trend in ridership over the same two years.

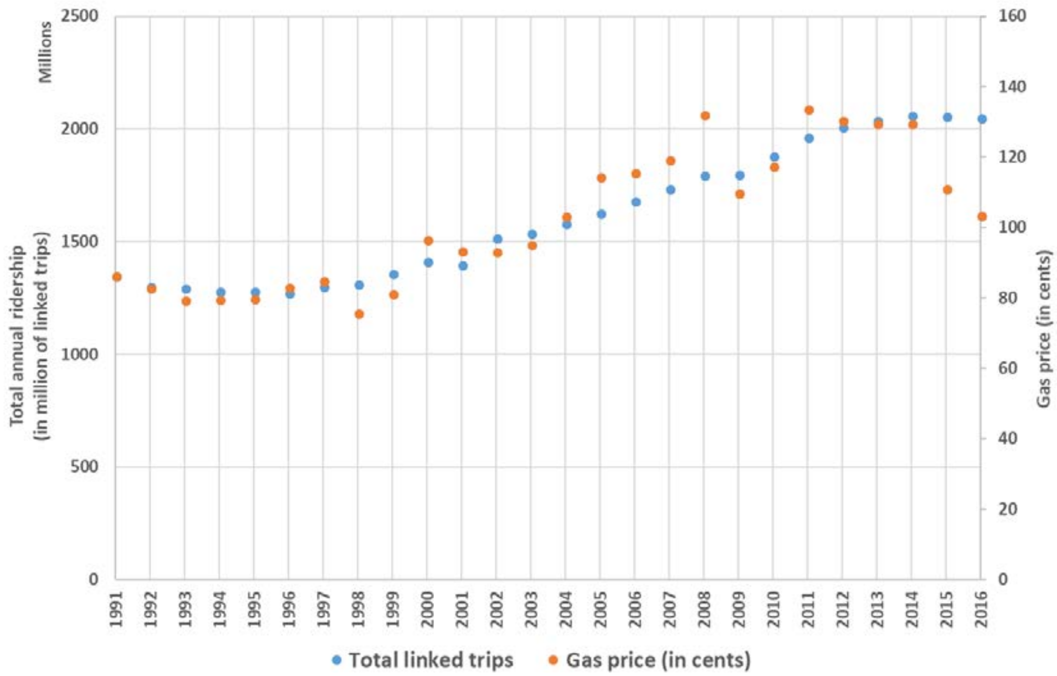


Figure 4-4: Trends of ridership and gas price from 1991 to 2016 at the national level.

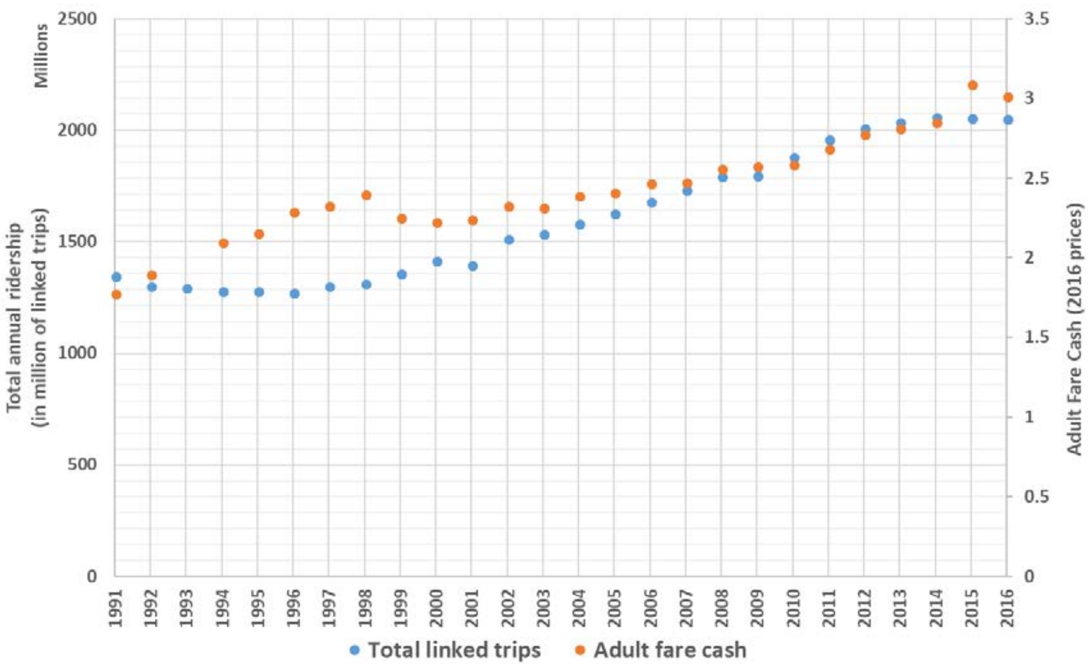


Figure 4-5: Trends of ridership and adult cash fare price (in 2016 dollars) from 1991 to 2016 at the national level.

Figure 4-6 illustrates the relationship between ridership and the median household income (in 2016 dollars). As shown in the figure, there were some fluctuations in the median household income over time. More specifically, the household income declined over the years to reach its lowest in 2006 and 2007 before it started to increase gradually in a slower rate. Nevertheless, it is difficult to visually observe a clear relationship between ridership and median household income over the years between 1991 and 2016. Similarly, no clear relationship can be visually identified between ridership and employment rate (Figure 4-7). In contrast, Figure 4-8 shows a considerable increase in the number of postsecondary students, which is visually associated with the increase in ridership.

Figure 4-9 illustrates the relationship between ridership and the percentage of households having a vehicle (owned or leased). As seen in the figure, the percentage of households having a vehicle did not change considerably between 1997 and 2016, while there was a general increase in ridership over the years. This shows a weak association between the two. Finally, Figure 4-10 illustrates the relationship between ridership and the total length of highways and major roads from 2002 to 2016. As seen in the figure, there was a slowing growth trend in the total length of highways and major roads over the span of several years (from 2012 to 2016). To better understand the previous results, a correlation analysis is presented in the following section to illustrate the association between the discussed factors and ridership.

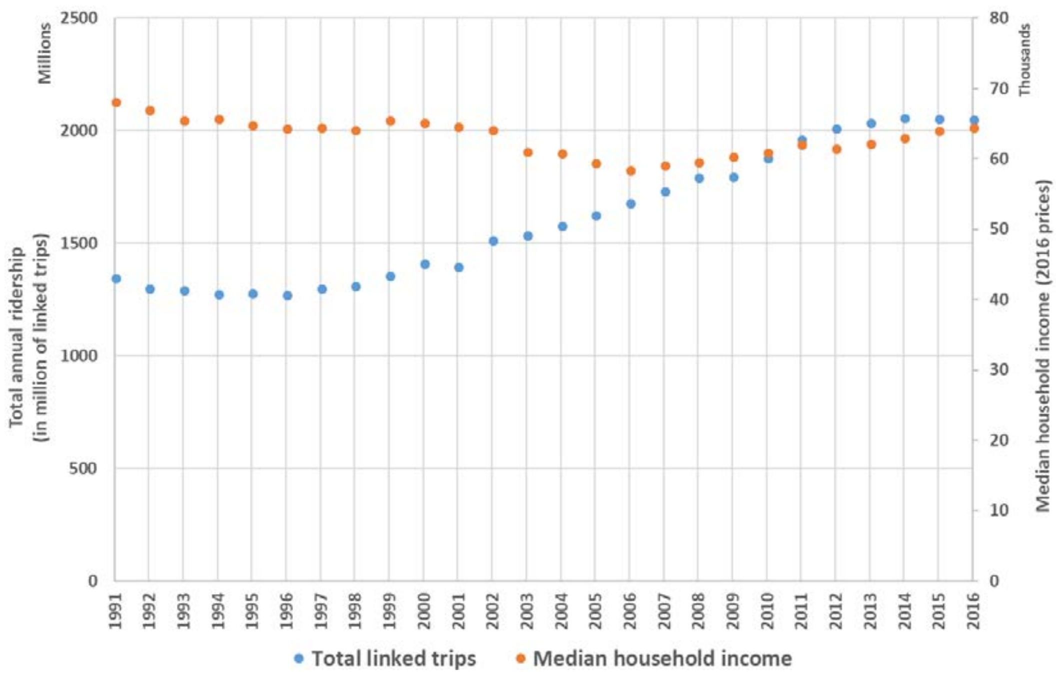


Figure 4-6: Trends of ridership and median household income (in 2016 dollars) from 1991 to 2016 at the national level.

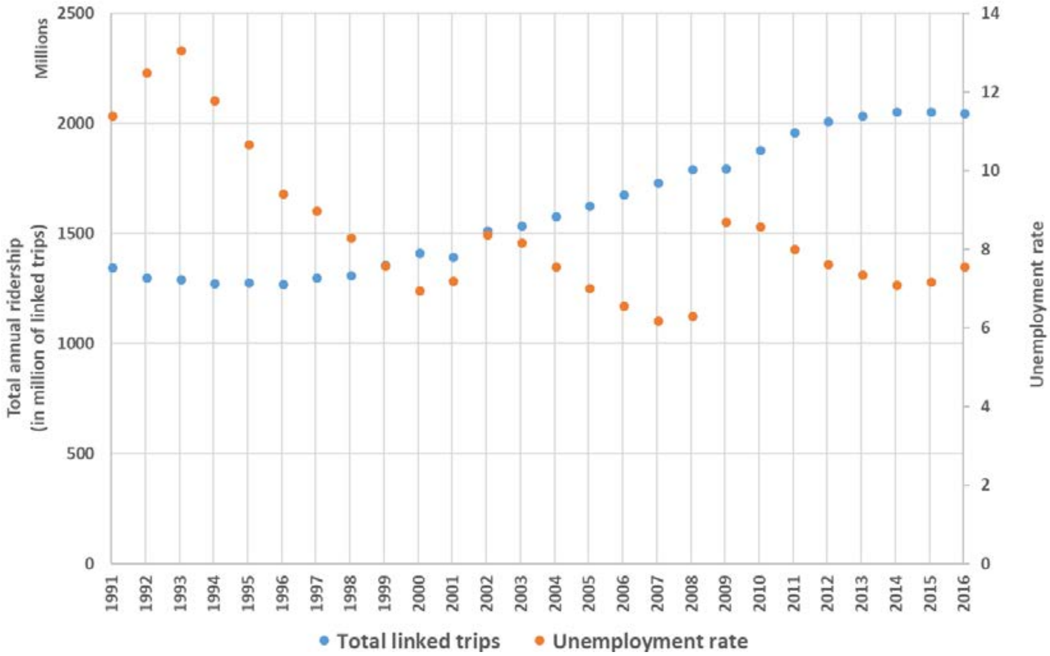


Figure 4-7: Trends of ridership and unemployment rate from 1991 to 2016 at the national level.

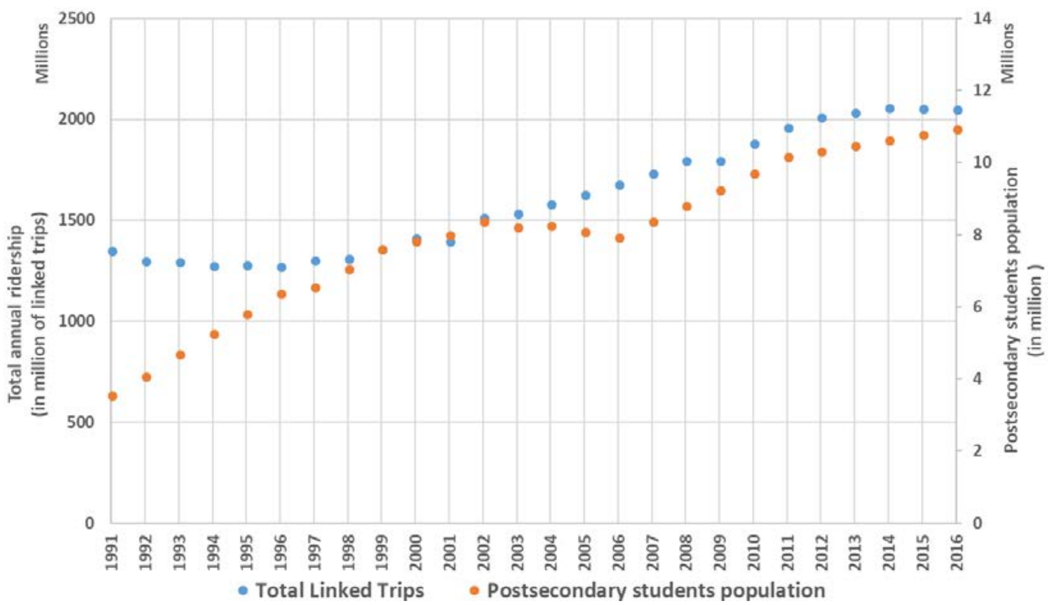


Figure 4-8: Trends of ridership and population of postsecondary students from 1991 to 2016 at the national level.

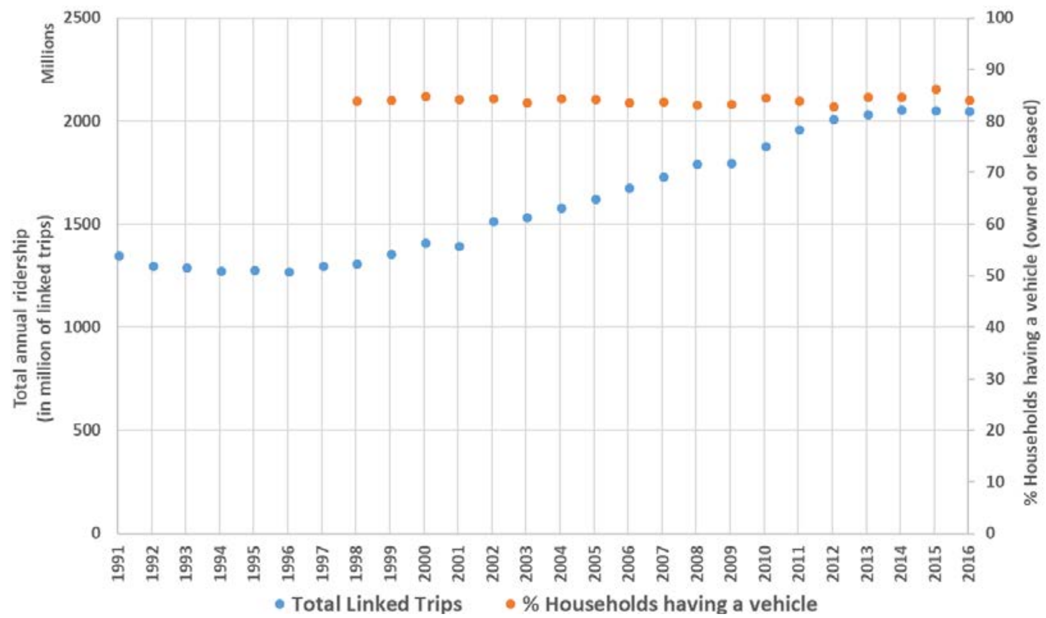


Figure 4-9: Trends of ridership and % of households having a vehicle (owned or leased) from 1991 to 2016 at the national level.

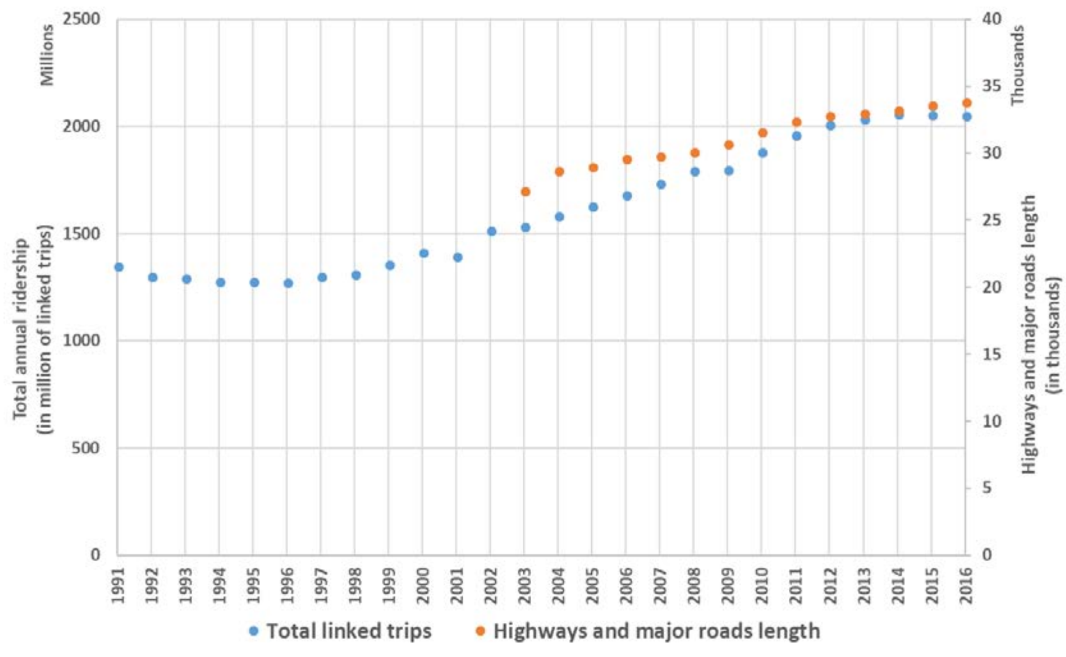


Figure 4-10: Trends of ridership and highways and major roads length from 2002 to 2016 at the national level.

4.2 Correlation analysis

Table 4-1 shows the results of correlation analysis between transit ridership and a number of explanatory variables as well as among the explanatory variables. The correlation analysis was performed in a dataset organized in a longitudinal manner. More specifically, each transit agency had 26 observations corresponding to the years between and including 1991 and 2016.

As seen in the table, most of the correlation signs are as expected: ridership is positively and highly related to service area population, number of postsecondary students, revenue hours and kilometres. The overall number of linked trips by transit agency is also positively associated with existing road and rail infrastructure, indicated by the total length in the service area, which could be proxies for the degree of urbanity. On the other hand, gasoline prices, the number of strike days, unemployment rate and the percentage of households having a vehicle are negatively related to ridership levels as expected. Adult fare prices and median household income have weak, but positive, correlation with ridership. Indeed, this relationship requires further investigation, which will be done in the next phase of the project.

Table 4-1: Correlations between transit ridership (total annual linked trips) and a number of internal and external indicators

	Total linked trips	Service area population	Revenue hrs	Revenue kms	Gas price	Adult fare cash	Median household income	Unemployment rate	Postsecondary students	% Households having a vehicle	Road length	Duration of strikes
Total linked trips	1.00											
Service area population	0.89	1.00										
Revenue hrs	0.97	0.94	1.00									
Revenue kms	0.98	0.94	1.00	1.00								
Gas price	-0.02	-0.04	-0.04	0.01	1.00							
Adult fare cash	0.14	0.19	0.15	0.14	0.21	1.00						
Median household income	0.08	0.16	0.09	0.10	-0.16	0.15	1.00					
Unemployment rate	-0.01	-0.02	-0.01	-0.02	-0.10	-0.02	-0.03	1.00				
Postsecondary students	0.91	0.98	0.95	0.96	-0.01	0.22	0.16	-0.03	1.00			
% Households having a vehicle	-0.10	-0.09	-0.08	-0.08	-0.06	-0.14	0.05	-0.14	-0.09	1.00		
Road length	0.67	0.87	0.73	0.75	-0.03	0.24	0.22	0.04	0.87	-0.08	1.00	
Duration of strikes	-0.18	-0.14	-0.19	-0.17	0.06	-0.14	-0.08	0.02	-0.13	0.12	-0.10	1.00

5. CONCLUSION

Based on the data obtained from CUTA, the temporal changes in the annual transit ridership in terms of linked trips were presented, from 1991 to 2016 at various levels. At the national level, a rather steady rise in ridership is observable since the mid 1990's but that trend levels off after 2014. There are drastic variations between ridership levels across provinces and transit agencies. At the provincial level, Ontario has the highest ridership level, followed by Quebec, British Columbia and Alberta. At the transit agency level, the highest five ridership levels belong to those in the largest urban regions, i.e., Toronto, Montreal, Vancouver, Calgary and Ottawa. The stabilization and slight decreases in transit ridership are observable in the majority of large transit agencies since around 2014.

Furthermore, the relationship between transit ridership and a number of explanatory variables was explored by plotting their trends and by correlation analysis in a longitudinally organized dataset. The results of the exploratory analysis confirm the expectations based on the literature review. Ridership is positively and highly related to service area population, number of postsecondary students, and revenue hours and kilometers. Ridership is also positively associated with existing road and rail infrastructure. On the other hand, gasoline prices, the number of strike days, unemployment rate and the percentage of households having a vehicle (owned or leased) are negatively related to ridership levels as expected.

The next chapter provides an in-depth empirical investigation of variables that explain variations in transit ridership among transit systems and over time. It outlines the correlations between ridership and a comprehensive set of indicators at the transit agency level between 2002 and 2016. Furthermore, the chapter presents the final subset of chosen variables incorporated in the longitudinal econometric models. Finally, it will present and discuss the findings of various models and their validation.

Part V

Modelling Ridership Trends

Table of Contents

List of Tables

List of Figures

1. INTRODUCTION

The goal of this chapter is to provide an empirical investigation of variables that affect variations in transit ridership among transit systems and over time, using data from CUTA member agencies and a comprehensive set of indicators related to a) built environment attributes, b) socioeconomic factors, c) transit service factors and d) other external/contextual factors. This study should improve our understanding of the association of these various factors with past ridership trends, which should consequently improve our ability to forecast future trends.

This chapter consists of three major sections. The first section presents the methodology used in the analysis. The second provides summary statistics of a number of key factors related to ridership trends and investigates several case studies. The third section presents the results and the interpretations of the implemented models. The chapter ends with some concluding remarks and policy implications.

2. METHODOLOGY

2.1 Model development and specifications

The dependent variable is the number of annual linked trips per transit agency. The independent variables include a number of factors related to ridership at the transit agency level. These variables were specified based on the literature review (Part II of this document) and consultation with transit agencies (Part III of this document). The list of variables considered in the analysis can be found in Table 3-1. They have been categorized into four main groups: a) built environment attributes, b) socioeconomic factors, c) transit service factors and d) other external/contextual factors.

To achieve the project goal of assessing the factors associated with the changes in transit system ridership trends across Canada, the modelling approach should meet several important requirements. First, the model should control for the longitudinal nature of the data. This means that the observations of a transit agency are not independent of each other, as they are related over time. In other words, observations are nested within transit agencies. To perform a longitudinal analysis, the data was first organized into a panel structure where each transit agency had 15 observations (from 2002 to 2016) with dependent and independent variables values corresponding to each year.

Second, the model should be able to measure both the between-group variation and within-group variation. The between-group variation is the variation that exists among transit agencies. In other words, this variation is due to differences between transit agencies at the cross-section level. The within-group variation is the variation over time, or between years. The structure of the investigated panel dataset demonstrates, for the most part, higher variations between the transit agencies rather than within the transit agencies. This is due to the observed higher values of the between-group than within-group standard deviations for the majority of variables. Accordingly, the estimator which can take account of both types of variations is the random effects estimator. This estimator can measure the role of independent variables which do not change over time (i.e., time-invariant) as it includes the higher-level (between-group) variation.

Third, the model should be able to account for the simultaneity between the transit service supply and demand, i.e., the two-way causal relationship between them. While the demand for transit is influenced by its supply, transit supply itself is adjusted by transit operators in response to demand. If this simultaneity is not addressed –which is the case in simple one stage ordinary least squares (OLS) regression models – the coefficient estimates could be biased and inconsistent. Unfortunately, the majority of existing studies has not taken this simultaneity into account (Jung, Yu, & Kwon, 2016; Lee & Lee, 2013; Taylor & Fink, 2013).

The endogeneity of transit supply can be accounted for by applying instrumental variables (IVs) in a two-stage least squares (2SLS) method. Here, first, the transit supply (e.g. vehicle revenue hours/miles) is regressed on a number of exogenous variables. Then, the predicted transit supply is incorporated as an independent variable in a second equation to estimate the transit demand (e.g. number of linked trips). Ideally, the instrumental variables should satisfy two conditions. They should be correlated with the endogenous variables (transit supply), but uncorrelated with the error term in the demand model. The IVs used by the few existing similar studies are:

- total population plus political culture (% voting Democrat) as IVs to predict transit vehicle revenue hours (Taylor et al., 2009)
- operating subsidies per capita from federal, state, and local governments (transit subsidy per 1000 people) plus urbanized area population as IVs to predict transit vehicle revenue miles (Storchmann, 2001)
- dummy variable indicating whether the transit service is directly provided by the transit agency or a private operator as an IV for passenger miles travelled (Lee & Lee, 2013)

In this project, transit supply, indicated by vehicle revenue hours, is first estimated by two variables that directly influence it: the population size and operating budget. In the second stage, the number of linked trips is estimated using the predicted value of the revenue hours from the first stage regression model and a set of other independent variables:

First stage:

$$RevHours_{it} = f(Pop_{it}, Operatingbudget_{it}) \quad (1)$$

Second stage:

$$LinkedTrips_{it} = f\left(\widehat{RevHours}_{it}, BuiltEnv_{it}, SocioEcon_{it}, TransitService_{it}, Other_{it}\right) \quad (2)$$

Where i and t respectively denote transit agency and time point. $RevHours_{it}$ is the vehicle revenue hours of transit agency i at time t. Pop is service area population, $OperatingBudget$ is transit agency's total operating budget (in 2016 dollars). $LinkedTrips$ is the number of annual linked trips. $\widehat{RevHours}$ is the predicted number of vehicle revenue hours from the first stage regression model. $BuiltEnv$, $SocioEcon$,$Transit\ Service$ and $Other$ are the built environment, socioeconomic, transit service and other variables, respectively.

Note that all non-dummy variables are transformed into the natural logarithmic form for the following reasons. First, this transformation will normalize the positively skewed distribution of a number of key variables such as the number of linked trips, population, and vehicle revenue hours. Second, this transformation is also applied to the remaining variables in order to facilitate the interpretation and comparisons between coefficients. Third, the log-log formulation allows for the interpretation of model results in terms of elasticities. A number of previous studies investigating transit ridership have also applied the log-log transformation (Boisjoly et al., 2018; Guerra & Cervero, 2011; Taylor et al., 2009).

In the end, three models were estimated using three different datasets. The first model was estimated for all CUTA member transit agencies in Canada, providing a general model that can be used at the national level. The second model was estimated for transit agencies with more than 1.2 million linked trips in 2016. The third model was estimated for transit agencies with less than 1.2 million linked trips in 2016. The second and third models were estimated to investigate the differences between large and small agencies with respect to the influential factors and their coefficients. It should be noted that the threshold of "1.2 million linked trips" represents the median value for linked trips in 2016. This classification is rather coarse, especially as it groups the top major metropolitan transit systems with the rest of the agencies with above 1.2 million linked trips in 2016. However, testing smaller subsets of agencies would have compromised the reliability of the statistical inference due to the reduction in the sample size which was further exacerbated by dividing each group into a training and a test sample for the purpose of validation (discussed in the following section).

2.2 Model validation

The original sample for each model was split into a training sample (90% of the transit agencies) and a validation sample (10%). The validation sample was selected randomly from the original sample. In other words, for the general model, 10 transit agencies out of the 103 transit agencies were selected randomly and their data were kept aside for model validation. Similarly, for the other two models, 5 transit agencies were selected randomly and their data were set aside for model validation. As such, each model was estimated based on the relevant training sample. The estimated model was subsequently used to predict the linked trips of each agency in the validation sample every year. Finally, the predicted yearly values of linked trips were compared to the corresponding actual values for each agency in the validation sample. This approach is a common method that has been extensively used in the literature to understand the quality of model predictions.

2.3 Variable selection methodology

The inclusion of variables in the models was based on a systematic process. First, a correlation analysis was performed for all the variables within each group. In order to do that, the dataset was organized in a longitudinal manner, with each transit agency having 15 observations corresponding to the years between and including 2002 and 2016. As expected, several independent variables were correlated with each other within and between the four variable groups. In order to choose a subset of variables which best explained variations in transit ridership, several steps were taken. First, correlations between the independent variables and the dependent variable were computed for each variable group (see the column *correlation with linked trips* in Table 3-1 for the results). Second, various models were estimated with different combinations of variables having high correlation with the dependent variable. Here, we included the variables emphasized by and in line with theory (or those debated, such as the presence of Transportation Network Companies) and controlled for multicollinearity between pairs of variables and the models' mean collinearity indicated by the variance inflation factor (VIF) indicator. The final choice of variables was based on theoretical expectations, while controlling for collinearity between variables, taking into account the number of missing values for different variables and assessing various model performance statistics such as the VIF, overall R-squared and Wald Chi-squared. Finally, only significant variables were kept in the models.

3. EXPLORATORY ANALYSIS

3.1 Summary statistics of variables tested for the inclusion in the final model

An exploratory analysis of the trends in ridership and key variables is presented in this section. A list of variables was prepared and tested for the inclusion in the final model(s). Table 3-1 presents the summary statistics of all the variables organized by variable group, while Appendix D provides an overview of the description of the tested variables, their sources and availability for various years and geographies (spatial units). In the table, all non-dummy variables were transformed into natural logarithmic form, while all monetary values were transformed to 2016 Canadian dollars. As discussed above, the table includes only observations for the years between 2002 and 2016 inclusive. It should be noted that this variable list was generated based on the results of the initial literature review (Part II of this report) and consultation with transit agencies (Part III of this report) as well as the availability of data sources.

Table 3-1: Summary of variables tested for inclusion in the final model per variable group

Variable	Obs.	Mean	Std. Dev.	Min.	Max.	Correlation with linked trips	Used in the models
a. Built environment factors							
Transit agency area	1387	5.44	1.47	0.96	11.06	0.58	No
Total population	1387	11.32	1.38	7.48	14.82	0.94	Yes

Variable	Obs.	Mean	Std. Dev.	Min.	Max.	Correlation with linked trips	Used in the models
Population density	1387	5.88	1.34	-0.36	8.32	0.41	No
Length of highways & major roads	1302	5.05	1.31	0.66	7.79	0.87	No
Length of railways	1163	3.50	1.09	-0.16	5.97	0.70	No
# Local opportunities	1387	8.52	1.28	4.48	11.93	0.94	Yes
# Occupied private dwellings	1387	10.39	1.41	0.00	13.92	0.95	No
# Rooms per dwelling (avg.)	1387	1.90	0.24	0.00	4.54	0.01	No
$ Dwelling value (avg.)	1387	12.61	0.56	0.00	15.75	0.43	No
# Band dwellings	1387	0.73	1.58	0.00	9.79	0.18	No
# Private dwellings in need of major repairs	1387	7.69	1.32	0.00	11.32	0.93	No
% Apartment dwellings	1387	3.20	0.48	1.09	4.39	0.57	Yes
% Row house dwellings	1387	2.01	0.47	0.38	3.31	0.02	Yes
% Single family dwellings	1387	4.12	0.24	2.72	4.52	-0.58	Yes
% Rented dwellings	1386	3.38	0.35	0.95	4.41	0.21	Yes
% Owned dwellings	1386	4.22	0.16	1.52	4.52	-0.33	No
Household density	1387	5.06	1.26	0.00	7.48	0.38	No
b. Socioeconomic factors							
% Female	1387	3.94	0.03	3.81	4.00	-0.18	No
% Child (age 0-15)	1387	2.84	0.15	2.47	3.25	-0.12	No
% Senior (age 65 and over)	1387	2.64	0.42	0.63	3.51	-0.17	Yes
% Canadian citizen	1386	4.53	0.10	2.94	4.58	-0.18	No
% Recent immigrant	1386	2.53	0.64	0.18	3.97	0.50	Yes
% Population working from home	1387	1.33	0.28	0.00	2.20	0.26	Yes
% Postsecondary students	1387	3.72	0.20	0.00	4.62	0.38	Yes
Unemployed rate	1371	1.92	0.35	1.10	4.24	-0.02	No
Participation rate	1386	4.21	0.26	2.42	6.88	0.16	No
# persons per household (avg.)	1387	0.94	0.25	0.32	3.68	0.16	No
# In the labour force	1386	10.69	1.42	5.08	14.21	0.94	No
$ Median income	1386	10.31	0.18	8.28	10.97	0.15	No
$ Median household income	1386	11.00	0.22	8.93	11.69	0.25	Yes
$ Average gross rent	1386	6.83	0.24	4.91	7.75	0.17	No
$ Average major payments for owners	1386	7.08	0.24	5.31	7.86	0.31	No
$ Household expenditure on purchase of automobiles	1358	7.57	0.17	7.02	8.11	0.05	No
$ Household expenditure on private transportation	1365	9.26	0.12	8.97	9.63	-0.22	No
$ Household expenditure on transit	1365	7.08	0.32	5.85	7.75	-0.08	No
$ Household expenditure on parking	1364	5.26	0.38	3.25	6.36	-0.06	No
$ Person expenditure on public transit	1365	6.14	0.39	3.64	6.75	-0.16	Yes
Number of vehicles per person	1363	3.88	0.08	3.69	4.16	0.05	No
% Households having a vehicle	1367	4.44	0.04	4.17	4.54	-0.08	No
% Households with 1 vehicle	1367	3.81	0.07	3.44	4.01	0.16	No
% Households with 2 or more vehicles	1367	3.66	0.14	3.12	4.11	-0.16	No
% of people work within their CSD of residence	1387	4.03	0.40	2.39	4.48	-0.02	No
% of people work outside their CSD of residence	1387	2.77	0.90	0.64	4.25	0.02	Yes

Variable	Obs.	Mean	Std. Dev.	Min.	Max.	Correlation with linked trips	Used in the models
% of workers with no fixed place of work	1387	2.44	0.27	1.37	3.06	0.19	No
c. Transit service factors							
Multi-modal system (dummy)	1387	0.07	0.25	0.00	1.00	0.61	No
Number of fixed bus routes	1387	2.63	1.19	0.00	5.51	0.93	No
Total operating expenses	1359	15.92	1.98	11.17	21.26	0.99	Yes
Total regular service passenger revenue	1238	14.89	2.27	8.87	20.84	0.99	No
Vehicle revenue Hours	1322	11.26	1.82	6.31	16.21	0.99	Yes
Vehicle revenue Kilometers	1122	14.60	1.82	9.86	19.23	0.98	No
Disruption >= 20 days (dummy)	1387	0.01	0.09	0.00	1.00	-0.02	No
# Buses	1347	3.56	1.68	0.41	7.58	0.98	No
# Total transit vehicles	1347	3.57	1.71	0.41	8.04	0.98	No
# Total low floor buses	1387	3.13	1.79	0.00	7.58	0.93	No
# Total articulated buses	1387	0.58	1.34	0.00	5.89	0.65	No
Service span Tuesday (hours)	1049	2.85	0.20	2.08	3.18	0.74	No
Service span Saturday (hours)	1008	2.78	0.28	1.61	3.18	0.70	No
$ Adult fare cash	1383	0.92	0.25	0.18	2.33	0.46	No
$ Adult fare unit price	1366	0.75	0.25	-0.01	1.97	0.38	No
$ Adult fare monthly pass	1328	4.23	0.28	3.40	5.22	0.59	No
$ Student fare cash	1326	0.78	0.31	-0.07	1.90	0.15	No
$ Student fare unit price	1262	0.56	0.24	-0.19	1.55	0.13	No
$ Student fare monthly pass	1272	3.92	0.33	3.00	4.86	0.29	No
$ Senior fare cash	1351	0.77	0.29	-0.25	1.90	0.14	No
$ Senior fare unit price	1295	0.52	0.24	-0.36	1.55	0.06	No
$ Senior fare monthly pass	1282	3.78	0.34	2.51	4.73	0.06	No
D. Other external/contextual factors							
$ Gas price	1375	4.75	0.14	4.40	5.00	0.10	Yes
Passenger vehicle registration fees (CPI)	1387	4.69	0.14	4.61	5.57	-0.03	No
Passenger vehicle insurance premiums (CPI)	1387	4.92	0.18	4.61	5.36	-0.02	No
Other passenger vehicle operating expenses (CPI)	1387	4.92	0.17	4.61	5.31	-0.04	No
Local and commuter transportation (CPI)	1387	4.84	0.14	4.61	5.09	0.05	No
Parking fees (CPI)	1387	4.94	0.20	4.61	5.20	0.02	No
Presence of Uber (dummy)	1387	0.07	0.26	0.00	1.00	0.12	Yes
Presence of bike-sharing systems (dummy)	1387	0.02	0.15	0.00	1.00	0.30	Yes
Automated fare collection system (dummy)	1387	0.12	0.32	0	1	0.20	Yes
Average annual temperature (F)	1352	3.78	0.14	3.02	4.00	0.22	No
Average annual rainfall precipitation (mm)	1146	6.54	0.57	4.36	7.90	0.15	No
Average annual snowfall (cm)	1142	4.89	0.90	0.90	0.91	-0.13	No
# Total road motor vehicles	1387	14.90	1.16	9.97	15.96	0.01	No

Notes: All variables were transformed to natural logarithm form; All monetary values were transformed to 2016 Canadian dollars.

3.2 Relationships between ridership and a number of indicators

This section explores the trends in ridership and key factors for a few individual transit agencies in Canada using the data prepared for the model development. Four transit agencies were selected for this analysis. Two performed poorly in terms of experiencing the sharpest decline in ridership between 2002 and 2016 (i.e., Saint John Transit, Cornwall Transit) while the other two (Brampton Transit, Comox Valley Transit) performed well in maintaining and growing their ridership levels during the same time period. Further investigation of the relationship between key variables and ridership trends at the national level can be found in Part II of this report.

Two transit agencies that experienced major declines in ridership between 2002 and 2016 are Saint John Transit, NB and Cornwall Transit, ON. They belong respectively to the groups of "large" and "small" transit agencies of this study (according to the ridership threshold of 1.2 million linked trips), each experiencing one of the sharpest declines in ridership within its group. In contrast, two transit agencies that experienced one of the largest increases in ridership in their respective groups between 2002 and 2016 are Brampton Transit, ON and Comox Valley Transit, BC. The purpose of this section is to describe visually the temporal trends in ridership and key associated factors in each of these four transit systems.

Figure 3-1 shows Saint John Transit's ridership trends in relation to each of vehicle revenue hours, total population, gas price and personal expenditure on transit. As shown in the figure, Saint John Transit experienced a large decrease in ridership, particularly since 2010. This trend is matched with a similar decline in vehicle revenue hours, albeit at a different rate. Both peaked between 2008 and 2011 before experiencing a decline. Furthermore, the decline in transit ridership after 2011is well matched with a drop in the transit agency's service area population, highlighting the association between the two factors. In addition, the decline in gas prices after 2011 is visually associated with the drop in ridership in Saint John, while there was an increase in the total paid transit fares per person.

Figure 3-2 shows Cornwall Transit's ridership trends in relation to the same set of factors. Cornwall Transit saw a sharp decline in ridership in the early years followed by a short period of growth until 2011, beyond which ridership levels have shown only small fluctuations. The vehicle revenue hours has shown a similar trend line. The service area population remained almost the same for a longer period of time (since 2006). Gas prices have similarly increased over time and then declined after 2014, which corresponds visually to changes in ridership after 2014, while there was some increase in the total paid transit fares per household over time, particularly after 2014.

Figures 3-3 and 3-4 show the trends in Brampton Transit and Comox Valley Transit, respectively. In the case of Brampton Transit, a steady increase in ridership is observed over time. This increase is visually associated with the increase in both vehicle revenue hours and service area population, while gas prices and personal expenditure on transit experienced some fluctuations. Similar results can be observed in Figure 3-4. Growth in ridership for Comox Valley Transit followed a trend that matches the growth in vehicle revenue hours and population. The decline in gas prices after 2014 is visually associated with a slight decline in ridership, while there were fluctuations in personal expenditure on transit.

To summarize, ridership is highly associated with both vehicle revenue hours and transit agency's service area population over time. Gas prices and expenditures on transit are also associated, but to a lesser visible degree. In this section, ridership association with four different factors were inspected visually using four individual case studies. Nevertheless, in order to better understand the presented results, while controlling for the impacts of different influential factors, the following section presents the results of three models.

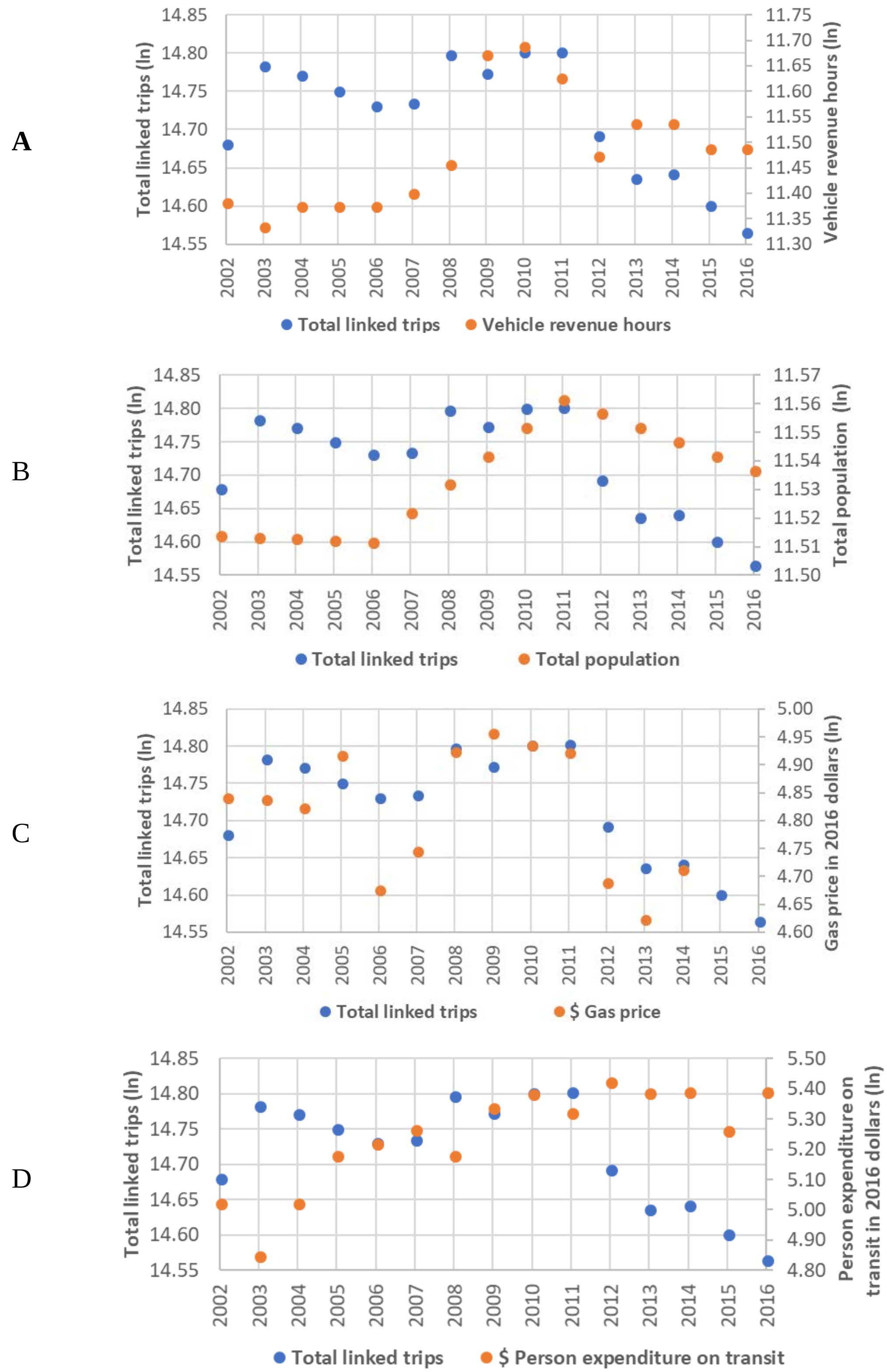


Figure 3-1: Saint John Transit trends of ridership in relationship to: A. Vehicle revenue hours, B. Total population, C. Gas price, and D. Person expenditure on transit.

A

B

C

D

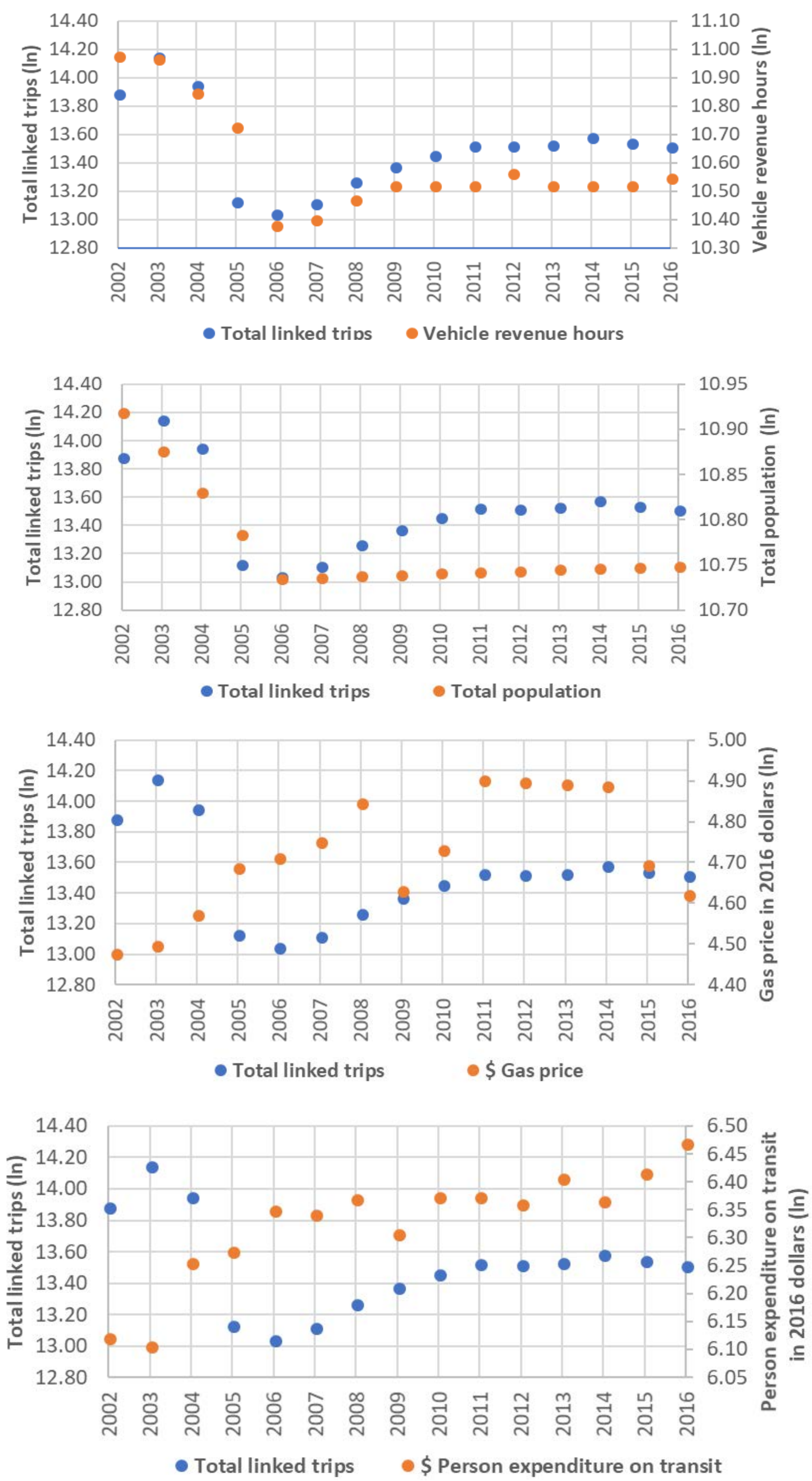


Figure 3-2: Cornwall Transit trends of ridership in relationship to: A. vehicle revenue hours, B. Total population, C. Gas price, and D. Person expenditure on transit.

A

B

C

D

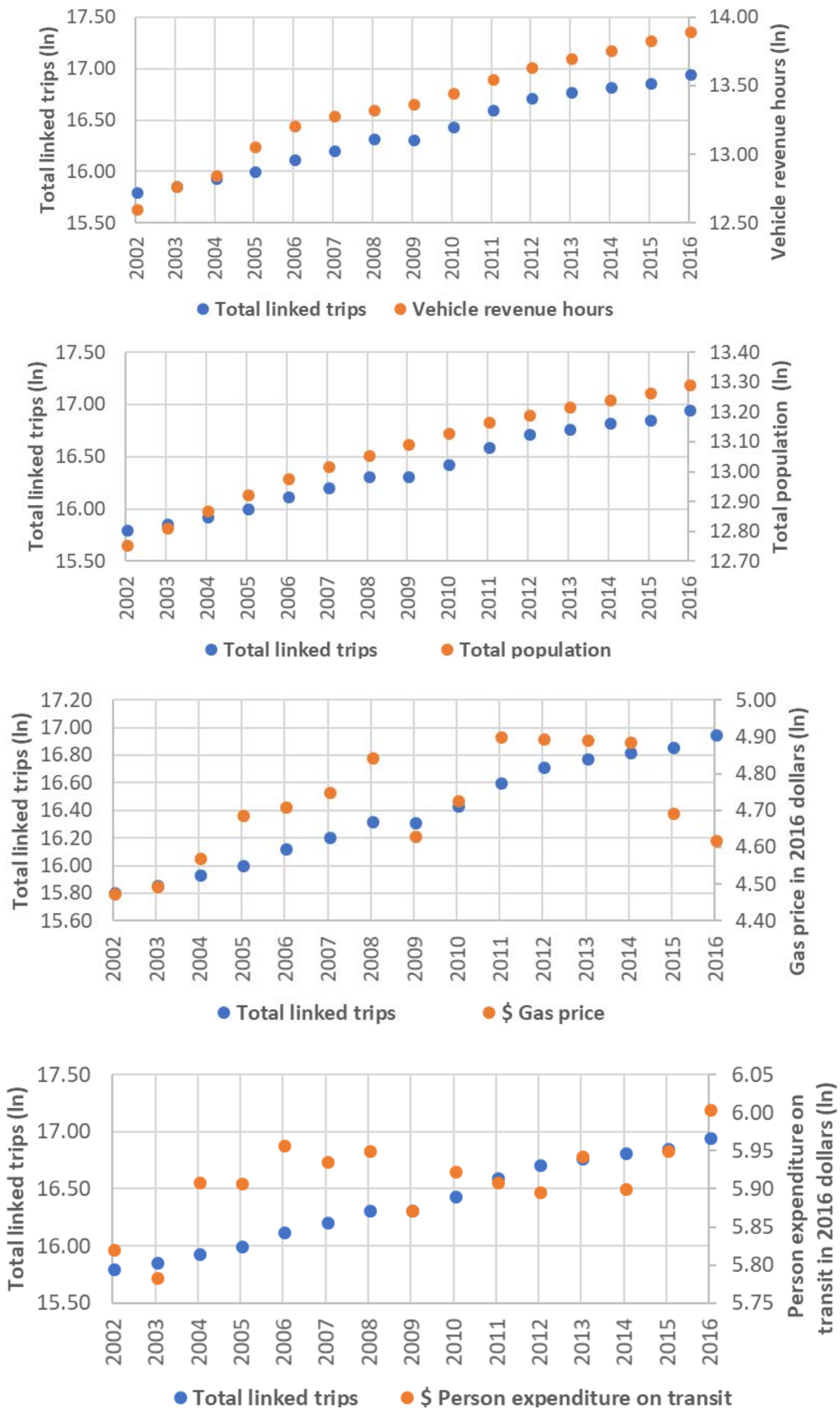


Figure 3-3: Brampton Transit trends of ridership in relationship to: A. vehicle revenue hours, B. Total population, C. Gas price, and D. Person expenditure on transit.

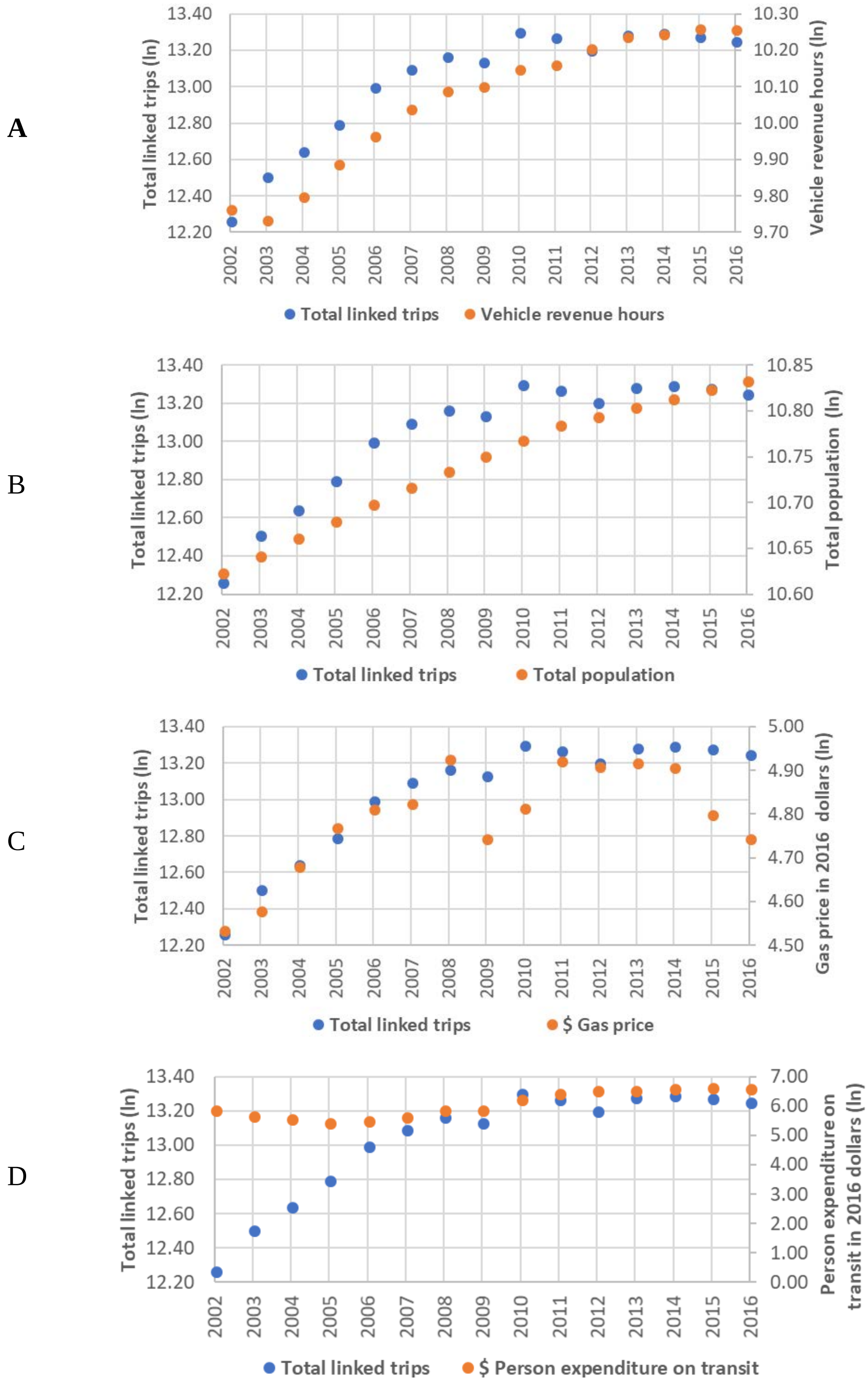


Figure 3-4: Comox Valley Transit trends of ridership in relationship to: A. vehicle revenue hours, B. Total population, C. Gas price, and D. Person expenditure on transit.

4. MODEL RESULTS

Three two-stage least squares (2SLS) models were estimated to account for the simultaneity of service supply and consumption while modelling determinants of transit use in Canada between 2002 and 2016. As discussed earlier, 2SLS models include two stages. Table 4-1 shows the results of the first-stage models, which use the log of the total vehicle revenue hours as the dependent variable. More specifically, Table 4-1.A presents the results of the general model that includes all CUTA member transit agencies in Canada, while Table 4-1.B. and Table 4.1.C present the results of the first group (>1.2 million linked trips) and second group (< 1.2 million linked trips) of transit agencies, respectively. A variety of models to predict the total vehicle revenue hours were initially tested and a simple two-variable model for the first stage was selected for each dataset. As seen in Table 4-1, all the models explain well the variation in the log of the total vehicle revenue hours. For example, the general model contains 1172 records and explains 97% of the variation in the log of the total vehicle revenue hours. This proportion of explained variance is considered relatively high in comparison with first-stage models presented in the literature (Taylor et al., 2009). The models include only two independent variables, namely the total population and total direct operating expenses. The logic here is that greater vehicle revenue hours are expected to be delivered by transit agencies serving larger populations and enjoying higher operating budgets. As seen in Table 4-1.A, for every 10% increase in the service total operating budget, a 5.50% increase in the vehicle revenue hours is expected. Similarly, for every 10% increase in the total transit agency service area population, a 4.70% increase in the vehicle revenue hours is expected. The other models show similar results in terms of the sign and direction with slightly different magnitudes, particularly for the third model (i.e., transit agencies with < 1.2 million linked trips). The predicted vehicle revenue hours variable from the first-stage model was then used as an instrumental variable in the second-stage model to predict transit ridership.

Table 4-1: First stage – estimation of vehicle revenue hours

	A. General model		**B. First group of transit agencies model**		**C. Second group of transit agencies model**	
	Coef.	z	Coef.	z	Coef.	z
Total operating budget	**0.55**	**35.36*****	**0.47**	**26.37*****	**0.56**	**25.17*****
Total population	**0.47**	**15.87*****	**0.46**	**12.81*****	**0.25**	**3.92*****
Constant	**-2.84**	**-13.65*****	**-1.39**	**-4.94****	**-0.90**	**-1.60**
N	**1172**		**662**		**485**	
Overall R-Squared	**0.973**		**0.967**		**0.854**	
Wald chi2	**(2) 6389.3**		**(2) 3553.2**		**(2) 998.4**	

*** Significant at 99% ** Significant at 95% * Significant at 90%

Table 4-2.A presents the results of the second-stage model predicting transit ridership for all transit agencies, while Table 4-2.B and Table 4-2.C present the results of the second-stage models for the first and second group of transit agencies, respectively. The dependent variable in these models is the total number of linked trips by transit agency per year. The z-value and the statistical significance are reported in the table along with the independent variable coefficients. The first model (i.e., the general model) is based on 1139 observations and explains 97% of the overall variation in ridership. This includes the variation between the transit agencies compared to each other as well as variations within transit agencies over time. The second and third models contain 653 and 485 observations and explain 96% and 73% of the overall variation in ridership for the first and second group of transit agencies, respectively. It should be noted that the number of observations in this table is different from the previous table due to missing values of variables used in different models.

As seen in Table 4-2.A, for the general model, the *Predicted vehicle revenue hours* variable, which accounts for the transit supply, has a positive and statistically significant association with ridership. More specifically, for every 10% increase in the predicted vehicle revenue hours a 10.0% increase in ridership is expected. By observing the other variables in the model, this variable is the largest contributor to transit ridership, which is consistent with previous research (Boisjoly et al., 2018; Lee & Lee, 2013; Taylor et al., 2009). It should be noted that this variable was predicted based on the total population (i.e., built environment factor) and total direct operating budget (i.e., transit supply factor). Additionally, other built environment factors are also important determinants of transit ridership. More specifically, transit usage is positively associated with apartments and row house dwellings, while negatively associated with single-family dwellings. As seen, a 10% increase in the numbers of apartments and row houses are associated with a 5.04% and 2.89% increase in ridership, respectively. In contrast, a 10% increase in the number of single-family households is associated with a 3.42% decrease in transit ridership, while keeping all other variables constant at their mean values. This finding is reasonable as single-family houses are associated with lower population densities and less mixed land-use patterns, which are associated with a lower level of transit supply, which in turn discourages users from using the service.

Transit ridership is positively associated with the number of local businesses/recreation opportunities within the transit agency's service area. A 10% increase in the number of opportunities is linked with a 1.22% increase in ridership. This means that service areas with more businesses/recreation opportunities are more likely to enjoy higher ridership rates, which is consistent with previous research findings (Chiou, Jou, & Yang, 2015). Transit usage is also positively associated with the percentage of postsecondary students within the transit agency's service area. More specifically, a 10% increase in percentage of postsecondary students is associated with less than 1.17% increase in ridership. Conversely, person expenditure on public transit, which represents the total amount of money spent on transit fares per household divided by the number of persons in the household, is significantly associated with a decrease in ridership, where a 10% increase in household expenditure on public transit is linked with a 1.43% decrease in ridership. The number of employed work force commuting to another Census Subdivision (CSD) is negatively associated with ridership. In other words, for every 10% increase in the number of workers commuting to a CSD that is different from their home origins' CSD, a 1.43% decrease in ridership is expected. This means that this group of users is more likely to use transit less than other groups of users who live and work in the same CSD. Gas prices have a positive and statistically significant relationship with transit service ridership, which is consistent with several studies in the literature (Guerra & Cervero, 2011; Lane, 2010; Lee & Lee, 2013; Litman, 2005). The model suggests that for every 10% increase in gas prices a 1.44% increase in transit ridership is expected. The model included a dummy variable to distinguish the transit agencies that utilized smart card automated fare collection systems to control the impact of data source quality on ridership estimation. As seen in the model, as expected, transit agencies that use AFC systems observe less ridership by 0.04% compared with time period before using these systems for the same transit agencies and other transit agencies that are not using these systems. This confirms the transit agencies' observations elicited in the survey. Finally, the model included a dummy variable to distinguish the transit agencies that have a bike-sharing system within their service area. The model suggests that bike-sharing systems have a negative association with transit usage. This association, however, is very limited in terms of magnitude: the introduction/presence of bike- sharing system within transit agencies service areas is associated with 0.07% decrease in transit service ridership. It should be noted that bike-sharing system are only located within six transit service areas (i.e., six cities), with a total records of about 30 records in our models. Therefore, while this variable's results are based on theoretically sound number of records, they should be treated with caution since they are presenting only few (and specific) systems.

Table 4-2.B shows the results of the second-stage model for the first group of transit agencies (with more than 1.2 million linked trips in 2016). Most of the model coefficients follow the same signs and magnitude as the previous model, with only a few exceptions. More specifically, the predicted vehicle revenue hours has the largest association (in terms of magnitude) with ridership. Transit usage is positively associated with the percentage of row house dwellings, number of businesses/recreation opportunities, and gas prices, while it is negatively linked with the percentage of single-family dwellings, person expenditure on public transit, the percentage of workers commuting to another CSD and the presence of bike-sharing systems. Interestingly, in this model, the percentage of apartment dwellings is negatively associated with transit usage. This may reflect the context of large cities in Canada where different type of apartment dwellings (i.e., condos) can exist in different urban contexts.

Regarding the new variables incorporated in the model, the number of rented dwelling units has a positive and statistically significant association with ridership, where a 10% increase in the number of rented dwelling units is linked with a 1.47% increase in ridership. The percentage of telecommuting population (i.e., people working from home) has also a negative association with transit usage. A 10% increase in the percentage of telecommuting population is associated with a 1.27decrease in ridership, respectively. Conversely, transit usage is positively associated with the percentage of seniors within transit agency's service area, where a 10% increase in the percentage of seniors is associated with 1.23% increase in ridership.

The presence of Uber systems within a transit agency's service areas is positively associated with transit ridership. This shows that in the context of large cities in Canada, Uber service plays a complementary role, helping transit service to gain more ridership, while keeping all other variables constant at their mean values (Boisjoly et al., 2018). This association, however, is very limited in terms of magnitude: the presence/introduction of Uber system is associated with 0.05% increase in ridership for the large transit agencies included in the model. Multiple factors could be causing this result. For example, some people might use transit for some trip purposes (e.g. shopping) while relying on using Uber in the return trip. Nevertheless, it should be stressed that the positive contribution of Uber found in this study was tested at an aggregate level using dummy variables in the model. Therefore, a better understanding of the detailed drivers behind this association at the disaggregate level is still needed, which could be done using more detailed data (e.g., number of trips made with Uber) which are not available for the study team.

Table 4-2.C presents the results of the second-stage model for the second group of transit agencies (with less than 1.2 million linked trips in 2016). As seen in the table, a fewer number of variables have a significant relationship with transit usage, with relatively lower overall R-squared compared to the other two models. This highlights a challenge in modelling ridership at these locations, where more context-specific issues can exist. Nevertheless, regarding the model results, as expected, most of the model coefficients follow the same signs and magnitude as the previous models. The predicted vehicle revenue hours variable has the largest association with ridership. Transit usage is positively associated with the percentage of row house and apartment dwellings, while it is negatively linked with the person expenditure on public transit and the percentage of workers commuting to another CSD. In the context of small transit agencies (or cities), the percentage of immigrants at the aggregate level has a positive and statistically significant association with ridership. More specifically, for every 10% increase in the percentage of immigrants, a 1.09% increase in ridership is expected. This highlights that immigrant population at these locations presents an important source of transit ridership. Conversely, the presence of Uber systems within these locations is negatively associated with transit ridership. More specifically, the introduction/presence of Uber system within small transit agencies service areas is associated with a 1.48% decrease in transit service ridership. This highlights the attractiveness of these ride-sharing system at these locations, which are normally characterized by poor transit service (e.g., long headways). Also, due to the residents' short trip distances in these contexts, users can travel across the city using ride-sharing services within a very reasonable and competitive cost. Taking the previous model results into account, this mixed impacts of ride-sharing system on transit usage was suggested before in the literature (Martin & Shaheen, 2011). Using a North American car sharing survey data, Martin and Shaheen (2011) indicated that some travellers decrease their use of transit as a result of using car sharing systems, while others increase their use of transit. Other variables were tested in our model and had a significant impact but were removed from the model due to the small number of records. For example, the service disruption dummy variable, which distinguished service disruptions (e.g., due to operator and union strikes) that exceeded three weeks, had a negative and significant impact on ridership. However, these results were based on less than 10 records.

Table 4-2: Second stage – estimation of the influential factors on transit ridership

	A. General model		B. First group of transit agencies model		C. Second group of transit agencies model	
	Coef.	z	Coef.	z	Coef.	z
Predicted vehicle revenue hours	**1.009**	**31.59 *****	**0.890**	**19.42 *****	**1.044**	**18.55 *****
% of apartments dwellings	**0.504**	**5.650 *****	**-0.404**	**-4.250 *****	**0.535**	**4.180 *****
% of row house dwellings	**0.289**	**4.910 *****	**0.113**	**1.900 ***	**0.591**	**5.870 *****
% of single family dwellings	**-0.342**	**-2.920 *****	**-0.960**	**-6.180 *****	**-0.341**	**-2.010 ****
Number of local opportunities	**0.122**	**2.460 ****	**0.295**	**4.600 *****		
% Rented dwellings			**0.147**	**2.730 *****		
% of population working from home			**-0.127**	**-1.710 ***		
% of population senior			**0.123**	**2.180 ****		
% of population postgraduate students	**0.117**	**2.940 *****				
% of population recent immigrant					**0.109**	**1.790 ***
% of people work outside CSD of residence	**-0.071**	**-2.130 ****	**-0.046**	**-1.650 ***	**-0.124**	**-1.660 ***
$ Person expenditure on public transit	**-0.143**	**-4.870 *****	**-0.147**	**-4.640 *****	**-0.162**	**-4.030 *****
$ Gas price	**0.144**	**3.500 *****	**0.222**	**4.520 *****		
Presence of Uber			**0.049**	**2.170 ****	**-0.148**	**-2.650 *****
Presence of bikesharing systems	**-0.066**	**-1.740 ***	**-0.057**	**-1.650 ***		
Constant	1.095	1.360	**6.302**	**6.050 *****	**2.109**	**1.740 ***
N	**1139**		**653**		**485**	
Overal R Square	**0.969**		**0.963**		**0.734**	
Wald chi2	**(11) 5469.3**		**(14) 3117.6**		**(8) 959.9**	

*** Significant at 99% ** Significant at 95% * Significant at 90%

5. MODEL VALIDATION

As discussed earlier, for each model, the original dataset was split into a training sample (90%) and a validation sample (10%) at the level of the transit agency. In other words, 10% of the transit agencies were randomly removed from the dataset used for estimating each model. Using the developed models, it was possible to estimate the total number of linked trips for each transit agency in the validation sample for each year and compare the results with the actual number of linked trips. Table 5-1 shows the predicted and actual number of linked trips for each model and their correlation.

As seen in Table 5-1, in the first model, the average log of the actual number of linked trips was 13.73, while the average log of the estimated number of linked trips was 13.54. This indicates a close relationship between both the actual and estimated number of linked trips. Furthermore, the standard deviation of the log of the actual number of linked trips was 1.73, while it was 1.75 for the estimated number of linked trips. This indicates very close similarity in the variation of the actual and estimated number of trips from the mean. Using a Pearson correlation test, a statistically significant positive correlation of 0.935 between the actual and estimated linked trips was detected, implying a very strong relationship between the two values. Very similar results were found for the two other models.

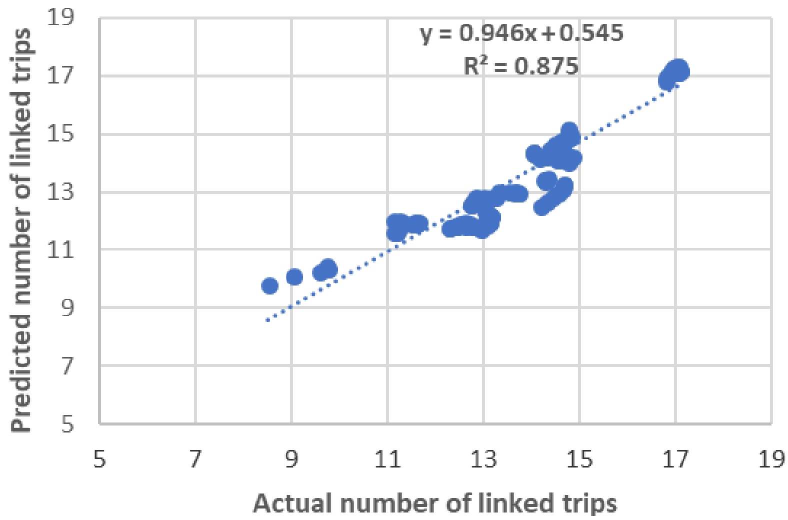
Table 5-1: Predicted and actual number of linked trips for each model

	A. General model		**B. First group of transit agencies model**		**C. Second group of transit agencies model**	
	Actual	Predicted	Actual	Predicted	Actual	Predicted
Average	13.73	13.54	15.84	15.89	12.74	12.73
Standard deviation	1.73	1.75	1.64	1.73	0.70	0.64
Pearson Correlation	**0.935***		**0.985***		**0.798***	
Number of observations	**140**		**75**		**67**	

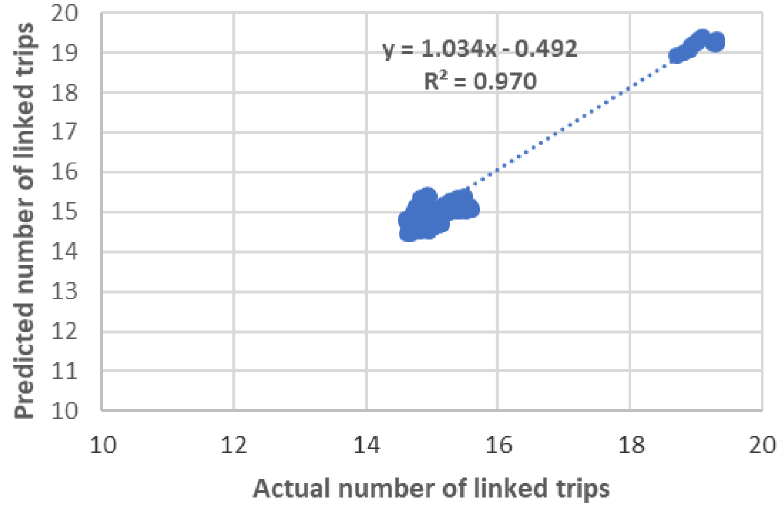
*** Significant at 99.9% ** Significant at 95% * Significant at 90%

Figure 5-1 shows the estimated and actual number of linked trips for the three randomly selected validation samples. The figure also shows the linear function of the relationship between the estimated and actual number of linked trips and its R-squared value. As we see in the figures, for first two models the slope of the linear function is between 0.95 and 1.03, indicating that for every 10% increase in the actual number of linked trips, a 9.5–10.3% increase in the estimated ridership is expected. This indicates a very strong correlation between the actual and estimated values. The third model shows a lower level of performance, with a 0.72 slope of linear function. This means that less number of trips could be explained directly by the slope, without considering the slope's intercept value (i.e., 2.109), which highlights a challenge in modelling ridership at these locations, where more context-specific issues can exist.

A



B



C

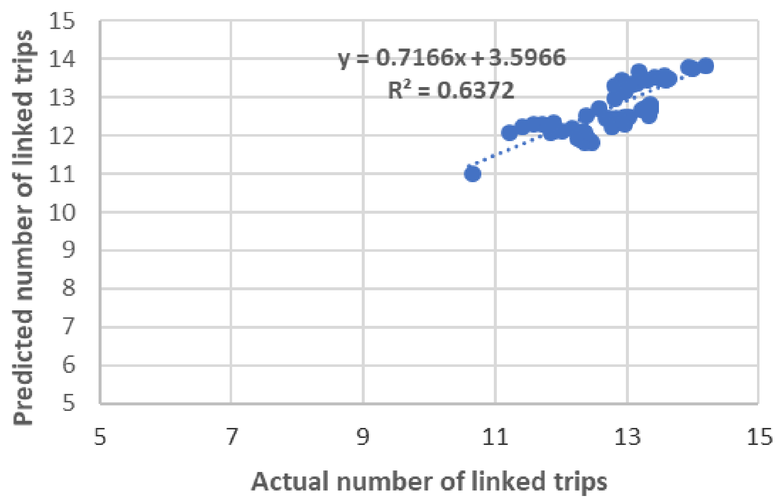


Figure 5-1: Actual and predicted number of linked trips for: A. All transit agencies, B. Transit agencies with more than 1.2 million linked trips in 2016, and C. Transit agencies with less than 1.2 million linked trips in 2016.

6. CONCLUSIONS AND POLICY IMPLICATIONS

The goal of this chapter is to provide an understanding of the relationships between key factors and transit ridership in Canada, as well as to provide an explanation of recent observed ridership trends (discussed in Part IV of this document). In order to do so, a comprehensive long-term spatio-temporal dataset was compiled from various sources. Each transit agency in the dataset (approx. 100 agencies in total) had multiple observations at annual intervals over the study period between 2002 and 2016. This allowed for the investigation of the variation in transit ridership in relation to variation in different influential factors across agencies and also changes in these factors over time, in contrast to the cross-sectional analyses of transit ridership commonly used in previous studies. The study team gathered data on four major sets of indicators: a) built environment attributes, b) socioeconomic factors, c) transit service factors and d) other external/contextual factors. These factors were tested for inclusion in three different models. The first model was estimated for all CUTA member transit agencies in Canada. The second and third models were estimated for transit agencies with more than or equal to 1.2 million annual linked trips and less than 1.2 million annual linked trips in 2016, respectively. Estimating these two models assisted in investigating the difference between a number of key factors and their contribution to transit ridership for larger and smaller transit agencies. To account for the causal and two-directional relationship between transit service supply and consumption, a two-stage least squares (2SLS) modelling approach using instrumental variables was implemented. Finally, model validation was performed by comparing the actual ridership with the predicted ones. The model validation indicated a strong and significant relationship between the model-predicted ridership and actual ridership. The overall findings of this chapter can be summarized as follows:

- Transit ridership (in terms of number of linked trips per year) in Canada is associated with several factors that differ according to the transit agency size. Nevertheless, internal transit service factors, namely, revenue vehicle hours and personal expenditure on public transit (which can be used as a proxy for transit fares), are among the main contributors to ridership across all transit agencies. This suggests that transit agencies and municipalities can improve their ridership by investing in improving the transit service as well as by reducing the associated cost of using transit (i.e., fares). Currently, several transit agencies are planning to reduce transit cost by implementing Universal Transit Pass (U-Pass) for students and by integrating fare systems (see Part III of this document). A more efficient strategy could be the introduction/increase of public transit subsidies, generated from toll roads and parking, thus charging the real cost of using private motor vehicles as an alternative to transit.
- Built environment factors, in terms of household type, number of businesses/recreation opportunities within the transit service area, and number of employers/workers commuting to a different Census Subdivision (CSD) outside their CSD of residence are

also strong and important determinants of transit ridership. As such, policies influencing the built environment can play an important role in maintaining/improving transit ridership. For example, municipal policies discouraging the construction of single family dwellings in large cities, while promoting row-house and apartment dwellings in small cities can have a positive effect on transit ridership. In addition, policies that encourage providing more suitable job opportunities within the residents' CSD of residence can be also recommended. The model also indicated that number of rented dwelling units has a positive and statistically significant association with ridership. Theretofore, policies that support providing more rental dwelling units can be recommended in order to increase ridership.

- Socioeconomic factors also play an important role in determining transit ridership. Generally, the percentages of telecommuting population (i.e., people working from home) have a negative association with transit usage. In contrast, the percentages of postsecondary students, seniors, and recent immigrants have a positive association with transit usage. Also, while taking into account that the influence of each factor varies according to scale (national level, large size and small size cities), policies targeting specific socioeconomic groups may be introduced in order to improve ridership levels.
- As expected, economic factors, in terms of gas prices, are also associated with transit ridership. More specifically, increases in gas prices contribute positively to increasing transit ridership. With the improvements in private automobiles fuel efficiency (which reduces the amount of money spent on gas), higher increases in gas and carbon taxes (e.g., to fund the public transit service) would help encourage higher transit ridership across Canada.
- Other external factors were also found to have a significant association with transit ridership. More specifically, the models suggest that bike-sharing systems have a negative association with transit usage. This association, however, is very limited in terms of magnitude. On the other hand, ride sharing systems (i.e., Uber service) have a mixed impact on ridership that varies according to the size of transit system. The presence of Uber within large transit service areas is shown to be positively associated with transit ridership, while keeping all other variables constant at their mean values. This finding indicates that in such contexts, Uber provides many residents with a transportation option that complements, rather than competes with, the use of public transit. It should be emphasized that this effect, though significant, is marginal. Conversely, the presence of Uber within small transit service areas is shown to be negatively associated with transit ridership. This highlights the need for formulating a new set of policies that mitigate the negative impacts of transportation network companies either by integrating them with the existing transit system or by making transit more competitive.

Notwithstanding the findings outlined above, the study has some limitations which should be acknowledged. First, due to data limitations it was not possible to investigate the association between ridership and several operational factors. For example, longitudinal data about service quality and reliability as well as changes in user satisfaction levels were not available. In addition, no accurate information on the types and timing of service improvement strategies or changes in the network design were readily available. Therefore, this study could not investigate, for example, the effects of changes in network design, stop re-location/consolidation, density of dedicated bus lanes and transit preferential treatment (e.g., transit signal priority), user satisfaction with the service, or service reliability issues (e.g., headway adherence, on-time performance).

Second, the changes over time to the spatial boundaries of transit networks were not available. Therefore, in this study, each transit agency service area was distinguished by consulting the current transit system maps available on the agency's website. Third, some of the existing data fields include many records with zero or missing values, which hindered our ability to test these variables in the model (e.g., service span (hours of service), fare structure (e.g., child and student fares), funding (e.g., federal and municipal funding)). Fourth, the impact of Uber and bike-sharing services was tested at the aggregate level using dummy variables in the models. However, it will be beneficial to use data about the number and average length of trips made by these services in order to understand the impact of different degrees of penetrations.

Part VI

Analytical Tool for Policy Analysis

Table of contents

1. INTRODUCTION

This section provides a description of the policy analytical tools that have been developed in MS-Excel to demonstrate the associations between various influencing factors and ridership in Canada. The tools, delivered to CUTA with the final report, are based on the models reported in the previous chapter. The main objective of the tools is to facilitate an improved understanding of the relative contribution of each factor to ridership variability, while keeping all other variables constant at their mean values.

2. TOOL DESCRIPTION

The first tool, named the "General Tool", was developed based on the general model results, which included all CUTA member transit agencies in Canada. The tool could be used to examine the association between various influencing factors and ridership at the national level. The second tool, named "Large Transit Agencies Tool", was developed based on the first group of transit agencies model, which included transit agencies with more than or equal to 1.2 million annual linked trips in 2016. This tool could be used to predict ridership changes for large transit agencies in Canada. The last tool, named "Small Transit Agencies Tool", was based on the second group of transit agencies model, which included transit agencies with less than 1.2 million annual linked trips in 2016.

Each tool was developed in MS-Excel and included four sheets. The first sheet, named "Factors Definitions", provides a precise definition of each factor used in the tool. The second sheet, named "*Policy analysis tool 1st stage*", replicates the first-stage of the model task in terms of estimating the total vehicle revenue hours based on total population and total direct operating expenses. This sheet directly calculates and inputs the value of *Predicted vehicle revenue hour* variable in the third sheet. The third sheet, named "*Policy analysis tool 2nd stage*", replicates the second-stage of the model task of estimating and predicting ridership based on the values of various factors.

The latter two sheets (i.e., *Policy analysis tool 1st stag* and *Policy analysis tool 2nd stage)* contain each factor's mean value as well as an empty cell to allow users to add any value according to their case study. The mean value for each factor was only provided for benchmarking and to understand the national average (based on the used sample). For dummy variables, instead of providing the average values, a value of zero was added. The two sheets also include some instructions about how each factor can be added and calculated to make it easier for users. The fourth MS-Excel sheet of the tool illustrates the final results by using a column chart while comparing them to the model average.

3. TOOL BENEFITS AND USAGE

Overall, the tools present practical instruments to estimate the relative changes in ridership due to changes in an individual factor (e.g., gas prices), while keeping all other variables constant at their mean values. Each tool enables a better understanding of the changes in ridership at the transit agency level or/and the national level. Therefore, these tools help with providing a meaningful comparison between different transit agencies and regions. One of the possible applications of the tools is examining the potential change in ridership (at the aggregate agency level) associated with new policies and projects, considered individually or in combination. These changes in the system can be related to internal transit service factors (e.g., operating budget), socioeconomic factors (e.g., the percentages of telecommuting population and postsecondary students) and built environment factors (e.g., percentage of single-family dwellings) and other contextual factors (e.g., the introduction of Uber-like services).

References and Appendices

REFERENCES

Abdel-Aty, M. (2001). Using ordered probit modeling to study the effect of ATIS on transit ridership. *Transportation Research Part C: Emerging Technologies, 9*(4), 265-277. doi: https://doi.org/10.1016/S0968-090X(00)00037-1

Alam, B., Nixon, H., & Zhang, Q. (2015). Investigating the determining factors for transit travel demand by bus mode in US Metropolitan Statistical Areas: Mineta Transportation Institute.

Arana, P., Cabezudo, S., & Peñalba, M. (2014). Influence of weather conditions on transit ridership: A statistical study using data from Smartcards. *Transportation Research Part A: Policy and Practice, 59*, 1-12. doi: https://doi.org/10.1016/j.tra.2013.10.019

Bernal, M., Welch, E., & Sriraj, P. (2016). The effect of slow zones on ridership: An analysis of the Chicago Transit Authority "El" Blue Line. *Transportation Research Part A: Policy and Practice, 87*, 11-21. doi: https://doi.org/10.1016/j.tra.2016.02.007

Boisjoly, G., Grisé, E., Maguire, M., Veillette, M., Deboosere, R., Berrebi, E., & El-Geneidy, A. (2018). *Invest in the ride: A longitudinal analysis of the determinants of public transport ridership in 25 North America cities.* Paper presented at the The 97th Annual Meeting of the Transportation Research Board, Washington D.C., USA.

Brakewood, C., Macfarlane, G., & Watkins, K. (2015). The impact of real-time information on bus ridership in New York City. *Transportation Research Part C: Emerging Technologies, 53*, 59-75. doi: https://doi.org/10.1016/j.trc.2015.01.021

Brown, J., & Thompson, G. (2008). Service orientation, bus—rail service integration, and transit performance: Examination of 45 U.S. metropolitan areas. *Transportation Research Record: Journal of the Transportation Research Board, 2042*, 82-89. doi: 10.3141/2042-09

Campbell, K., & Brakewood, C. (2017). Sharing riders: How bikesharing impacts bus ridership in New York City. *Transportation Research Part A: Policy and Practice, 100*, 264-282. doi: https://doi.org/10.1016/j.tra.2017.04.017

Cervero, R., Murakami, J., & Miller, M. (2010). Direct ridership model of bus rapid transit in Los Angeles County, California. *Transportation Research Record: Journal of the Transportation Research Board, 2145*, 1-7. doi: 10.3141/2145-01

Chakour, V., & Eluru, N. (2016). Examining the influence of stop level infrastructure and built environment on bus ridership in Montreal. *Journal of Transport Geography, 51*, 205-217. doi: https://doi.org/10.1016/j.jtrangeo.2016.01.007

Chakraborty, A., & Mishra, S. (2013). Land use and transit ridership connections: Implications for state-level planning agencies. *Land Use Policy, 30*(1), 458-469. doi: https://doi.org/10.1016/j.landusepol.2012.04.017

Chava, J., Newman, P., & Tiwari, R. (2018). Gentrification of station areas and its impact on transit ridership. *Case Studies on Transport Policy*. doi: https://doi.org/10.1016/j.cstp.2018.01.007

Chen, C., Varley, D., & Chen, J. (2011). What affects transit ridership? A Dynamic analysis involving multiple factors, lags and asymmetric behaviour. *Urban Studies, 48*(9), 1893-1908. doi: 10.1177/0042098010379280

Chiang, W., Russell, R., & Urban, T. (2011). Forecasting ridership for a metropolitan transit authority. *Transportation Research Part A: Policy and Practice, 45*(7), 696-705. doi: https://doi.org/10.1016/j.tra.2011.04.016

Chiou, Y., Jou, R., & Yang, C. (2015). Factors affecting public transportation usage rate: Geographically weighted regression. *Transportation Research Part A: Policy and Practice, 78*, 161-177. doi: https://doi.org/10.1016/j.tra.2015.05.016

City of Edmonton. (2016). Factors affecting transit ridership.

Clewlow, R. , & Mishra, G. (2017). Disruptive transportation: the adoption, utilization, and impacts of ride-hailing in the United States: Research Report–UCD-ITS-RR-17.

Crowley, D. (2000). STRP Synthesis S1: Profiling Transit Ridership. Toronto, Ontario: Canadian Urban Transit Association.

Currie, G., & Delbosc, A. (2011). Understanding bus rapid transit route ridership drivers: An empirical study of Australian BRT systems. *Transport Policy, 18*(5), 755-764. doi: https://doi.org/10.1016/j.tranpol.2011.03.003

Curry, B. (2017). Where have all the transit riders gone?, *The Globe and Mail*. Retrieved from https://www.theglobeandmail.com/news/politics/drop-in-transit-ridership-has-officials-across-canadastumped/article30178600/

CUTA. (2007). A profile of Canadian transit ridership.

Durning, M., & Townsend, C. (2015). Direct ridership model of rail rapid transit systems in Canada. *Transportation Research Record: Journal of the Transportation Research Board, 2537*, 96-102. doi: 10.3141/2537-11

Feigon, S., & Murphy, C. (2018). Broadening understanding of the interplay between public transit, shared mobility, and personal automobiles. Pre-publication draft of TCRP Research Report 195. Washington, D.C.: Transportation Research Board.

Guerra, E., & Cervero, R. (2011). Cost of a ride:The effects of densities on fixed-guideway transit ridership and costs. *Journal of the American Planning Association, 77*(3), 267-290. doi: 10.1080/01944363.2011.589767

Gutiérrez, J., Cardozo, O., & García-Palomares, J. (2011). Transit ridership forecasting at station level: an approach based on distance-decay weighted regression. *Journal of Transport Geography, 19*(6), 1081-1092. doi: https://doi.org/10.1016/j.jtrangeo.2011.05.004

Hemily, B. (2004). Trends affecting public transit's effectiveness, a review and proposed actions.

Jun, M., Choi, K., Jeong, J., Kwon, K., & Kim, H. (2015). Land use characteristics of subway catchment areas and their influence on subway ridership in Seoul. *Journal of Transport Geography, 48*, 30-40. doi: https://doi.org/10.1016/j.jtrangeo.2015.08.002

Jung, H., Yu, G., & Kwon, K. (2016). Investigating the effect of gasoline prices on transit ridership and unobserved heterogeneity. *Journal of Public Transportation, 19*(4), 4.

Kain, J., & Liu, Z. (1999). Secrets of success: Assessing the large increases in transit ridership achieved by Houston and San Diego transit providers. *Transportation Research Part A: Policy and Practice, 33*(7), 601-624. doi: https://doi.org/10.1016/S0965-8564(99)00009-9

Kohn, H. (2000). Factors affecting urban transit ridership: Statistics Canada, Transportation Division.

Kuby, M., Barranda, A., & Upchurch, C. (2004). Factors influencing light-rail station boardings in the United States. *Transportation Research Part A: Policy and Practice, 38*(3), 223-247. doi: https://doi.org/10.1016/j.tra.2003.10.006

Lane, B. (2010). The relationship between recent gasoline price fluctuations and transit ridership in major US cities. *Journal of Transport Geography, 18*(2), 214-225. doi: https://doi.org/10.1016/j.jtrangeo.2009.04.002

Lee, B., & Lee, Y. (2013). Complementary pricing and land use policies: Does it lead to higher transit use? *Journal of the American Planning Association, 79*(4), 314-328. doi: 10.1080/01944363.2014.915629

Lee, K., Eom, J., You, S., Min, J., & Yang, K. (2015). An empirical study on the relationship between urban railway ridership and socio-economic characteristics. *Procedia Computer Science, 52*, 106-112. doi: https://doi.org/10.1016/j.procs.2015.05.037

Litman, T. (2005). Appropriate response to rising gasoline prices: Citizens should demand,“Raise my prices now!” Victoria Transport Policy Institute, accessed via website.

Litman, T. (2017). Transit price elasticities and cross-elasticities: Victoria Transport Policy Institute.

Manville, M., Taylor, B., & Blumenberg, E. (2018). Falling transit ridership: California and Southern California: UCLA Institute of Transportation Studies.

Martin, E., & Shaheen, S. (2011). The impact of carsharing on public transit and non-motorized travel: an exploration of North American carsharing survey data. *Energies, 4*(11), 2094-2114.

Mees, P., Stone, J., Imran, M., & Nielson, G. (2010). Public transport network planning: A guide to best practice in NZ cities *Research report 396*. New Zealand: New Zealand Transport Agency.

Miller, C., & Savage, I. (2017). Does the demand response to transit fare increases vary by income? *Transport Policy, 55*, 79-86. doi: https://doi.org/10.1016/j.tranpol.2017.01.006

NACTO. (2018). Case study: Metro bus network redesign, Houston.   Retrieved February 28, 2018, from https://nacto.org/case-study/metro-bus-network-redesign-houston/

Rojo, M., Gonzalo-Orden, H., dell’Olio, L., & Ibeas, Á. (2012). Relationship between service quality and demand for inter-urban buses. *Transportation Research Part A: Policy and Practice, 46*(10), 1716-1729. doi: https://doi.org/10.1016/j.tra.2012.07.006

Schaller, B. (2017). Unsustainable? The growth of app-based ride services and traffic, travel and the future of New York city: Schaller Consulting.

Shalaby, A. , Woo, K., & Currie, G. . (2010). *Exploring transit ridership drivers in Toronto and Melbourne.* Paper presented at the Transportation Research Board (USA) Annual Meeting, Washington DC, United States.

Singhal, A., Kamga, C., & Yazici, A. (2014). Impact of weather on urban transit ridership. *Transportation Research Part A: Policy and Practice*, *69*, 379-391. doi: https://doi.org/10.1016/j.tra.2014.09.008

Smith, A. (2016). Shared, collaborative and on demand: The new digital economy: Pew Research Center.

Storchmann, K. (2001). The impact of fuel taxes on public transport — an empirical assessment for Germany. *Transport Policy, 8*(1), 19-28. doi: https://doi.org/10.1016/S0967-070X(00)00033-0

Sung, H., & Oh, J. (2011). Transit-oriented development in a high-density city: Identifying its association with transit ridership in Seoul, Korea. *Cities*, *28*(1), 70-82. doi: https://doi.org/10.1016/j.cities.2010.09.004

Tang, L., & Thakuriah, P. (2012). Ridership effects of real-time bus information system: A case study in the City of Chicago. *Transportation Research Part C: Emerging Technologies, 22,* 146-161. doi: https://doi.org/10.1016/j.trc.2012.01.001

Tao, S., Corcoran, J., Rowe, F., & Hickman, M. (2018). To travel or not to travel: 'Weather' is the question. Modelling the effect of local weather conditions on bus ridership. *Transportation Research Part C: Emerging Technologies*, *86*, 147-167. doi: https://doi.org/10.1016/j.trc.2017.11.005

Taylor, B., & Fink, C. (2013). Explaining transit ridership: What has the evidence shown? *Transportation Letters, 5*(1), 15-26. doi: 10.1179/1942786712Z.0000000003

Taylor, B., Haas, P., Boyd, B., Hess, D., Iseki, H., & Yoh, A. (2002). Increasing transit ridership: Lessons from the most successful transit systems in the 1990s: Mineta Transportation Institute.

Taylor, B., Miller, D., Iseki, ., & Fink, C. (2009). Nature and/or nurture? Analyzing the determinants of transit ridership across US urbanized areas. *Transportation Research Part A: Policy and Practice, 43*(1), 60-77.

TCRP. (2007). Elements Needed to Create High Ridership Transit Systems (Vol. 66): Transportation Research Board.

TCRP. (2016). Shared mobility and the transformation of public transit (TCRP research report 188).

Thompson, G., & Brown, J. (2006). Explaining variation in transit ridership in U.S. metropolitan areas between 1990 and 2000: Multivariate Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, *1986*, 172-181. doi: 10.3141/1986-23

Thompson, G., & Brown, J. (2012). Making a successful LRT-based regional transit system: Lessons from five new start cities. *Journal of Public Transportation, 15*(2), 8.

Thompson, G., Brown, J., Sharma, R., & Scheib, S. (2006). Where transit use is growing: surprising results. *Journal of Public Transportation, 9*(2), 2.

TRID. (2013). Home page. Retrieved 15 November, 2013, 2013, from http://trid.trb.org/

Wang, K., & Woo, M. (2017). The relationship between transit rich neighborhoods and transit ridership: Evidence from the decentralization of poverty. *Applied Geography, 86*, 183-196. doi: https://doi.org/10.1016/j.apgeog.2017.07.004

York Region Transit. (2017). Ridership Growth Study.

Appendix A: Review of the academic literature

Study	Sample	Investigated factors	Significant factors	Analysis methods	Key findings/recommendations
City level and multi-city studies					
Boisjoly et al. (2018)	• 25 transit authorities in 2015 in US and Canada • 2002 - 2015	• Y = unlinked passenger trips • Xs = - vehicle revenue kilometers - average fare - external variables: (population, area, proportion of carless households , unemployment rate, GDP per capita, gas price, highway mileage, presence of private bus operator/Uber/bicycle sharing system)	• Vehicle revenue km (+) • Average fare (-) • Presence of private bus operator (+) • Proportion of carless households (+) • Gas price (+) • Population (+) • Area (-)	• Aggregated (transit agency level) • Longitudinal multilevel mixed-effect regression	• In addition to the characteristics of the metropolitan area, internal factors (VRK and average fares) as well as car ownership are the main contributors of ridership. • Vehicle revenue kilometers is by far the largest contributor to ridership.
Durning and Townsend (2015)	• 342 stations • 2012 data • Five Canadian cities (Montreal, Toronto, Calgary, Edmonton, and Vancouver)	• Y = average daily boardings • Xs = 44 variables - socioeconomics (unemployed (%), median household income ($), renter households (%), age) - station attributes (bus connections, park-and-ride spaces, terminal station, transfer station, distance to terminus, bike parking dummy, car share dummy) - built environment attributes (population density, job density, total links, nodes, road length, street density, average block length, intersection density, open area, park area, residential area, dwelling density, industrial area, government–institutional area, commercial area, CBD and university dummy, land use mix, land use entropy, walkability index, commercial site density) - service attributes (peak only fare, pass cost, fare)	• Population density (+) • Intersection density (+) • Street density (-) • bus connections (+) • Parking spaces (+) • Transfer dummy (+) • Peak only service (-) • Commercial site density (+) • Residential ratio (+) • Commercial ratio (+) • Government–institutional ratio (+)	• Aggregated (rail station level) • Regression analysis	• Densities, land uses, and station amenities have a statistically significant association with station ridership. • Socioeconomic factors do not appear to be significant.

Study	Sample	Investigated factors	Significant factors	Analysis methods	Key findings/recommendations
Lee and Lee (2013)	• 67 urbanized areas • 2002 to 2010 data	• Y = Log of monthly number of unlinked trips • Xs = - transit system factors (transit service supply, public subsidy, transit fare) - external factors (population, population density, gasoline price, compactness index, containment policy dummy, college students share, unemployment rate, freeway lane miles, trendPost-peak dummy, month dummies) - Interactions (population density × gas price; compactness index × gas price, etc.)	• Population/population density (+) • Gasoline price (+) • Interactions (+/-) • Fare (-) • Freeway lane miles (+) • Percent college students (+) • Unemployment rate (-) • Trend (-) • Months (+/-)	• Aggregated (route and vehicle levels) • Two-stage least squares (2 SLS) regression analysis	• Gas pricing schemes will be more effective where alternatives to automobility and supportive land use policies exist. • The impacts of urban form on travel behavior are also strengthened when driving externalities are correctly priced.
Currie & Delbosc (2011)	• 77 bus routes • 2008 data • Four Australian cities (Melbourne, Brisbane, Adelaide, Sydney)	• Y = boardings per route km/boardings per vehicle km • Xs = - transit system factors (vehicle trips per annum, weekday frequency (buses/h), low-floor buses (%), weekday service span (h), average peak speed, share separate right of way (%), stop spacing) - external factors (residential density, employment density, car ownership, location dummy variables)	• Weekday frequency (+/ns) • Average peak speed (-) • Low-floor buses (+) • Share separate right of way (+) • Car ownership (-) • location dummy variables (+/-)	• Aggregated (route and vehicle levels) • Regression analysis	• Service level factors have a considerable impact on boardings per route. • Some BRT infrastructure treatments such as right of way have a significant impact on ridership but the influence of infrastructure is within the context of high service levels.
Chiang et al. (2011)	• 20 bus routes • 1998 to 2008 data • Tulsa, US	• Y = number of passengers per month • Xs = - fare and transit operating funds - gas price and number of individuals receiving food stamps - months of the year (dummy variables)	• Fare (-) • Gas price (+) • Food stamps (-) • Transit operating funds (-) • Months of the year (+/-)	• Aggregated (route level) • Regression analysis, artificial neural-network and univariate time-series model	• A simple combination of three forecasting methodologies (regression analysis, neural networks, and ARIMA models) yielded greater forecast accuracy. • Operating funds and price of gas has a significant and positive impact on ridership.

Study	Sample	Investigated factors	Significant factors	Analysis methods	Key findings/recommendations
Guerra & Cervero (2011)	• 50 fixed-guideway transit projects in 23 transit systems • 2000 and 2008 data	• Y = passenger miles traveled • Xs = - external facotrs (population/Jobs within 1/2 mile of stations, population/Jobs within 5 mile of stations, metropolitan economic growth, gas price) - park & ride spots and number of bus routes, fare, average speed and frequency - new projects dummy variable	• Jobs (+) • Population (+) • Fare (-) • Gas price (+)	• Aggregated (project level) • Regression analysis, fixed effect and random effects estimators, two stage least squares model	• Controlling for neighborhood, regional, and transit service attributes, population and job density are positively correlated with both ridership and capital costs.
Lane (2010)	• January 2002 to April 2008 • Nine US cities	• Y = monthly unlinked passenger trips • Xs = - average and standard deviation of monthly gas price - vehicle revenue miles - vehicles operated in maximum service - seasons dummy variables - time (continuous variable)	• Average monthly gas price (+) • Standard deviation of gas price (-/+) • Vehicle revenue miles by mode (+) • Vehicles operated in maximum service (+) • Seasons dummy variables (-/+) • Time (-/ns)	• Regression analysis • Ridership analysis	• A small but statistically significant amount of ridership fluctuation is due to changes in gasoline prices.
Taylor & Miller (2009)	• 265 urbanized areas (UZAs) • 2000 • US	• Ys = i) total urbanized area ridership ii) relative (per capita) ridership • Xs = - regional geography (area of urbanization, population/population density, regional location in the US) - metropolitan economy (personal/household income, unemployment) - population characteristics (% college students, % population in poverty, % population recent immigrants, political party affiliations, racial/ethnic composition) - auto/highway system (freeway lane miles, fuel prices, non-transit/non-SOV trips, % carless household, total lane miles of roads, vehicle miles per capita) - transit system characteristics (total revenue vehicle hours, dominance of single transit	• Area of urbanization (+) • Population density (+) • UZA in the South (-) • Personal/household income (+) • % college students (+) • % population in poverty (+) • % population recent immigrants (+) • % democratic voters (+) • % African-American (-) • No transit/non-SOV trips (+) • % carless households (+) • Freeway lane miles (+) • Average gas price (+) • Predicted transit service levels (+) • Dominance of primary transit operator (+)	• Aggregated (system level) • Cross-sectional • Two-stage simultaneous equation regression models	• It is important to account for simultaneity between transit service supply and consumption. • Most of the variation in transit ridership between MSAs can be explained by factors outside the control of public transit systems. • Population characteristics were the most significant of the external factors. • Controlling for the fact that public transit use is strongly correlated with urbanized area size, about 26% of the observed variance in per capita transit patronage across US urbanized areas is explained by service frequency and fare levels.

Study	Sample	Investigated factors	Significant factors	Analysis methods	Key findings/recommendations
		operator, fare levels, headways/service frequency, predicted transit service levels, route coverage/density)	• Transit fares (-) • Headways/service frequency (+)		
Thompson & Brown (2006)	• 1990-2000 • United States (82 MSAs)	• Y = passenger miles per capita • Xs = 22 variables - internal factors (% change in service frequency, % change in service coverage, change in rail ratio, multidestination system layout) - external factors(% change in MSA density, west region, ~~%~~ change in MSA population, % change in unemployment rate, % change in Hispanic population, % change in African American, population share)	• % change in service Frequency (+) • % change in service Coverage (-) • West Region (+) • % of Routes that do not serve the CBD (+)	• Aggregated (MSA level) • Regression analysis	• Service coverage and frequency are the most powerful explanatory variables for variation in ridership change among MSAs with 1 million to 5 million people, whereas a multidestination service orientation is the most important explanation for variation in ridership change among MSAs with 500,000 to 1 million people.
Kuby & Barranda (2004)	• 268 stations in nine cities representing a variety of urban settings • 2000 • US	• Y = average weekday light-rail boarding • Xs = - external factors (Population, % employment/population, airport, international border, college enrollments, CBD, park-and-ride spaces, heating and cooling degree-days, accessibility,% PMSA employment, % renters) - internal factors (bus connections, stations connecting to other rail lines, terminal station, station spacing , designated transfer station)	• Employment (+) • Population (+) • % renters within walking distance (+) • Bus lines (+) • Park-and-ride spaces (+) • Centrality • Terminal and transfer stations (+) • Heating and cooling degree days (-)	• cross-sectional • multiple regression	• Land use and accessibility are significant factors in explaining LRT boardings. • More extreme temperatures discourage LRT ridership.
Kain & Liu (1999)	• 2 transit agencies • 1980 - 1990 • Houston and San Diego, US	• The choice of factors is based on an earlier work (Kain & Liu, 1995) which investigated ridership data from the 75 largest transit operators in US	• Central city population (+) • Metropolitan employment (+) • Bus and rail miles (+) • Fares (-)	• Cross section and time series ridership models	• The difference of ridership increase between the cities is due to a combination of different land use characteristics and transit policy decisions.

Study	Sample	Investigated factors	Significant factors	Analysis methods	Key findings/recommendations
Within-city studies					
Tao et al. (2018)	• Three-month period from 4th February to 28th April 2013 • Brisbane, Australia	• Y = hourly ridership • Xs = - five weather variables (temperature, rainfall, humidity, wind and apparent temperature) - Location cluster	• Weather variables (+ / -)	• Aggregated (destination-based and stop level) • Time-series modelling	• Hourly bus ridership on weekends was more affected by changing weather conditions than weekdays. • Weather impacts on bus ridership varied not only between weekdays and weekends, but also across trip destinations.
Miller & Savage (2017)	• 1043 observation collected in 2004, 2006, 2009 and 2013 along 110 rail stations • Chicago, US	• Y = ratio of the average daily ridership • Xs = - year of fare increase (dummy variables) - day of the week (dummy variables) - annual per capita income - population density - distance from downtown (miles) - proportion of males - Proportion of elderly (aged 65+) - Proportion of children (aged 0–14)	• Annual per capita income (varies) • Population density (+) • Distance from downtown (-) • Proportion of males (varies) • Proportion of elderly (varies) • Proportion of children (varies)	• Aggregated (station level) • Different fixed effects regression models according to the day and year	• Fare changes have a mixed impact on ridership. • For example, for one-year fare change, there was a decline in ridership which was greater in lower-income neighborhoods than it was in higher-income neighborhoods. However, the reverse was found for another year.
Wang & Woo (2017)	• 2485 census block groups between 2000 and 2009 • Atlanta, US	• Y = transit ridership ratio relative to all modes • Xs = - socioeconomic characteristics (i.e., race, marital status, income, and employment) - physical characteristics (i.e., renter-occupied housing, density, land use, and distance to the CBD) - transportation variables (i.e., mode of commuting, travel time for commuting, car ownership, and locations of bus stops) - Locations and poverty rate	• Race (+) • Income (-) • self-employed workers (-) • Proportion of renters (+) • Employment density (-) • High-density residential use (+) • Distance to CBD (-) • Location dummy variables (- / +) • Poverty * suburban areas (+)	• Aggregated (census block level) • Regression analysis	• Poverty rates in suburban areas, compared to areas in downtown and inner-city, positively influence the percentage of transit ridership.

Study	Sample	Investigated factors	Significant factors	Analysis methods	Key findings/recommendations
Campbell & Brakewood (2017)	• 58851 route level observations • New York, US	• Y = unlinked bus trips • Xs = - interaction variable (bike opening date, bus route within a bike station area, number of bike docks near a bus route) - scheduled revenue miles and rapid bus service - real-time bus information - boro Taxi indicator - bike lanes within 0.25 mi of bus route	• Interaction variable (-) • Scheduled revenue miles (+) • Rapid bus service (-) • Real-time bus information (-/not) • Bike lanes within 0.25 mi of bus route (-)	• Aggregated (route level) • Difference-in-differences design • Regression analysis	• Bike-sharing system in New York City has a negative impact on bus ridership. • Every thousand bike-sharing docks along a bus route is associated with a 1.69 to 2.42% fall in daily unlinked bus trips
Chakour & Eluru (2016)	• 8000 stop level observations • Montreal, Canada	• Y = hourly boardings/alightings • Xs = - headway - transit around stop (number of bus, metro and train stops/stations; buses/metro/train line lengths) - infrastructure around the stop (major roads length and highway length) - built environment around stop (parks, commercial enterprises, and residential areas)	• Headway (-) • Number of bus, metro and train stops/stations (+/ns) • Major roads length (+/ns) • Highway length (-/ns) • Built environment (varies)	• Aggregated (stop level and time period) • Ordered response probit model	• Stops are categorized into three groups – low, medium, and high ridership. • Headway affects ridership negatively, while the presence of public transportation around the stop has a positive and significant effect on ridership. • Parks, commercial enterprises, and residential area, amongst others, have various effects across the day on boardings and alightings at bus stops.
Bernal et al. (2016)	• 430 records • Chicago, US	• Y = total ridership • Xs = - slow zone delay - number of scheduled trains - reliability - gas price - holidays, seasons, days of the week	• Reliability (+)- • Slow zone delay (-) • Number of scheduled trains (-) • Gas price (+) • holidays, seasons, days of the week (varies)	• Aggregated (route level) • Regression analysis	• Train slow zones (used for safety during construction) increase headway deviation which reduces ridership
Lee et al. (2015)	• 278 records • Busan, Korea	• Y = Number of boarding • Xs = - number of household - number of business - population in workforce - dummy variable representing regions	• Number of business (+) • Population in workforce (+) • Locations (-/+)	• Aggregated (stop level) • Regression analysis	• Ridership is highly correlated with numbers of business and number of people active in workforce

Study	Sample	Investigated factors	Significant factors	Analysis methods	Key findings/recommendations
Jun et al. (2015)	• 442 records • Seoul, Korea	• Y = total ridership • Xs = - population/employment density - % of the elderly - % of unmarried persons - housing information (small housing units, old housing units, one-person households, rental housing, household size) - land use factors (intersection density, mixed land use, commercial area, department stores) - transit factors (number of bus stops, number of transfers, headway, subway line dummy variables)	• Population/Employment density (+) • % of apartment units (-/ns) • % of rental housing (-/ns) • Old housing units (-/ns) • One-person households (-/ns) • % of unmarried persons (+) • Number of bus stops (+) • commercial area (+) • Intersection density (-) • Mixed land use (+) • Number of transfers (+) • Department store (+) • Headway (-)	• Aggregated (stop level) • Regression analysis and mixed geographically weighted regression (MGWR)	• Population and employment densities, land use mix diversity, and intermodal connectivity all have a positive impact on subway ridership, but differ in their spatial ranges. • In particular, the influence spans of residential and commercial development patterns and mixed land use on ridership were limited to only the core and primary catchment area (within 600 meters).
Brakewood et al. (2015)	• 185 bus routes • January 2011 to December 2013 • New York, USA	• Y = Average weekday route-level unlinked bus trips per month • Xs = - availability of real-time information - scheduled revenue miles - select bus service - bus and rail base fare - rail actual vehicle revenue miles - rail vehicles in peak service - availability of bike-sharing - borough population - gas price - unemployment rate - total monthly snowfall/precipitation - hot/cold month - hurricane Sandy	• Real-time information (+) • Fare ($) (-) • Rail revenue miles (+) • Rail vehicles in peak service (-) • Bike-sharing (-) • Unemployment rate (-) • Cold/hot month (-) • Total snowfall/precipitation (-) • Hurricane Sandy (+) • location dummy variables (varies)	• Aggregated (route level) • Fixed effects regression analysis	• Average weekday unlinked bus trips per month has increased by 118 trips per route per weekday (median increase of 1.7% of weekday route-level ridership) attributable to providing real-time information. • Further refinement of the model suggested that this ridership increase may only be occurring on larger routes.

Study	Sample	Investigated factors	Significant factors	Analysis methods	Key findings/recommendations
Singhal et al. (2014)	• 124,000 hour-based records • Subway 2010 and 2011 AFC data • New York, USA	• Y = hourly ridership residuals (the percentage difference between the actual hourly ridership for a given day and a 9-term moving average) • Xs = - weather variables (snow, rain, heavy rain (1.0 inch), heavy snow (>1.0 inch), wind speed, strong breeze (>25 miles/h), temp. deviation, hot/cold day (10 F higher or lower the 30 year average), Fog, snow last 24 h) - year seasons	• Rain (-) • Snow (-) • Heavy rain (-) • Wind speed (-) • Strong breeze (-/ns) • Temp. deviation (+) • Hot/cold day (+) • seasons (+) compared to summer	• Aggregated (stop level) • Regression analysis	• The time of day models indicate that under the given weather conditions, for any day of the week, the ridership during the PM time period is most affected, followed by midday period and least affected during AM period. Rain and snow affect transit ridership particularly on weekends • Rain, heavy rain, wind speed and hot days are found to have a higher negative impact on ridership at elevated stations than on ridership at underground stations.
Arana et al. (2014)	• 674 daily records • Saturdays and Sundays data in 2010 and 2011. • Cash and AFC data • Gipuzkoa, Spain	• Y = number of daily trips • Xs = - mean wind speed (km/h) - mean air temperature (C) - relative air humidity (%) - rain (l/m2) - day dummy variable	• Rain (-) • Wind (-) • Temp (+)	• Aggregated (stop level) • Regression analysis	• Wind and rain could result in a decrease in the number of trips, while a temperature rise caused an increase in the number of trips
Chakraborty & Mishra (2013)	• 1151 statewide zones • 2000 data • zone data • Maryland, USA	• Y = total daily ridership (boarding and alighting) • Xs = - household/ Employment density - drive alone density - household without cars - household workers density - income less than 60,000 - number of school enrollment - total freeway distance - average free flow speed -accessibility to transit stop (0,1) - housing/Health care/Recreation square feet	• Household density (+) • Employment density (+) • Drive alone density (-) • Carless household (+) • Workers density (+/ns) • Income less than 60,000 (+) • Total freeway distance (-) • Average free flow speed (-) • Accessibility to transit stop (+) • Housing square feet (+) • Health care/Recreation s square feet (-)	• Aggregated (zone level) • Regression analysis & spatial error model	• Land use type, transit accessibility, income, and density are strong predictors of transit ridership for the statewide. • Theses determinants and their coefficients vary across urban, suburban and rural areas.

Study	Sample	Investigated factors	Significant factors	Analysis methods	Key findings/recommendations
Tang & Thakuriah (2012)	• 14,540 records • Longitudinal data 2002 to 2010 • Chicago, US	• Y = monthly average weekday bus ridership • Xs = - real-time bus information system dummy - system internal factors (bus fare, rail fare, key service route dummy, 24-hour service route, frequency, vehicle revenue hours, rail vehicles peak hour) - system external factors (average gas price, unemployment rate, population, snow fall, rain fall, very cold(<32 F), cold(33 to 47 F), Chilly (47 and 62 F), hot (>77 F) - month of the year	• Real-time bus info system (+) • Bus fare (-), • Rail fare (+) • Key service route (+) • 24-hour service route (+) • Frequency, (+) • Vehicle revenue hours (-) • Rail vehicles peak hour (-) • Gas price (+), • unemployment rate (-/+) • Population (+) • Snow, Rain, Very cold, Cold, Chilly (-)	•Aggregated (route level) • Linear mixed-effects model	• The introduction of real-time bus information system has increased bus ridership, although the average increase is modest.
Sung & Oh (2011)	• 214 rail stations • AFC data for two days • Seoul, Korea	• Y = transit ridership • Xs = - Transport system factors (number of bus routes, average headways, number of short bus route, number of bus stops, distance between stations, number of existing stations) - Land use factors (residential, commercial, and business density; land use and commercial/business mix index; Seoul subway accessibility, and regional subway accessibility) - Design characteristics (total road length, average road width, percentage of drive way, four-way intersection density, dead end road, average building group area, average building area)	• Number of bus routes (+) • Average headways (-) • Short bus route (+/ns) • Number of bus stops (+) • Distance between stations (+/-) • Number of stations (+/ns) • Residential/business density (+) • Commercial density (+/ns) • Land use mix (+/ns) and commercial/business mix index (+/ns) • Seoul subway accessibility and regional subway accessibility (varies) • Road length and average road width and % of drive way (-/ns) • Four-way intersection density (+/ns) • Dead end road (+/ns) • Average building group area (-/ns) • Average building area (+/-)	• Aggregated (station level) • Regression analysis	• The study suggests that a mixed land-use pattern has almost the same relationship as high-density development with increased transit ridership.

Study	Sample	Investigated factors	Significant factors	Analysis methods	Key findings/recommendations
Gutiérrez et al. (2011)	• 158 metro stations • November 2004 • Madrid, Spain	• Y = monthly boardings • Xs = - stations characteristics (nodal accessibility, number of lines, accessibility within the network) - areas the stations serve (population, workers, foreigners, population, under 20 years old, population over 60 years old and non-car owning households), employment (in commercial, administration, education, health and industrial sectors), street density, land use mix, urban bus lines, suburban bus lines, parking)	• Nodal accessibility (+) • Number of lines (+) • Foreign population (+) • Workers (+) • Employment in commercial sector (+) • Employment in educational sector (+) • Land use mix (+) • Urban bus lines (+) • Parking (+)	• Aggregated (station level) • Regression analysis	• The paper shows that weighting the variables according to the distance-decay functions provides systematically better results. • The number of non-car households is not highly significant, probably because the Madrid Metro network covers very dense areas. • The number of urban bus lines near the station is highly associated with metro station boardings.
Cervero et al. (2010)	• 69 bus stops • October 2008 • Los Angeles, US	• Y = average number daily boardings • Xs = 22 variables - service attributes (number of buses, hours of service, number of feeder buses, rail connections, rail feeder trains, number of rail lines, bus lanes) - location and neighborhood attributes (population density, employment density, urban density, street connectivity, distance to the nearest BRT stop) - bus stop and site (terminal stop, park-and-ride lot, parking spaces, bus benches, bus shelter, bus schedule information, real-time info system, far-side stop, BRT-branding)	• number of buses (+) • number of feeder buses (+) • number of rail feeder trains (+) • population density (+) • distance to nearest BRT stop (+) • bus lane* number of feeder buses (+) • bus lane* rail feeder trains (+) • bus lane* parking spaces (+) • bus lane* population and employment density (+)	• Aggregated (stop level) • Regression analysis	• Service frequency strongly influences BRT patronage. • High intermodal connectivity, population densities, exclusive-lanes and high employment densities are also associated with high daily boardings.

Study	Sample	Investigated factors	Significant factors	Analysis methods	Key findings/recommendations
Chen et al. (2011)	• 1996 - 2009 • New York City region, US	• Y = rail ridership (linked trips) • Xs = - lagged ridership - seasonal effects - gasoline price - transit fare - labour force - service level (vehicle revenue miles)	• Lagged ridership (+) • Seasonal effects • Gasoline price (+) • Transit fare (-) • Labour force (+) • Vehicle revenue miles (+)	• Dynamic time series models	• The effect of gasoline price, albeit small, is significant, extends over a year and mainly derives from its rise not fall. • Fare is most influential both in terms of short-term and long-term elasticities and its effect is largely contributed by fare increases.

Appendix B: Review of transport authorities and research centres reports

Study	Data sources/sample	Investigated factors	Significant factors	Analysis methods	Key findings/recommendations
Manville et al. (2018)	• Data sources: - US census summary files - Integrated Public Use Microdata (IPUMS) of the Census - State and national travel diary data - National Transit Database (NTD) - Data and rider survey • Periods within 1998 and 2016 • Southern California, US	• Ys = - unlinked passenger trips - transit trip per capita - mean and total daily transit trips - net change in ridership by operators • Xs = - transit service levels - transit service quality - fares - fuel prices - Lyft and Uber - neighbourhood change and migration - vehicle ownership	• The most significant factor: - motor vehicle access (+) • Other: - driver's license (-) - non-whites (+) - foreign born (especially being both foreign born and a new arrival)(+)	• Aggregated (system level) • Descriptive analysis • Regression Analysis • Application of regression parameters to time-series data from the census (years 2000, 2010 and 2015)	• The Transportation Network Companies do not appear to have cannibalized transit. • The substantial increase in private vehicle access particularly among low-income households was the main contributor to falling transit ridership • Instead of trying to win back former riders who now travel by auto, transit agencies better convince the vast majority of people who rarely or never use transit to begin riding occasionally instead of driving.
Feigon & Murphy (2018)	• 1.3 million TNC trip origins and destinations in Chicago, Los Angeles, Nashville, Seattle, Washington, D.C., US, for May 2016 • >10,000 transit & other shared mobility users in 8 metropolitan areas • Transit riders in Atlanta, the Bay Area, New Jersey, and Washington, D.C.	• TNC trips' spatial and temporal distribution • Travel behaviour of transit and TNC users • For exploratory regression analysis: Ys= transit and TNC trip frequencies Xs = - socio-economic characteristics: (university grads (%), households with no vehicle (%), housing tenure: owner/Renter (%), job density, median household income, % population within various ages bands, population density, population earning no more than twice the poverty line (%), unemployed population (%), white/non-white population (%)) - transit and TNC trip frequency (transit stop and schedule, frequency of TNC pick-ups)	–	• Spatio-temporal analysis of TNC trip data • Regression Analysis • Descriptive statistics and tabulations	• Most TNC trips are short and concentrated in downtown core neighborhoods, during evening hours and weekends. • There is no clear relationship between the level of peak-hour TNC use and longer term changes in the study regions' public transit usage. • TNC use is associated with decreases in respondents' vehicle ownership and single-occupancy vehicle trips.

Study	Data sources/sample	Investigated factors	Significant factors	Analysis methods	Key findings/recommendations
Clewlow & Shankar Mishra (2017)	• 4094 respondents (representative sample of urban and suburban populations) • Two phases from 2014 to 2016 • Seven major metropolitan areas (Boston, Chicago, Los Angeles, New York, San Francisco/ Bay Area, Seattle, and Washington, D.C.), US	• Socio-economic characteristics and travel behaviour of TNC users, e.g.: - reasons for/frequency of/trip purpose of ride-hailing - location of respondents - car ownership - use of various transportation modes including public transit	–	• Descriptive statistics	• Ride-hailing users report a net decrease (average reduction of 6%) in their transit use. • Ride-hailing has substituted bus services (a 6% reduction) and light rail (a 3% reduction) while serving as a complementary mode for commuter rail services (a 3% net increase in use).
Schaller (2017)	• New York, US • 2012 - 2016	• Electronic trip logs • FHV (for hire vehicle) and taxi trip volumes • Vehicle mileage, etc.	–	• Descriptive statistics	• TNCs have become the leading source of growth in non-(personal) auto travel in the city, and have pulled more people away from public transit, especially bus, rather than adding riders.
Victoria Transport Policy Institute (2017)	• Transit agencies surveyed by APTA (2008) • US	–	• User type • Trip type • Geography • Type of price change • Direction of price change • Time period • Transit type	• Summary of APTA (2008) elasticity studies	• Elasticity of transit ridership with respect to fares is usually in the –0.2 to –0.5 range in the short run (first year), and increases to –0.6 to –0.9 over the long run (five to ten years). • A relatively large fare reduction is generally needed to attract motorists to transit, since they are discretionary riders. Such travelers may be more responsive to service quality (speed, frequency and comfort), and higher automobile operating costs through road or parking pricing.

Study	Data sources/sample	Investigated factors	Significant factors	Analysis methods	Key findings/recommendations
York Region Transit (2017)	• 2001-2016 and 2016-2020 • York region, Canada	population and employment growth, travel patterns, ridership trends, revenue to cost ratio, etc.		• Descriptive trends	• York Region's transit ridership experienced a rather significant growth, however since 2013 can be due to a reduction in service hours since 2011. • Some suggested initiatives with highest potential for ridership growth: - Build on a strong policy foundation - Consider regional trends and initiatives - Continue implementing YRT strategic initiatives
TCRP (2016)	• 4,500 shared mobility users • Seven cities (Austin, Boston, Chicago, Los Angeles, San Francisco, Seattle, Washington, D.C.), US • Interview with public agency officials and private mobility operators	• Uber API data to understand ridesourcing availability and demand across time and geography (1.07 million observations for the 7 study regions) • Transit agencies' General Transit Feed Specification (GTFS) service information	–	• Descriptive statistics and tabulations • Ridesourcing and public transit capacity and demand analysis	• Ridesourcing appears more likely to substitute for automobile trips than public transit. • Shared modes will continue to grow in significance, and public entities should identify opportunities to engage with them to ensure that benefits are widely and equitably shared.
Smith (2016)	• 4,787 respondents • December 2015 wave of American Trends Panel (ATP) • US	• Socio-demographics, travel behaviour and attitudes towards ride-hailing regulation, e.g.: - frequency of Ride-hailing use - location of respondents - car ownership - use of various transportation modes - political affiliation - user experience	–	• Descriptive statistics and tabulations	• Frequent ride-hailing users are less likely to own a vehicle and more likely to use a range of transportation options (56% regularly take public transportation). • Residential location matters: those Americans who live in an urban center are much more likely to have greater access to ride-hailing services, alongside a range of transportation alternatives that allow

Study	Data sources/sample	Investigated factors	Significant factors	Analysis methods	Key findings/recommendations
					them to live a car-free (or car-light) lifestyle.
The City of Edmonton (2016)	• Edmonton, Canada	–	• Demographics • Car ownership • Trip purpose • Transit type • Transit operation factors: - scheduling & service hour changes - frequency - reliability - bus routing and coverage • Land use and the built environment - density - diversity - design - distance to transit • Natural environment • Attitudes, perceptions and customer experience	• Review of literature and local data (ETS Customer Service Satisfaction)	• Transit ridership is largely a product of factors outside the control of transit systems and the single largest factor affecting transit use is automobile ownership. • Improvements to quality of transit service and targeted pricing schemes have shown to be the most effective mechanisms for increasing ridership. • Outside of the transit system, it is important to develop dense land use and built form, transit-supportive community designs, transit supportive and walkable street networks and high quality LRT facilities to entice new transit riders.

Study	Data sources/sample	Investigated factors	Significant factors	Analysis methods	Key findings/recommendations
Alam et al. (2015)	• 358 Metropolitan Statistical Areas (273 MSAs for final regression model) • 2010 • US	• Y = passenger miles per capita (passenger trip = boarding) • Xs = - MSA population - MSA size - population density - median household income - total households - % African American population - % carless households - vehicles per household - % college population - % population in poverty - % immigrant population - gas price - metropolitan sprawling index - MSAs in the South - rail transit presence - vehicle miles per capita - revenue miles - route miles - service intensity - vehicle hours - revenue hours - average headway - safety - transit fare - transit coverage - transit orientation pattern	• Internal factors: - transit fare (-) - transit supply (+) - revenue hours (+) - average headway (-) - safety (reported number of incidents/accidents involving transit vehicles) (+) - transit coverage (+) - service intensity (-) • external factors: - gas price (+)	• Literature review • Multi-variate regression (year 2010)	• Internal factors were the predominant significant predictors of transit travel demand by bus mode in 2010. Thus the managers and operators are likely to be able to control ridership without depending on outside factors. • The only significant external predictor was gas price.

Study	Data sources/sample	Investigated factors	Significant factors	Analysis methods	Key findings/recommendations
CUTA (2007)	• 46 (of 71) Canadian transit agencies • Available O-D survey data from across Canada (Victoria, Vancouver, Edmonton, Greater Toronto area municipalities, Independent Communities in Southern Ontario, Greater Montreal Area Communities, National Capital Region) • National survey of conventional Canadian transit agencies	–	• External factors: - vehicle access/availability - characteristics of destination land use/location - characteristics of origin or residential land use/location - age/stage of life cycle - employment status - changes in population composition (due to increase in immigrant population living in large Canadian cities) • Internal factors: - transit level of transit service (access/wait/in-vehicle/transfers time) - transit fares	• Literature review • Descriptive statistics and tabulations	• The best and most widely available transit ridership profiling information available would be provided by household telephone origin-destination surveys such as those carried in the Greater Toronto and Montreal area. • Persons aged 15-24 are much more likely to use transit than other cohorts • Some key areas of focus for increasing future of transit ridership: - land use and density planning - immigrants and aging population - specific markets (e.g. commuter corridors to the downtown core, BRT, university or employer-based passes) - transportation demand management policies
Hemily (2004)	• US	• demographic, social, transportation, and land-use trends	–	• Literature review	• Significant trends that will affect transit's effectiveness in the medium-to-longer term: - growing sprawl, in terms of both population and employment - growing auto fleet, use, and distances traveled, - growing congestion - changing travel patterns resulting decreasing traditional work trips and increasing trip chaining.

Study	Data sources/sample	Investigated factors	Significant factors	Analysis methods	Key findings/recommendations
Taylor et al. (2002)	Analysis 1: • >500 transit agencies in the National Transit Database (NDT) • 1991 - 1999 Analysis 2: • 227 transit agencies that increased ridership levels • 1995 - 1999 • US	• Y = - unlinked trips - unlinked trips per person • Xs = - unemployment rate - total employment - real GDP - real GDP/person - per capita income - real hourly wage - vehicle revenue miles - vehicle revenue miles/person - real average fare (per unlinked trip) - vehicle revenue hours	• Unemployment rate (+/-) • Total employment (+) • Real GDP (+) • Real GDP/person (+) • Per capita income (-) • Real hourly wage (+) • Vehicle revenue miles (+) • Vehicle revenue miles/person (+) • Real average fare (-) • Vehicle revenue hours (+)	• Aggregated (system level) • correlation, surveys of transit agency managers	• Factors with the highest correlation to ridership increases (for those agencies that have increased ridership) are increases in revenue service and total employment. • Overall, service improvements were the most frequently cited factors by the successful transit systems. • The report recommends a balance of external and internal adjustments – increasing gas prices and parking costs, combined with improved quality and quantity of transit service – to attract more transit riders.
Kohn (2000)	• 85 transit agencies • 1992 - 1998 • Canada	• Y = no. of urban transit passengers • Xs = - population variables - revenue vehicle hours - revenue vehicle kilometers - average fare - dummies for cities with high/low passenger amounts	• Average fare rate (-) • Revenue vehicle hours (+)	• Aggregated (system level) • Regression Analysis	• In the 1990s, commuters have continued to use urban transit services even though fares increased. • The fare increases may be seen as marginal, when compared to the costs of operating automobiles and downtown parking. • Different policies affect cities in varying ways and there is diversity in urban transit ridership across Canada.

Appendix C: Survey of Canadian ridership prediction practice

INTRODUCTION

The purpose of this survey is to obtain information on the current state of practice in fixed-route transit ridership prediction in Canada. More specifically, it aims at developing a better understanding of the extent to which Canadian transit agencies use models to predict future ridership changes, and the ridership factors incorporated in these models. The survey includes five main sections. The first gathers information on the ridership prediction typology and general practice. The second and third sections include questions on the ridership data sources and types of prediction methods used, respectively. The fourth section elicits information on the explanatory factors and associated data inputs considered for each ridership prediction method, and it also measures the agency's satisfaction with these methods. Finally, the last section inquires about the agency's requirements for robust ridership prediction modelling. The direct benefit of this survey is to help CUTA and the University of Toronto develop an appropriate modelling approach and analytical tool for ridership forecasting as a function of various influencing factors. The term "prediction" used in this survey refers to the estimation or forecasting of ridership at a future time period.

The entire survey should take about 20-30 minutes. We greatly appreciate your time and feedback in completing this survey. Your participation is of course voluntary and you are free to skip some questions during the study. A potential phone interview will be undertaken if additional details are required to complete the study, according to your willingness to answer further questions about the used prediction methodology and the requirement of a robust prediction model. Your verbal consent will be required at the time of the phone interview.

All information gathered will be stored securely at the University of Toronto and will be accessible only by members of the research team and CUTA. Your identity will be linked to your responses through a numerical code. The file linking responses and respondents will be stored securely and separately from all other data. During the course of the study, your responses will be shared with other study participants in a non-identifiable manner. We intend to publish the results of the study with responses presented in an aggregate or non-identifiable manner. The survey involves no identifiable risk to you or your organization.

You can withdraw from the online survey and/or from the phone interview at any point with no further action required. There is absolutely no consequence related to withdrawing from the survey/phone interview. Withdrawal is also possible after completing the survey or the phone interview; respondents can email me to withdraw their submission within a week of the survey deadline completion or interview date. If you wish to take part in the study, and with your superior's approval, please click the following link to start the survey.

If you would like more information about the study, please contact Prof. Shalaby at (416) 978-5907 or amer@ecf.utoronto.ca. If you have any concerns, you are also free to contact our Ethics Review office at 416-946-3273 or ethics.review@utoronto.ca.

Thank you for your participation!

RESPONDENT INFORMATION

Date:

Transit agency name:

Name of the respondent:

Title of the respondent:

Respondent's telephone number:

Respondent's e-mail address:

Respondent's mailing address:

A. RIDERSHIP PREDICTION TYPOLOGY

This section includes general questions about the use of ridership prediction methods.

1. In which case do you predict ridership?
Check all applicable answers

- ☐ Minor scheduling or route adjustments*
- ☐ Major scheduling or route adjustments**
- ☐ The addition of a new route or transit corridor
- ☐ Service improvements (e.g. introduction of transit signal priority (TSP), reserved lanes)
- ☐ Changes in fare
- ☐ Partial network redesign
- ☐ Complete network redesign
- ☐ The introduction of a new mode
- ☐ Specific project evaluation
- ☐ For the next fiscal year (e.g., annual planning for budgeting purposes)
- ☐ Long-term ridership prediction (e.g., next 5 or 10 years) for planning
- ☐ Other (please list):

Notes:
** Minor adjustments refer to changes affecting less than or equal to 25% of a route schedule or structure*

*** Major changes refer to changes affecting more than 25% of a route schedule or structure.*

2. Do you have formal guidelines that define the cases in which ridership prediction is required and the specifications of prediction models?
Check only one answer

- ☐ Yes, we have formal guidelines and we follow them strictly
- ☐ No formal guidelines exist; the need and type of prediction methods are decided on a case by case basis
- ☐ We follow both formal guidelines and informal practices

2.1 Provide comments on your guidelines and practices.

3. Do you predict ridership in terms of linked* or unlinked** trips?
Check only one answer

- ☐ Linked ridership
- ☐ Unlinked ridership
- ☐ Both linked and unlinked ridership

Notes:
** Linked trips refer to trips from origin to destination under one transit agency, where individual trips involving transfers are only counted once.*
*** Unlinked trips refer to the number of times passengers board public transportation vehicles. Here, passengers are counted each time they board vehicles no matter how many vehicles they use to travel from their origins to destinations.*

B. RIDERSHIP DATA SOURCES AND QUALITY

This section focuses on understanding the sources of ridership data

4. What data sources do you currently use for ridership volumes?
Check all applicable answers
- ☐ Ridership data from the farebox
- ☐ Ridership data from automated fare collection (AFC) system
- ☐ Ridership data from ride checks
- ☐ Ridership data from spot checks
- ☐ Ridership data from automated passenger counting (APC) system
- ☐ New technologies (e.g., smart phone data and smart phone applications data)
- ☐ Origin/destination data from on-board surveys
- ☐ Origin/destination data from other household travel surveys
- ☐ Census travel data
- ☐ Other (please list):

5. If you selected more than one data source in the previous question, what are the main and supporting data sources? How are they integrated?

6. Have you changed the source of ridership data over the past 20 years?
Check only one answer
- ☐ Yes, there was a change in the source of ridership data
- ☐ No, there was no change in the source of ridership data

6.1 If yes, when did you change the source of ridership data?

6.2 If yes, which data source(s) have you added and which source(s) have you discontinued?

6.3 If yes, did you notice any difference in data quality and did that require any adjustments?

7. Are you satisfied with the reliability and quality of the ridership data? If not, why?

8. What amount of data do you aim for to support your prediction process and methodology (e.g., sufficient number of records (e.g., 30 trips per route and day)? Do you have access to that amount of data?

C. RIDERSHIP PREDICTION METHODOLOGY

This section includes more detailed questions about the types of ridership prediction methods used.

9. What types of prediction methods* do you use?
Check all applicable answers

- ☐ Professional judgment**
- ☐ Rules of thumb or similar-route cases and analysis ***
- ☐ Trend line analysis****
- ☐ Elasticity based methods
- ☐ Four-step travel demand forecasting model (trip generation, trip distribution, mode choice and transit assignment)
- ☐ Econometric model (regression equations, etc.)
- ☐ Other (please list):

Notes:
** In the following sections, prediction methods will be broken down to:*

1. Qualitative/judgement-based/simple quantitative methods, which include the use of professional judgment; rules of thumb or similar-route cases and analysis; trend line analysis; and elasticity based methods.

2. Quantitative methods, which incorporate the use of more than one explanatory factor in a model. They include four-step travel demand forecasting models and econometric models.

*** "Professional judgment" relies on the judgment and experience of the analyst. For example, an analyst might use professional judgment to adjust a ridership estimate developed by means of another technique depending on his/her subjective expectation.*
**** "Rules of thumb or similar-route analysis" predicts ridership on a given route based on the experiences on other routes with similar service areas and frequencies.*
*****"Trend lines analysis" refers to observing the historical trends of ridership and predicting the change in ridership accordingly.*

10. Does your methodology for ridership prediction change according to **<u>the scale/scope of the change or forecast</u>** (e.g., stop, route or network level)?
Check only one

- ☐ Yes
- ☐ No

11. If your agency operates more than one mode, does your methodology for ridership prediction change according to **the mode of service** (e.g., bus, light rail transit, subway services)?
Check only one
- ☐ Yes
- ☐ No
- ☐ Not applicable. My agency operates only one mode

12. Do you use different methods for **short-term** and **long-term predictions**?
Check only one
- ☐ Yes
- ☐ No

12.1 How do you define short-term and long-term predictions?

13. When do you use these different methods*?

Method	Using it for: (e.g., type of change/mode/horizon)
A. Qualitative/ Judgment-based/Simple quantitative methods	
1. Professional judgment	☐ … ☐ …
2. Rules of thumb or similar-route cases and analysis	☐ … ☐ …
3. Trend line analysis	☐ … ☐ …
4. Elasticity based methods	☐ … ☐ …
B. Quantitative methods which include more than one explanatory factor	
5. Four-step travel demand forecasting model	☐ … ☐ …
6. Econometric model	☐ … ☐ …

Method	Using it for: (e.g., type of change/mode/horizon)
7. Other (please specify):	☐ … ☐ …

Notes:

** Add a few words to summarize when these methods are used according to the type of change, mode and/or horizon. For example, a respondent may select "elasticity based methods" and add the following comment "used for short-term prediction of ridership changes due to fare increases across the bus network"*

14. What specialized software do you use for ridership prediction using any given method?

D. DATA INPUTS AND SATISFACTION WITH PREDICTION METHODS

This section focuses on understanding the inputs used in your prediction methods and your level of satisfaction with these methods.

Subsection 1:

Qualitative/simple prediction methods

These methods include professional judgment; rules of thumb or similar-route cases and analysis; trend line analysis; and elasticity based methods

[If the respondent indicates that he/she uses Elasticity based method in Question 9, Question 15 will show up]

15. What are the inputs to the Elasticity based method you currently use*??

Method	Method inputs
4. Elasticity based methods	☐ … ☐ …

Notes:
** E.g., existing ridership, fare and/or other sources (reference books & manuals)*

[If the respondent indicates that he/she uses any type of qualitative/simple methods in Question 9, questions 16 to 18 will show up]

16. Regarding all the **qualitative/simple methods you use**, on a scale of one to five, indicate how you are **satisfied with**:

Issue	Not satisfied at all 1	Not satisfied 2	Neutral 3	Somewhat satisfied 4	Very satisfied 5	Not applicable
The used ridership prediction qualitative/simple method(s) (e.g., overall satisfaction)	☐	☐	☐	☐	☐	☐
Availability of input data at the appropriate level	☐	☐	☐	☐	☐	☐
Accuracy of input data at the appropriate level	☐	☐	☐	☐	☐	☐
Accuracy of the prediction results	☐	☐	☐	☐	☐	☐
Time required to generate ridership predictions	☐	☐	☐	☐	☐	☐
Effort and experience required to generate ridership predictions	☐	☐	☐	☐	☐	☐
Flexibility of the method to be used in a wider variety of cases	☐	☐	☐	☐	☐	☐
Suitability for long-term ridership prediction	☐	☐	☐	☐	☐	☐
Suitability for short-term ridership prediction	☐	☐	☐	☐	☐	☐
Suitability for implementation across different modes	☐	☐	☐	☐	☐	☐

17. Regarding the **qualitative/simple methods**, on a scale of one to five, indicate to what extent each issue is important:

Issue	Not important at all 1	Not important 2	Neutral 3	Somewhat important 4	Very important 5	Not applicable
Current ridership qualitative/simple prediction method(s)	☐	☐	☐	☐	☐	☐
Availability of input data at the appropriate level	☐	☐	☐	☐	☐	☐
Accuracy of input data at the appropriate level	☐	☐	☐	☐	☐	☐
Accuracy of the prediction results	☐	☐	☐	☐	☐	☐
Time required to generate ridership predictions	☐	☐	☐	☐	☐	☐
Effort and experience required to generate ridership predictions	☐	☐	☐	☐	☐	☐
Flexibility of the method to be used in a wider variety of cases	☐	☐	☐	☐	☐	☐
Suitability for long-term ridership prediction	☐	☐	☐	☐	☐	☐
Suitability for short-term ridership prediction	☐	☐	☐	☐	☐	☐
Suitability for implementation across different modes	☐	☐	☐	☐	☐	☐

18. How do you **calibrate and validate** your methods? How do you assess the **reliability and accuracy** of the prediction results?

Subsection 2:

Quantitative prediction methods which include more than one explanatory factor

These methods incorporate the use of more than one explanatory factor in a model (e.g., four-step travel demand forecasting models and econometric models). If you use more than one prediction model in this category, please add the inputs for each model individually.

[If the respondent indicates that he/she uses any type of quantitative methods in Question 9, this subsection will show up]

19. Please specify the first type of model you use:

Notes:
*You can **add up to six models**; therefore start by adding the **models with the highest number of explanatory factors**.*

20. Check the box of the **<u>relevant factors*</u>** considered in the above specified model and add a brief description of **<u>how these factors are calculated or defined</u>**.

Note:* *These factors are classified into five categories: A. Existing ridership factors, B. Built environment factors, C. Socioeconomic factors, D. Transit service factors and E. Other external/contextual factors*

Factors	Brief description/ definition
A. Existing ridership	
☐ Network ridership	……………
☐ Route, route segment or stop ridership	……………
☐ Similar route ridership	……………
☐ Other (please specify):	……………
B. Built environment factors	
☐ Population density	……………
☐ Urban land area	……………
☐ Green space	……………
☐ Local opportunities: businesses	……………
☐ Local opportunities: recreation	……………
☐ Freeway network length and exits	……………
☐ Highway network length and exits	……………
☐ Street network length and number of intersections	…………… ……………
☐ Railway lines and stations	……………
☐ Private dwellings by type	……………
☐ (e.g. single-detached, apartment)	……………
☐ Dwelling characteristics (e.g., period of construction, condition)	…………… ……………
☐ Dwelling tenure	……………
☐ Property value	……………
☐ Work location (in/outside census subdivision of residence)	…………… ……………
☐ Distance to downtown	……………
☐ Employment in downtown	……………
☐ Other (please specify):	……………
C. Socioeconomic factors	
☐ Age	……………
☐ Gender	……………
☐ Student population	……………
☐ Senior population	……………
☐ Workforce	……………
☐ Unemployment rate	……………
☐ Household size	……………
☐ Car ownership rates	……………

Factors	Brief description/ definition
☐ Ownership of driver's license	……………
☐ Household composition (e.g. couples with (out) children, singles, etc.)	…………… ……………
☐ Household disposable income	……………
☐ Household's expenditure on transport	……………
☐ Average rent/shelter cost	……………
☐ Employment status and type (part-time/full time)	…………… ……………
☐ Employment sector (e.g. agriculture, utilities, construction)	……………
☐ Employment/population ratio	……………
☐ Mode of transport	……………
☐ Education (highest certificate)	……………
☐ Immigration status (citizen or not)	……………
☐ Immigration period	……………
☐ Other (please specify):	…………… …………… ……………
D. Transit service factors	……………
☐ Service frequency	……………
☐ Service reliability (e.g., headway adherence, on-time performance)	……………
☐ Service network coverage	……………
☐ Service span/hours	……………
☐ Vehicle revenue hours	……………
☐ Vehicle revenue miles	……………
☐ Fare	……………
☐ Fare/income	……………
☐ Availability of integrated-fare payment systems	……………
☐ Composition of fleet and modes (bus, subway or LRT)	……………
☐ Density of dedicated bus lanes and transit preferential treatment	……………
☐ Availability of real-time information	……………
☐ Transit service accessibility	……………
☐ Average network load	……………
☐ Transit funding	……………
☐ User satisfaction level	……………
☐ Other (please specify):	…………… ……………
E. Other factors (external/contextual)	
☐ Weather (temperature, snow on ground and precipitation)	……………

Factors	Brief description/ definition
☐ virtual connectivity (telecommuting, online shopping)	……………
☐ Air quality (Air Quality Index and Air Quality Health Index)	……………
☐ Price of car ownership (fuel/energy, insurance, maintenance)	……………
☐ Congestion (average level/cost)	……………
☐ Active transportation support systems (availability and promotion, bike-sharing scheme)	…………… ……………
☐ Vehicles for hire/ride-sharing availability (Uber, Lyft, Taxis)	……………
☐ Other (please specify):	……………

21. Do you want to add **<u>any remarks</u>** about this model?

Notes:
** Remarks can be about the model's <u>level of aggregation, base year and forecasted years</u>. Also, they can be about how the model is <u>integrated and used</u> with other models.*

22. If you use **<u>another model</u>** in this category for predicting ridership, check "Yes" to add the model's inputs. If not, you will proceed to the satisfaction question:

- ☐ Yes
- ☐ No

23. If you can recommend other **<u>new factors</u>** that should be used in ridership prediction, but currently not used, please list and add a brief description of how these factors could be calculated or defined.

24. Regarding all the **<u>quantitative methods</u>** you use**,** on a scale of one to five, indicate how you are **<u>satisfied with</u>**:

Issue	Not satisfied at all 1	Not satisfied 2	Neutral 3	Somewhat satisfied 4	Very satisfied 5	Not applicable
The used prediction model(s) (i.e., overall satisfaction)	☐	☐	☐	☐	☐	☐
Availability of input data at the appropriate level	☐	☐	☐	☐	☐	☐
Accuracy of input data at the appropriate level	☐	☐	☐	☐	☐	☐

Accuracy of the prediction results	☐	☐	☐	☐	☐	☐
Number of used predictive variables	☐	☐	☐	☐	☐	☐
Time required to generate ridership predictions	☐	☐	☐	☐	☐	☐
Effort and experience required to generated ridership predictions	☐	☐	☐	☐	☐	☐
Flexibility of the method to be used in a wider variety of cases	☐	☐	☐	☐	☐	☐
Suitability for long-term ridership prediction	☐	☐	☐	☐	☐	☐
Suitability for short-term ridership prediction	☐	☐	☐	☐	☐	☐
Suitability for implementation across different modes	☐	☐	☐	☐	☐	☐

25. Regarding the **quantitative methods,** on a scale of one to five, indicate to what extent an **issue is important**:

Issue	Not important at 1	Not important 2	Neutral 3	Somewhat important 4	Very important 5	Not applicable
Current ridership prediction model(s)	☐	☐	☐	☐	☐	☐
Availability of input data at the appropriate level	☐	☐	☐	☐	☐	☐
Accuracy of input data at the appropriate level	☐	☐	☐	☐	☐	☐
Accuracy of the prediction results	☐	☐	☐	☐	☐	☐
Using more predictive variables	☐	☐	☐	☐	☐	☐
Time required to generated ridership predictions	☐	☐	☐	☐	☐	☐
Effort and experience required to generated ridership predictions	☐	☐	☐	☐	☐	☐
Flexibility of the method to be used in a wider variety of cases	☐	☐	☐	☐	☐	☐
Suitability for long-term ridership prediction	☐	☐	☐	☐	☐	☐
Suitability for short-term ridership prediction	☐	☐	☐	☐	☐	☐
Suitability for implementation across different modes	☐	☐	☐	☐	☐	☐

26. How do you **calibrate and validate** your models? How do you assess the **reliability and accuracy** of the prediction results?

E. CONCLUDING QUESTIONS

27. In the last three years, what is **one of the major initiatives** to increase the transit ridership in your network and how was **the ridership predicted for this initiative**?

28. How do you define the **requirements of a robust ridership** prediction model?

29. From your experience, what are the **most important lessons** that would benefit other transit agencies regarding ridership prediction practices and methods?

30. Are there any documents or guidelines about ridership prediction methodology used in your agency that you can share with us? If "Yes", we will follow up and email you.

- ☐ Yes
- ☐ Yes, but I can't share
- ☐ No

31. Did your transit agency produce any recent study (within the past 5 years) about the factors affecting transit ridership?

- ☐ Yes
- ☐ No
- ☐

31.1 If yes, can you share it with us? (If "Yes", we will follow up and email you.)

- ☐ Yes
- ☐ No

32. Would you be willing to answer further questions about the used prediction methodology and the requirement of a robust prediction model through a telephone interview?

- ☐ Yes
- ☐ No

33. Thank you for answering this survey and if you have any feedback please use the following box.

Thank you again for completing this survey!

Appendix D: Description of variables tested for inclusion in the final model

Variable	Source	Years available*	Geography/Level	Variable definition/construction
A. Built environment factors				
Transit agency service area	Statistics Canada	Census years	Census Subdivision	Transit agency geographic service area
Total population	Statistics Canada	Census years	Census Subdivision	Transit agency service area total population
Population density	Statistics Canada	Census years	Census Subdivision	Total population/ transit agency service area
Length of highways & major roads	DMTI Spatial Inc.	2001-2016 annually	Canada	Total length of highways & major roads (CARTO<= 4) within transit agency service area
Length of railways	DMTI Spatial Inc.	2002-2016 annually	Canada	Total length of railways within transit agency service area
# Local opportunities	DMTI Spatial Inc.	2002-2013annually	Canada	Total number of businesses and recreation opportunities within transit agency service area
# Occupied private dwellings	Statistics Canada	Census years	Census Subdivision	Total number of occupied private dwellings
# Rooms per dwelling (avg.)	Statistics Canada	Census years	Census Subdivision	Average number of bedrooms per dwelling
$ Dwelling value (avg.)	Statistics Canada	Census years	Census Subdivision	Average value of dwelling
# Band dwellings	Statistics Canada	Census years	Census Subdivision	Number of or historical or on reserve dwellings
# Private dwellings in need of major repairs	Statistics Canada	Census years	Census Subdivision	Number of private dwellings in need of major repairs
% Apartment dwellings	Statistics Canada	Census years	Census Subdivision	Number of apartment dwellings/total number of dwellings
% Row house dwellings	Statistics Canada	Census years	Census Subdivision	Number of row house dwellings/total number of dwellings
% Single-family dwellings	Statistics Canada	Census years	Census Subdivision	Number of single-detached & semi-detached dwellings/total number of dwellings
% Rented dwellings	Statistics Canada	Census years	Census Subdivision	Number of rented dwellings/total number of dwellings
% Owned dwellings	Statistics Canada	Census years	Census Subdivision	Number of owned dwellings/total number of dwellings
Household density				Number of household /transit agency service area
B. Socioeconomic factors				
% Female	Statistics Canada	Census years	Census Subdivision	Number of females/ total population
% Child (age 0-15)	Statistics Canada	Census years	Census Subdivision	Number of children/ total population
% Senior (age 65 and over)	Statistics Canada	Census years	Census Subdivision	Number of seniors/ total population
% Canadian citizen	Statistics Canada	Census years	Census Subdivision	Number of Canadian citizens/total population
% Recent immigrant	Statistics Canada	Census years	Census Subdivision	Number of recent immigrants/total population

Variable	Source	Years available*	Geography/Level	Variable definition/construction
% Population working from home	Statistics Canada	Census years	Census Subdivision	Number of people working from home/total population in workforce
% Postsecondary students	Statistics Canada	Census years	Census Subdivision	Number of postsecondary students/total population
% Unemployed population	Statistics Canada- Labour Force Survey	Yearly	Census Metropolitan Area	Unemployment rate
Participation rate	Statistics Canada	Census years	Census Subdivision	Total labour force/working-age population
# Persons per household	Statistics Canada	Census years	Census Subdivision	Average number of persons per household
# In the labour force	Statistics Canada	Census years	Census Subdivision	Number of persons in labour force
$ Median income	Statistics Canada	Census years	Census Subdivision	Person median income
$ Median household income	Statistics Canada	Census years	Census Subdivision	Median household total income
$ Average gross rent	Statistics Canada	Census years	Census Subdivision	Average gross rent
$ Average major payments for owners	Statistics Canada	Census years	Census Subdivision	Average major payments for household owners
$ Household expenditure on purchase of automobiles	Statistics Canada	1997-2016 annually	Province	Average household expenditure on purchase of automobiles
$ Household expenditure on private transportation	Statistics Canada	1997-2016 annually	Province	Average household expenditure on private transportation
$ Household expenditure on transit	Statistics Canada	1997-2016annually	Province	Average household expenditure on public transport (in terms of fares)
$ Household expenditure on parking	Statistics Canada	1997-2016 annually	Province	Average household expenditure on parking
Person expenditure on public transit	Statistics Canada	1997-2016 annually	Province/ Census Subdivision	Average household expenditure on public transport/number of people per household (*# Persons per household*)
Number of vehicles per person	Statistics Canada	1997-2016 annually	Province	Number of vehicles/total population
% Households having a vehicle (owned or leased)	Statistics Canada	1997-2016 annually	Province	Number of households having a vehicle/total number of households
% Households with 1 vehicle	Statistics Canada	1997-2016 annually	Province	Number of households with 1 vehicle total number of households
% Households with 2 or more vehicles	Statistics Canada	1997-2016 annually	Province	Number of households with 2 or more vehicles/total number of households
% of people work within their CSD of residence	Statistics Canada	Census years	Census Subdivision	Number of workers who work and live within the transit agency core' CSD (with the largest population)/ Number of employed labour force of the same CSD
% of people work outside their CSD of residence	Statistics Canada	Census years	Census Subdivision	Number of workers who work outside the transit agency core' CSD (with the largest population)/ Number of employed labour force of the same CSD

Variable	Source	Years available*	Geography/Level	Variable definition/construction
% of workers with no fixed place of work	Statistics Canada	Census years	Census Subdivision	Number of workers with no fixed workplace address / total employed workforce
C. Transit service factors				
Multi-modal system (dummy)	CUTA	1991-2016 annually	Transit agency	Presence of more than one of mode of transit (dummy variable, 1=present, 0=not present)
# Fixed bus routes	CUTA	1991-2016 annually	Transit agency	Number of fixed bus routes
Total operating budget	CUTA	1991-2016 annually	Transit agency	Total operating budget
Total regular service passenger revenue	CUTA	1991-2016 annually	Transit agency	Total regular service passenger revenue
Vehicle revenue hours	CUTA	1991-2016 annually	Transit agency	Hours travelled by vehicle in revenue service
Vehicle revenue kilometers	CUTA	1991-2016 annually	Transit agency	Number of kilometers travelled by vehicle in revenue service
Disruption >= 20 days (dummy)	CUTA	1991-2016 annually	Transit agency	Service disruptions that exceeded 20 days (median duration)
# Buses	CUTA	1991-2016 annually	Transit agency	Number of buses in revenue service
# Total transit vehicles	CUTA	1991-2016 annually	Transit agency	Total number of transit vehicles in revenue service
# Total low-floor buses	CUTA	1991-2016 annually	Transit agency	Total number of low-floor buses
# Total articulated buses	CUTA	1991-2016 annually	Transit agency	Total number of articulated buses
Service span Tuesday (hours)	CUTA	1991-2016 annually	Transit agency	Span of service in hours on Tuesday (a typical weekday) with least number of missing records)
Service span Saturday (hours)	CUTA	1991-2016 annually	Transit agency	Span of service in hours on Saturday
$ Adult fare cash	CUTA	1991-2016 annually	Transit agency	Adult fare cash price
$ Adult fare unit price	CUTA	1991-2016 annually	Transit agency	Adult fare unit price
$ Adult fare monthly pass	CUTA	1991-2016 annually	Transit agency	Adult fare monthly pass price
$ Student fare cash	CUTA	1991-2016 annually	Transit agency	Student fare cash price
$ Student fare unit price	CUTA	1991-2016 annually	Transit agency	Student fare unit price
$ Student fare monthly pass	CUTA	1991-2016 annually	Transit agency	Student fare monthly pass price
$ Senior fare cash	CUTA	1991-2016 annually	Transit agency	Senior fare cash price
$ Senior fare unit price	CUTA	1991-2016 annually	Transit agency	Senior fare unit price
$ Senior fare monthly pass	CUTA	1991-2016 annually	Transit agency	Senior fare monthly pass price
D. Other external/contextual factors				
$ Gas price	Statistics Canada	1991-2016 annually	Province & city level	Average retail prices for gasoline
Passenger vehicle registration fees (CPI)	Statistics Canada	1991-2016 annually	Province & city level	Passenger vehicle registration fees (CPI)
Passenger vehicle insurance premiums (CPI)	Statistics Canada	1991-2016 annually	Province & city level	Passenger vehicle insurance premiums (CPI)
Other passenger vehicle operating expenses (CPI)	Statistics Canada	1991-2016 annually	Province & city level	Other passenger vehicle operating expenses (CPI)

Variable	Source	Years available*	Geography/Level	Variable definition/construction
Local and commuter transportation (CPI)	Statistics Canada	1991-2016 annually	Province & city level	Local and commuter transportation (CPI)
Parking fees (CPI)	Statistics Canada	1991-2016 annually	Province & city level	Parking fees (CPI)
Presence of Uber (dummy)	Manual data collection		Transit agency	Presence of Uber system (dummy variable, 1=present, 0=not present)
Presence of bike-sharing systems (dummy)	Manual data collection		Transit agency	Presence of fixed bike-sharing systems (dummy variable, 1=present, 0=not present)
Automated fare collection system (dummy)	Manual data collection		Transit agency	Presence of smart card fare system in operations (dummy variable, 1=present, 0=not present)
Average annual temperature (F)	Environment Canada	Yearly	Transit agency	Average annual temperature (F)
Average annual rainfall precipitation (mm)	Environment Canada	Yearly	Transit agency	Average annual rainfall precipitation (mm
Average annual snowfall (cm)	Environment Canada	Yearly	Transit agency	Average annual snowfall (cm)
# Total road motor vehicles	Statistics Canada	1999-2016	Province	Total number of registered road vehicles