Determining the Required Operating Subsidy of Canadian Transit Agencies:

An Examination of Public Transit Operating Deficits in Canada

By

John Hunter Haché

Thesis submitted in partial fulfillment of the

requirements for the Degree of

Bachelor of Science with

Honors in Economics Acadia University

May, 2023

This thesis by John Hunter Haché is accepted in its present form by the

Department of Economics

as satisfying the thesis requirements for the degree of

Bachelor of Science with Honours

Approved by the Thesis Supervisor

______________________________ ____________________

Dr. Justin Beaudoin Date

Approved by the Head or Director of the Department or School

____________________________ ____________________

Dr. Andrew Davis Date

Approved by the Chair, Senate Honours Committee

______________________________ ____________________

Dr. Matthew McSweeney Date

I, John Hunter Haché, grant permission to the University Librarian at Acadia University to reproduce, loan, or distribute copies of my thesis in microform, paper, or electronic formats on a non-profit basis. I, however, retain the copyright in my thesis.

____

Signature of Author

____

Date

Acknowledgments

I want to thank my supervisor, Dr. Justin Beaudoin, for all of his assistance and guidance throughout the data collection and thesis writing process. I would also like to thank Dr. Brian VanBlarcom for his suggestions to improve this paper during my editing process. I am also thankful for the collaborative efforts of a fellow student, Swiler Boyd, who contributed significantly to this paper's data curation and cleaning process.

I also acknowledge the Economics Department at Acadia University for positively impacting my academic growth and accomplishments over the past three years.

Table of Contents

List of Tables and Figures

Abstract

Public transit systems can reduce road congestion, decrease air pollution, and provide affordable transportation within the areas they service. Despite continued growth in Canadian urban transit ridership levels over the past years, public transit systems across Canada consistently earn less revenue than they spend to remain operational. The difference between transit agencies' operating revenue and operating expenditures (known as the operating deficit) is covered by government subsidies which allow public transit to continue working. As more funding becomes available for new transit projects and current transit services, it is essential to understand how much each transit agency requires to remain in operation. This thesis conducts a regression analysis on the determinants of transit demand to estimate the expected per-trip operating deficit a transit agency will experience. Data collected by Statistics Canada and the Canadian Urban Transit Association (CUTA) is used to generate two models which can predict the per-trip subsidy a public transit agency will require based on transit demand levels in the region it services. Data collected from the years 1981, 1986, 1991, 1996, 2001, 2006, and 2016 from 104 Canadian Transit agencies is used in this study.

Transit demand is accounted for with Canadian census data, and findings show that demographic and socioeconomic factors that impact ridership levels among transit users have varying effects on the per-trip subsidy a transit service requires. These effects reveal diseconomies of scale within large, urban transit agencies and explain the relationship between transit demand and transit costs. The results from this analysis allow policymakers to predict expected costs for current transit infrastructure, and determine expected future costs for new transit projects.

1 Introduction

When determining how subsidies are allocated to transit services it is important to understand the required costs faced by transit agencies and the factors which affect demand for their services. This paper examines the connections between transit costs and demand to explain why certain transit services may require more funding than others. Specifically, the relationships between transit agency operating deficits and determinants of transit demand will be examined. By understanding these relationships, a predictive model can be created to estimate the subsidy required to keep a transit system operational.

Public transit is considered a merit-based good because of the positive externalities it provides in the areas it operates. When implemented correctly, transit services can reduce congestion, lower carbon emissions, and provide cost-effective transportation (Buchanan, 2020). Despite this, all government-owned transit systems operating in North America fail to make a profit, and most only generate enough revenue through public fares to cover 33% of their operating costs, (Hannan, 2012). The difference between a transit agency's operating revenue and operating costs is known as an operating deficit. This deficit must be covered by the government subsidies for transit systems in North America to remain operational. Subsidies can either be paid to transit agencies as a lump sum in order to cover operating costs, or they can be used to subsidize user's fares which lowers the consumer cost of transit in the form of reduced trip fares, Hidalgo, Dario (2015). Not only are subsidies able to keep public transit operational, but they are also able to increase ridership by reducing the cost of transit for users.

For the purpose of this study, the per-trip subsidy that a transit agency requires will be investigated instead of the total required subsidy. This is because the total subsidy is primarily determined by the size of the transit service while the per-trip subsidy will reflect the inefficiencies of transit systems that serve different demographics of transit users.

This research is relevant given the recent Canadian Public Transit fund announced by Canada's Federal government in 2019. This fund will provide $14.9 to transit services in 2026 to cover operating costs and investment for new transit projects. Of the $14.9 billion, $5.9 billion will be made available as short-term funding to transit projects all across Canada. The remaining $9 billion will be used to create a permanent transit fund that aims to disburse $3 billion per year to transit services starting in 2026. These funds will be disbursed to municipalities and their transit system based on their needs and priorities; however, it is unclear how these services will be assessed when determining how much money needs to be spent on them. Another fund developed in 2021 by the Canadian government is the Rural Transit Solutions Fund (Infrastructure Canada, 2023), which is providing $250 million of funding towards transit projects specifically in rural areas over 5 years.

The development of these funds makes it important to understand how the demand for transit across Canada is influenced by the demographic, social, and economic factors of the populations using transit in order to ensure that transit agencies receive the correct subsidy corresponding to the level of transit demand they service. An efficient way of determining how funds are disbursed to transit projects is by understanding these trends, comparing them with the costs of transit services, and constructing a predictive model that would suggest the required transit subsidy allocated to a service/project based on the characteristics of the transit demand it supplies.

This paper will also examine whether transit systems operate with diseconomies of scale. Contrary to economies of scale, firms that face diseconomies of scale experience increasing costs per unit as they grow and service higher levels of demand (Perez, 2021). Coulombel, et al. (2019) find that public transit can experience diseconomies of scale when demand thresholds are exceeded. When demand is too high for the supply of public transit in an area, that transit agency will experience increased congestion causing decreased ridership and a decline in service frequency. Large public transit systems operating in metropolitan areas can be inefficient in how they provide service to users, and Min

(2017) finds that transit systems operating in large highly populated cities are often more inefficient than transit systems operating in smaller cities due to less collaborative transit planning which attempts to service a wider demographic of transit users. Because of this, when examining the determinants of ridership demand it is important to understand that an increase in ridership may not always be consistent with an observed reduction in per-trip operating cost.

Most of the existing research on transit investment in Canada has been conducted using individual case studies and focuses on the supply of transit, rather than the demand for transit service, and few studies have examined how diseconomies of scale may impact the operating costs that transit agencies experience, so this research is unique and provides new analysis to the literature.

This paper will conduct a cross-sectional panel analysis of Canadian transit systems, making it easier for policymakers to determine how funds should be distributed to Canadian transit agencies. First, this paper will provide a summation of the current literature available on this topic, then the methodology of the research will be explained, and finally, the findings from the model will be discussed.

2 Literature Review

This section examines the results of similar investigations into the effects that transit supply and transit demand have on transit costs. Additionally, research about the variables describing the socio-economic characteristics of transit users will be discussed to form expectations about what the models generated in this study will show.

2.1 Transit Demand

One of the main points of research when determining how transit subsidies should be distributed is whether or not transit in rural areas should receive more or less funding than transit that operates in urban areas. This is tested by analyzing the demand for transit in urban and rural areas, and by determining whether or not transit agencies exhibit economies of scale.

It is important to note that while there have been many studies conducted into the demand for transit, most of these papers only examine data collected from individual cities, and most studies focus on urban transit systems rather than rural transit agencies. Because of this, there are few studies that examine the differences in transit demand between urban and rural areas, and this could pose a problem for policymakers trying to create transit funds available to both urban and rural transit systems. Furthermore, few take into account the trade-off that occurs when increasing funding to rural transit systems at the cost of decreasing funding to urban transit systems. In the context of the recently created Canadian Public Transit Fund, this is a problem as policymakers have insufficient information on which to decide where transit subsidies are allocated.

Additionally, the existing literature lacks time series data analysis from multiple Canadian cities. This poses a problem for policymakers as impacts on transit demand from changes in transit supply and funding cannot be researched. It is also interesting to note that there have been no studies that attempt to explain the relationship between the factors which affect transit demand, and the subsidy required by transit agencies to maintain operations.

While the articles studying transit demand use different methods, they draw many of the same conclusions about the demographic, economic, and social factors which determine transit demand in rural areas. Borjesson (2020) examines the optimal level of transit supply and subsidies in rural areas. Because there is no data on the willingness to pay of transit users, their demand is instead determined through a utility function created from revealed preference data. This research is done through the use of a theoretical model which accounts for the demand for transit services in rural areas by using demographic variables such as population density, household income, and car ownership (Borjesson, 2020). Through this model, it was determined that areas with lower population densities and higher levels of car ownership have lower expected demand for transit, and therefore the transit subsidy should be reduced in these areas. Additionally, it is suggested that rural transit should be designed with the interests of rural residents in mind by targeting routes that provide access to employment opportunities and regional centers (Borjesson, 2020). This emphasizes the importance of taking the demand for transit into account when planning for the level of transit that will be supplied to an area. This theoretical model and research are important because it explains the reasons behind transit demand in rural areas, and helps quantify the demand. Knowing the level of demand for transit in an area reveals the optimal level of transit that should be supplied, and so the level of investment needed can be better estimated. One limitation of this article is that it only uses data collected from 4 Swedish towns, (Karlstad, Kil, Sunne, and Torsby), and is limited to data collected in one year.

The most recent example of a study that analyzed data from multiple different Canadian cities is Diab et al (2020). This research used data collected from 103 Canadian transit agencies between 2002 and 2016 and was conducted to investigate the factors that impact transit ridership through time in different spatial contexts. The research is based on a two-stage least squares model that explains the relationship between environmental and socioeconomic variables, and levels of transit ridership. Similar to the paper studying optimal levels of transit (Borjesson, Maria 2020), this article finds that lower population densities are more commonly associated with low levels of transit ridership and that having greater access to cars also decreases the level of transit ridership. Interestingly, this analysis factors in the use of Uber among its variables explaining ridership, and found that in rural areas with small transit systems, the use of Uber had a large negative impact on the level of transit ridership while it had a small positive impact in urban areas with larger transit systems. These findings suggest that in rural settings public transit and Uber are substitutes, but in urban areas they are complements. This supports the idea that the variables impacting demand for transit in rural and urban areas can have different impacts given the presence of other demographic factors and reinforces the idea that transit solutions and funding must account for the specific variables present within that transit system's operating region. This analysis also suggests that demographic changes can also impact ridership levels. Specifically, the presence of aging populations and declining birth rates within the service area of a transit agency resulted in lower levels of transit ridership.

The idea that a higher population density supports a higher level of transit ridership is also supported by Taylor, et al. (2013) in a study conducted in the United States across 265 urban areas. The study found that an increase in median household income of 1% would lead to ridership levels increasing by 0.65%. Additional factors that were found to be linked with increased levels of transit demand were high proximity to transit stops and a higher availability/frequency of transit service. These findings suggest that the demand for transit is linked not only to the demographics of transit users but also to the quality of the transit service.

A study conducted by Gomez and Ibanez (1996) into the factors impacting transit demand in Boston found that an increase in income decreased transit ridership. The study found that for a 1% increase in real income, the expected level of transit ridership decreased by 0.75% suggesting that transit was an inferior good. This disparity in findings between the two studies was a point of interest in a study conducted by Guerra and Cervero (2011) into the efficacy of models which used cross-sectional data collected across different transit systems. They argued that when researchers conduct cross-sectional system-level studies, it is possible that they obtain biased and inconsistent coefficient estimates. This is because relevant variables may be omitted from the model, and these omitted variables may be correlated with both the dependent variable and other independent variables. Essentially, the study suggests that when researchers do not account for all relevant demand variables, their results may be misleading. This is because the missing variables may affect both the outcome being studied and the other factors that the researchers are examining. Even variables that are statistically insignificant to the regression can serve as controls for other significant variables making their estimates more efficient within the model. Therefore, it is important to carefully consider all relevant variables when conducting cross-sectional system-level studies. This appears to be supported by the contradicting findings from Gomez and Ibanez (1996) and Taylor et al. (2013), although it is addressed by Taylor et al. (2013). Taylor et al. (2013) suggest that while models based on cross-sectional data can lead to under-specification bias; this bias can be reduced as long as the number of independent explanatory variables within the model is much larger than the number of dependent variables they are trying to explain. This is because the model will better capture the complexity of the relationship between the independent and dependent variables, which will reduce the impact of under-specification bias on the results.

Based on these findings, panel data was used in this investigation because it controls for time-invariant systematic differences across transit agencies. This was done to reduce any under-specification issues that coefficient estimates may develop when based solely on cross-sectional data.

The literature also distinguishes descriptive and causal forms of analyses of transit ridership and demand. Taylor et al. (2013) describes a descriptive analysis of transit demand conducted with qualitative data collected from surveys of transit operators and users to describe the factors that impact transit demand. This method can be flawed as the data collected from these surveys is often subjective and open to multiple interpretations because the data collected is internal to the transit agency it was collected from. Taylor et al. (2013) suggest a more effective way of analyzing transit demand is by examining the social, demographic, and economic factors that impact transit users' demand for transit. Such studies often include a wider range of variables than the descriptive surveys can include, and the data used is less subjective and is external to the transit agency itself, as it is not collected solely from individuals either using the transit system or working for it. It is also noted that studies that use causal analysis as a research method can generate more robust results if a larger sample of transit agencies is used in the investigation, and these results are more accurate in predicting the expected results of other transit systems (Taylor et al. (2013)). For this reason, this study will conduct a causal analysis of the impacts of transit operating costs using, “regression analysis and a multitude of quantitative demographic, economic, and transportation variables, many of which (population density, employment levels, land use, etc.) are ‘external’ to the transit system and its managers.” Taylor et al. (2013). Additionally, this study will include a large sample of transit agencies in order to create a model that is more robust and accurate at predicting the operating costs that different transit agencies are expected to have.

2.2 Independent Variables

In a regression, the independent variables are used to predict an expected outcome from the dependent variable. In this paper, a combination of variables taken from Canadian Urban Transit Association (CUTA) data and Canadian census data was used to express transit demand in the model. The variables used in this analysis were limited by their existence across all years observed in this study within the collected data and were selected by their relevance to transit demand according to the existing literature.

2.2.1 Population Density

Population density has been used as a dependent variable in similar regression studies attempting to predict ridership levels (Borjesson, 2020; Taylor et al., 2009). The findings from these papers suggest that areas with higher population densities are expected to have higher transit demand which is reflected in higher ridership levels. This is consistent with other factors influencing transit demand which could not be accounted for in this investigation as they were not included in all years of the census data used for this research. One factor that influences demand for transit is the distance to transit agency stops. The literature finds that as the distance from users' homes to available transit stops decreases, the level of ridership is expected to increase (El-Geneidy et al.(2013), Taylor et al. (2009)). This is because ease of access to transit has a positive impact on the demand for transit. In areas with a lower population density, it can be expected that the average distance to transit stops for users will be greater than the average distance to transit stops for users in areas with population density. So, having population density in the model helps account for the impact of transit stop distance on transit demand which could not have been included in the model due to data limitations. Population density will be included as a dummy variable within the model in order to classify service populations as either urban or rural. This form of analysis has been used by Ong (2021), which found that rural transit agencies were expected to have 1.5% higher per-trip costs than urban transit systems; however, the rural variable was also deemed statistically insignificant within that model. Controlling for population density as a dummy variable will allow this investigation to examine the differences in expected per-trip subsidy between urban and rural transit systems.

2.2.2 Car Ownership

An important factor in the demand for transit is car ownership within the community the transit agency operates. This data was taken from the Canadian census variable “total labor force population by the method of transportation", and specifically, the values for truck, car, or van ownership by driver were used. This data was then converted into a percentage of the population by dividing the total number of vehicle owners in the census division, by the population of that census division for that year. This is an important variable to include in the calculation of transit demand because high levels of vehicle ownership would be associated with lower levels of transit demand. This is because public transit and private vehicle ownership can be considered substitutionary goods as they are often used in place of one another (Hayes, Adam 2022).

2.2.3 Income

Household income's effect on transit ridership will be included in the model through the annual average family income variable collected from the Canadian Census. This is an important variable to include because there are different arguments in the literature about what impact family income has on public transit ridership. An article by Wang, et al. (2017) examines the impacts of changing socioeconomic factors on transit ridership levels, with an emphasis on evaluating the impact of poverty. This article finds that transit-rich areas are typically located in densely populated, downtown areas in cities, and as high-income families move into these areas, they drive lower-income residents out. These lower-income families are forced to relocate to suburban neighborhoods where they lower the average family income and increase the ridership levels in the area. Conversely, the high-income families moving into downtown city cores increase the average family income while not necessarily increasing the transit ridership levels. This is because higher-income families are more likely to afford private vehicles rather than rely on public transit relative to lower-income families. So, it is important to include this variable in the model to determine whether or not this impact is observable across Canadian transit agencies through time. If this is found, then the conclusion of Wang, et al. (2017) would suggest that increased funding be allocated to suburban transit systems to support the increased demand for ridership there and to increase the accessibility of these transit agencies for low-income users.

2.2.4 Labor force participation rate

The labor force participation rate (LFP) is included in the model as a factor of transit demand. The current literature suggests that LFP has mixed impacts on transit demand for a variety of reasons. Sanchez (1998) explores the connection between employment and transit by using a regression model to estimate the relationship between labor force participation rates and transit accessibility. This study found that levels of employment and transit availability might not be positively correlated in part because transit may not be a cost-effective form of transportation if the transit service is not properly implemented in an area. This finding is backed up by the idea that transit services in areas with low LFP rates are often inefficient in providing access to workplaces for workers (Sanchez, 1998). It also found that in areas with high levels of transit accessibility and ridership, LFP rates tend to increase as accessibility to workplaces increases. As the LFP rates increase, individuals are more likely to purchase private transportation, and so transit ridership can be expected to decrease. Additionally, employed individuals may actively choose to live near transit systems with the intent of using them to commute to work, so there is a possible positive correlation between ridership demand and LFP rates as more workers use transit to commute. Sanchez, (1998) indicates that the results of the study do not indicate a conclusive causal relationship between LFP rates and transit access; however, he does note that the positive relationship between the two may not be purely coincidental because of the aforementioned reasons. As such, it is important to include LFP rates in this model to better understand the relationship between employment and the expected operating deficit faced by a transit agency to ensure that transit agencies operating in areas with low LFP rates can provide better access to workplaces.

2.2.5 Housing Costs

While there is extensive research within the literature on transit accessibility's impact on property value (Cao, et al. (2008)), there are few articles that explore the impact housing costs has on transit demand (Ong, (2021)). The research conducted in this study will use variables taken from the Canadian census data to quantify the percentage of both homeowners and apartment renters whose mortgage/rent payments exceed 25% of their annual income. This variable is similar to the income variable as it explains how much income an individual can spend on transit; however, this variable also takes into account the cost of housing in an area being serviced by transit. The effects of household payments exceeding 25% of owners' income is expected to have mixed impacts on the expected cost of transit in an area, as the number of individuals living in an area with lower rates of payments exceeding 25% of owners' income is expected to have a negative correlation with transit ridership; however, this relationship will be influenced by the same factors that led to LFP rates and transit accessibility being positively related.

2. 3 Transit Supply

Most of the literature on transit demand studies includes an analysis of the factors which impact transit supply, and the effects transit supply can have on transit ridership. Policymakers and transit agencies need to understand the factors that influence transit supply so that they can determine what the most effective use of investment will be in transit supply to increase ridership. By determining the impacts of improved transit frequency, coverage, or other aspects of transit supply, policymakers can make informed decisions about how to effectively finance transit projects that service different levels of transit demand in urban and rural areas.

The method used to quantify transit supply by Taylor et al. (2009) and Diab et al. (2020) are through total vehicle revenue hours. This is a measure of the total time a transit agency operates, and is able to earn revenue through providing commuters with service. Both of these studies use 2 stage least squares regressions to determine the effect certain predictors have on transit supply. Diab et al. (2020) used total direct operating expenses and total population as the independent variables in the model and found that for a 10% increase in total population, the expected increase in revenue vehicle hours would be 5.5%, and for a 10% increase in total operating expenses, the expected increase in revenue vehicle hours would be 4.7%. Taylor et al. (2009) used two different variables to predict revenue vehicle hours; however, both models were able to account for over 80% of the variation in the independent variable. Taylor et al.’s (2009) model found that areas in which a higher percentage of the population voted for Democrats in the 2000 presidential election were more likely to support transit subsidies funded through public expenditures, and so predict a higher level of transit supply in those areas. They also found that populations based in urban areas were linked to 11.5% higher expected levels of vehicle revenue hours. This means that vehicles of urban transit fleets are expected to operate and generate revenue along longer routes than rural transit vehicles.

Berechman and Giuliano (1985) find that decreasing returns to scale are often observed in large public transit services due to congestion issues when transit output is high. Furthermore, transit agencies that operate primarily with buses experience diseconomies of scale due to complexities surrounding their fleet organization and garage capacities. Another study conducted by Viton (1981) examined economies of density by calculating the short-run cost curve of bus transit systems and determining their cost curve elasticity. The long-run envelope cost curve was then calculated in order to examine economies of scale when using vehicle miles as a measure of a transit agency’s output. Viton concluded that diseconomies of scale were observed in transit agencies with higher output levels which served high levels of ridership demand.

3 Methodology

3.1 Data Collection

This investigation aims to explain the factors that impact transit demand in different regions and compare them with the required costs faced by the transit agencies that serve them. This meant collecting data on the demographic, social, and economic variables of transit users, as well as the expenditures and revenues of transit agencies across Canada.

Two data resources were used to construct this model. First, was data collected by the Canadian Urban Transit Association (CUTA) about Canadian transit systems. This source provided information on ridership levels, fare structures, and expenses for more than 100 transit agencies across Canada. This data would be used to determine the differences in revenues and expenditures of transit agencies as well as information on car ownership within the populations of interest. The second resource was Canadian census data collected by Statistics Canada. Every five years, the Canadian census is conducted by Statistics Canada through a survey to collect demographic, social, and economic data on the Canadian population to be used for statistical analysis. This data would provide detailed information about the populations within the service area of each transit agency and would provide the model with demographic, social, and economic descriptors of the average rider for each transit agency. These variables were chosen based on whether or not they were expected to have a measurable impact on transit operating cost or transit demand based on similar studies reviewed in the literature.

For the scope of this investigation, census data collected every five years since 1981 would be used along with the corresponding CUTA data from each year. So, the statistics from each data set would have to be curated into one singular data set which could be inputted into a statistical analysis software. The first step in creating this resource was to determine the correct area of census responses that would be used to determine the population characteristics of each transit agency's users. To do this, a list of all 104 transit agencies across Canada was taken from the CUTA data and inputted into a spreadsheet along with the population size within each transit agency's service area.

To determine the demographic makeup of populations living within these service areas, the correct area in which survey responses were recorded had to be used. This meant matching the recorded population within the service area of each transit agency from the CUTA data with the population surveyed from different census units in the transit agency's location. Census units are subdivisions created by Statistics Canada to break larger regions into small areas within which to conduct their surveys. These subdivisions correspond to Canadian municipalities and are divided into smaller regions so that researchers can use data collected from specific areas rather than the entire subdivision as a whole. Because the service area of each transit area varied, this meant using a combination of smaller census subdivisions, and larger census metropolitan areas when finding the correct set of survey responses to use for each transit area population.

It should be noted that this form of analysis assumes that individuals living in one transit agency's service area solely use that transit agency. This does not account for individuals who commute outside of their census unit with a neighboring transit agency, but for the point of this analysis, the census data collected from individuals living within a particular transit agency's service area should provide an accurate representation of the population using said transit agency.

After the correct census units had been determined for each transit area's location, the survey data from those census units was uploaded into 16 separate spreadsheets. Eight spreadsheets contained the responses to census subdivisions for the years: 1981, 1986, 1991, 1996, 2001, and 2006, and the other eight contained the responses from census metropolitan areas from the same eight years. These spreadsheets were generated from the information collected by Statistics Canada and contained information from every census area in Canada. Next, new spreadsheets were created for each of the eight years which would contain only the data from census areas corresponding to the 104 transit agencies taken from the CUTA data. This meant searching through the larger data files from Statistics Canada and copying over only the census areas corresponding to transit agency service areas. For each year, this data was collected from both the census subdivision and census metropolitan area data sets depending on the transit agency, and the size of its service population. These new spreadsheets contained all of the survey response data corresponding to each transit agency's service population for the years: 1981, 1986, 1991, 1996, 2001, 2006, 2011, and 2016.

3.2 Data Curation

With these newly constructed data sets, the social, demographic, and economic data describing each transit agency's population could be analyzed; however, the surveys conducted by Statistics Canada contained many more population variables than were needed for this research, so the unnecessary variables were deleted. The remaining population variables would be used as independent variables when constructing the regression model.

The CUTA data was sourced directly from the Canadian Transit Fact Book (CUTA (2016)), for each relevant year to this study (1981, 1986, 1991, 1996, 2001, 2006, 2011, 2016). These files contained information on transit ridership from all of the agencies operating within Canada. The relevant variables were then extracted from these files, including the total operating expenditures and operating revenues received in a year by each transit agency, as well as data on car ownership of individuals living within the transit agency's service areas. Once the data had been collected from both the Canadian Census and Canadian Transit Fact Book, the resulting data sets were merged into one file which was then used in the regression analysis.

It is important to note that all variables measured in monetary terms had to be adjusted for inflation so that they would be consistent across all years when used to run regressions. This was done by finding the Consumer Price Index (CPI) value for each relevant year of study dating back to 1981, and then comparing it to the base year CPI, which was 2016 for this investigation. The calculation to find a conversion factor to modify each year's prices to account for inflation is by dividing the 2016 CPI by the CPI of the year of interest. Then the prices from the year of interest could be multiplied by this conversion factor to account for inflation.

3. 3 OLS Regression

This analysis used a one-step OLS regression to estimate the operating deficit a transit agency would be expected to have given certain demographic factors. Regression is a tool used by researchers to quantify the impacts a set of independent variables will have on one dependent variable. By defining the relationship between independent variables and the dependent variable, regressions can find the predicted value of the dependent variable for given values of a set of input independent variables. The simple one-step OLS regression uses the following basic mathematical equation to describe the relationship between two variables in the following equation:

$$y_i = c + \beta x_i + e_i \qquad (1)$$

Where y_i is the dependent variable, x_i is the set of independent variables, and c and β describe the parameters of the regression. The β coefficient shows the effect a one-unit change in x will have on the dependent y variable. For example, if β=1, then for every 1 unit increase in x_i, the expected value of y_i will increase by 1, holding all other variables in the equation constant. The error term, e_i, is the difference between the expected value of y_i (the prediction made by the model) and the actual value of y_i . Graphically, the output of the regression is shown below in Figure 1 where the green dots represent the actual values of y, and the blue slope is the regression's prediction of expected y values:

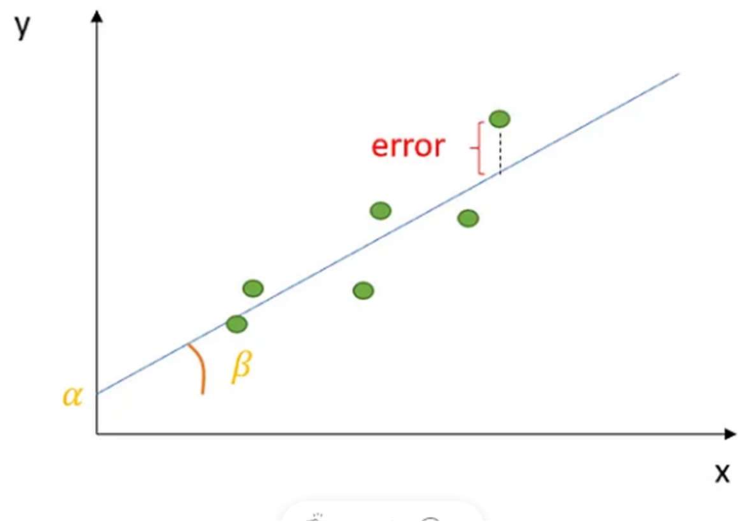

Figure 1: Diagram of an OLS Regression

The goal of the regression is to set parameters (c and β) to minimize the value of this error term. This is done through the method of ordinary least squares which attempts to minimize the value of the sum of squared error terms produced by the regression. For this research, a multiple linear regression was used, which follows the explained methodology of the simple OLS model, but includes additional independent variables to explain the different impacts on the dependent variable of multiple factors.

4 Data limitations

When identifying which census unit to use for each transit agency, certain agencies had to be excluded from the research due to the lack of a clear corresponding census unit. For example, the GO transit system operates within both the Greater Toronto Area, as well as in Hamilton Ontario (GO Transit (2023)); however, there are also municipal transit systems that operate separately within Hamilton and Toronto (CUTA (2016)). The Canadian census includes Toronto and Hamilton as separate census divisions and therefore collects responses specific to the two separate populations. Because this investigation aims to explain the operating costs of a transit agency based on the demographic factors of the population it services, this study is only conducted on transit agencies that service populations that only exist within one census division, and so GO transit was removed from the dataset. Additionally, the areas of some census units were combined for some years and were then separated in later years, meaning there would be no consistent measurement of the same population to describe a transit agency's population over time. Because of this, any transit agency within a census division that spatially changed within the timeframe of this investigation had to be excluded from the dataset.

Another issue with the data was that the data collected by the 2011 Canadian census contained far fewer population variables than the other surveys. This was because in 2010 the Canadian Federal Government decided to get rid of the long-form questionnaire from the census and replace it with a voluntary survey (Okanagan College Library (2022)). So, the 2011 Canadian census dataset is missing several of the key variables to this investigation. If the data from the 2011 Canadian census had been included in the dataset used to run this study's regressions, it would have prevented the model from taking the missing variables into account for any year they were included, significantly reducing the number of observations in the model, and so 2011 was omitted from the dataset.

One of the limiting factors on which variables could be used in the regression was whether or not they were present and had observable, collected data from all the transit locations included in the analysis and for all years included in the analysis. This meant that certain variables were omitted from the model because data were not consistently collected to describe these variables over the years. This is because the survey conducted by Statistics Canada changes from year to year, and different questions and methods of measuring similar trends are used for different years. This meant that certain variables used in other models throughout the literature could not be included in this model because of the wide time range this investigation covered.

Additionally, certain variables used to create the regression models had high degrees of correlation with each other. Variables with high degrees of correlation could not be used as independent variables in the same models or else there would be collinearity within the regression, UCLA (2021). Highly correlated independent variables would be unable to independently determine the expected values of the dependent variable and would lead to coefficient estimates that are statistically insignificant to the model. While this may not cause an issue in studies which had many observations, the relatively small sample size used in this investigation would allow collinearity to generate large standard errors and imprecise coefficient estimates. A correlation matrix was generated to deal with collinearity within regression outputs. Figure 2 displays the correlation coefficients between independent variables. Figure 3 provides a legend for interpreting the independent variables.

```
. correl poplog rural drivers avgmort2 rvklog rvhlog popimm partic gasprice2log operfu
> el2 opertrans2 unemp medinc tripslog
(obs=140)
```

	poplog	rural	drivers	avgmort2	rvklog	rvhlog	popimm	partic
poplog	1.0000							
rural	-0.3532	1.0000						
drivers	0.8149	-0.2298	1.0000					
avgmort2	0.1989	-0.2310	0.2240	1.0000				
rvklog	0.9443	-0.3299	0.7673	0.2080	1.0000			
rvhlog	0.9460	-0.3398	0.7406	0.1974	0.9958	1.0000		
popimm	0.6983	-0.1867	0.8837	0.2953	0.7067	0.6876	1.0000	
partic	-0.0705	-0.2439	-0.0073	0.5100	-0.0556	-0.0571	-0.0501	1.0000
gasprice2log	0.0181	0.1093	0.0587	-0.0900	0.0066	0.0081	0.0206	-0.1036
operfuel2	0.6530	-0.1784	0.7334	0.2197	0.7215	0.7098	0.8994	-0.0565
opertrans2	0.6642	-0.1829	0.7689	0.2387	0.7273	0.7150	0.9229	-0.0686
unemp	-0.0438	0.3734	-0.0316	-0.1254	-0.0376	-0.0476	0.0466	-0.5054
medinc	-0.1031	-0.1348	-0.1173	0.6674	-0.0821	-0.0957	-0.1231	0.3634
tripslog	0.9337	-0.3229	0.7355	0.1120	0.9784	0.9839	0.6682	-0.0567

	gaspri~g	operfu~2	opertr~2	unemp	medinc	tripslog
gasprice2log	1.0000					
operfuel2	0.0433	1.0000				
opertrans2	0.0530	0.9763	1.0000			
unemp	-0.0238	0.0722	0.0672	1.0000		
medinc	-0.0649	-0.0925	-0.0902	0.0155	1.0000	
tripslog	0.0245	0.7008	0.7015	-0.0578	-0.1578	1.0000

Table 1: Correlation Matrix

poplog	Population variable
rural	Rural dummy variable
drivers	Vehicle ownership variable
avgmort2	Mortgage payments >25% of household expenditures
rvklog	Revenue Vehicle Kilometers
rvhlog	Revenue Vehicle Hours
popimm	Immigrant population
partic	Participation rate
gasprice2log	Price of gas
operfuel2	Amount of fuel used
opertrans2	Transit operations
unemp	Unemployment rate
medinc	Median income

Table 2: Legend of Independent Variables

Correlation coefficients range from -1, which represents perfect negative correlation, to +1, which represents perfect positive correlation between variables. For this research, for pairs of independent variables with correlation coefficients with absolute values greater than 0.5, only one variable will be include in the same model to limit collinearity.

5 Results

5.1 Model 1 Specification

The Model 1 specification in Equation 1 explains the impacts of independent transit demand variables on the required subsidy variable. The required subsidy variable was generated first by finding the difference between annual operating expenditures and annual operating revenues for each transit agency across all years to determine the operating loss for each transit service. Then, the loss value was divided by the total number of trips conducted by a transit agency for each year to generate the per-trip loss each transit agency had for each survey year. This loss per trip is equivalent to the required subsidy per trip that a transit agency needs in order to cover its expenditures and continue operating. The log of the cost per trip variable was used as the dependent variable to estimate the percent change in required subsidy per trip a transit agency will face based on the transit demand factors of its users. Table 3 is the regression output generated by running the following specification:

Cost per Trip = f(population, labour force participation rate, revenue vehicle kilometers, number of automobile drivers, average mortgage affordability) (2)

The variation in the dependent variable is measured in partial impacts from each variation in the independent variables, holding all other independent variables constant.

Model 1 uses 217 observations to estimate the impacts of population size, participation rates, revenue vehicle kilometers, number of owned vehicles, and the average mortgage payments of transit users to explain the variation of a transit agency's required subsidy. The r-squared value generated by the regression output details the percentage of variation in the dependent variable explained by the independent variables within the model. As can be seen from the r-squared value of 0.866, Model 1 can explain 86.6% of the variation in the required per-trip subsidy a transit agency will have based on the included independent variables. Model 1 also includes dummy variables for each transit agency included in the regression. This allows the model to control for systemic differences between each transit agency which otherwise could not be controlled for within the regression. Furthermore, the specification for Model 1 includes a code that generates robust standard errors to account for heteroscedasticity.

```
inear regression                                  Number of obs     =        217
                                                  F(68, 139)        =          .
                                                  Prob > F          =          .
                                                  R-squared         =     0.8666
                                                  Root MSE          =      .2765
```

reqsublog	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
poplog	.1201686	.2151242	0.56	0.577	-.3051702	.5455074
partic	-.0701609	.0214548	-3.27	0.001	-.1125807	-.027741
rvklog	.1036621	.0967693	1.07	0.286	-.087668	.2949922
drivers	-3.93e-06	1.46e-06	-2.69	0.008	-6.83e-06	-1.04e-06
avgmort2log	2.122924	.4669447	4.55	0.000	1.199691	3.046157

Table 3: Model 1 Regression Output

5.1.1 Population

From model 1, the population is positively associated with the required per-trip subsidy a transit agency will face. Because the log of the population variable is used in the model, a 10% change in population is shown to have a 1.2% increase in the required per-trip subsidy holding all other variables constant. For a p-value of 0.57, we cannot reject the null hypothesis that the change in required per-trip subsidy will be equal to 0 for a change in population so this variable may be statistically insignificant within the model. This finding is consistent with the literature as transit agencies that service larger populations are found to have higher vehicle revenue hours consistent with larger transit fleets and higher operating costs, Amer, S. et al. (2016). Although a larger population size is also correlated with higher levels of ridership, the increased revenue generated from this additional ridership is unable to offset the additional operating expenditures required to maintain a larger transit infrastructure.

5.1.2 Labor Force Participation Rate

Model 1 shows the labor force participation rate to be negatively associated with the required per-trip subsidy. For a 1% increase in the labor force participation rate, the expected per-trip subsidy required by a transit agency will decrease by 6.8% holding all other variables constant. For a p-value of 0.001, we can reject the null hypothesis that the impact on the dependent variable is equal to 0, and can determine that this variable is statistically significant to the model at all levels of significance. This finding is consistent with the observation of Sanchez (1999) that the effects of labor force participation rates on transit demand are often mixed. In this case, it was observed that the required per-trip subsidy fell for increasing labor force participation rates. This could be explained by the finding that transit services in areas with low labor force participation rates are often inefficient in their implementation and service routes (Sanchez, 1999), and so the operating costs would be relatively high compared to the operating revenues received by these transit agencies resulting in a higher required per trips subsidy. As labor force participation rates increase, so does the level of transit demand for individuals commuting to work, and so transit agencies can develop routes that target specific business districts, and become more efficient in the routes that they offer. This decreases the difference between operating expenditures and revenues and reduces the overall subsidy per trip a transit agency requires.

5.1.3 Revenue Vehicle Kilometers

From model 1, the revenue vehicle kilometers variable is positively associated with the required per-trip subsidy of transit services. The revenue vehicle kilometer variable tracks the number of kilometers each transit agency's vehicles travel while earning revenue for the transit agency. In Table 1, it is observed that a 10% increase in revenue vehicle kilometers is associated with a 1% increase in required per-trip subsidy holding all other variables constant. For a p-value of 0.286, the null hypothesis that this independent variable's impact on the dependent variable is equal cannot be rejected, so revenue vehicle kilometers may be statistically insignificant within this specification. The positive relationship between RVK and per-trip subsidy size is consistent within the literature which finds that larger transit agencies that cover greater service areas will have higher operating costs, and therefore require larger per-trip subsidies to function, Amer, S. et al. (2016).

5.1.4 Car Ownership

Model 1 depicts a slight negative relationship between the number of car owners and the per-trip subsidy required to keep a transit system operating. This specification found that for a 10% increase in vehicle ownership, the expected per-trip subsidy would decrease by 0.000039% holding all other variables constant. The p-value associated with this coefficient is 0.008 so the effect is significantly different from 0 and is statistically significant within the model; however, this finding shows the effect of vehicle ownership on required per-trip subsidy to effectively be 0. This is most likely because of the mixed effects vehicle ownership has on transit demand. The existing literature finds that transit and private vehicles can often be considered substitutionary goods, and so in areas where vehicle ownership is high the corresponding level of demand for public transit would be low, Hayes, Adam (2022). Because of this, it would make sense for transit systems in areas with high levels of private vehicle ownership to be small and require smaller subsidies to operate. However, because of this low level of transit demand induced by high levels of vehicle ownership, the operating revenues received by these transit agencies will be smaller because fewer people are using public transit. Because of this, the impact on the required per-trip subsidy is very close to 0 as there are a variety of factors that influence whether or not it will positively or negatively affect the required subsidy.

5.1.5 Average Mortgage Payments

The average mortgage payments variable denotes the number of people in a service area whose spending on monthly mortgage payments exceeds 25% of their monthly expenditures. Essentially, it is a measure of how many people in the service area have a high or low amount of additional money for the consumption of goods and services after making monthly housing payments. In the Model 1 specification, the relationship between the avgmort2 variable and the required per-trip subsidy is positive. For a 10% increase in avgmort2, the per-trip subsidy is expected to increase by 21%. With a p-value of 0, the null hypothesis can be rejected, and it is determined that this variable is statistically significant within the model. This is consistent with the existing literature which finds that areas in which housing costs are a larger percentage of household expenditures are forced to spend less money on additional expenditures, such as transit use. Because of this, transit agencies operating within service areas characterized by high avgmort2 values are expected to generate less operating revenue than in areas with low avgmort2 values. This supports the findings of Wang, et al. (2017) who suggest that increased funding through subsidies should be spent on transit infrastructure in low-income neighborhoods to increase the accessibility and efficiency of these transit systems.

5.2 Model 2 Specification

The second model was generated using a two-stage least squares (2sls) regression and used the population variable as an instrument to correct for correlation between the drivers variable and the error term of the dependent variable within the model. Other independent variables were selected for this regression based on their coefficients of variation in order to minimize collinearity. The number of vehicle owners in an area is expected to be correlated with the error term of the expected cost per trip of a transit agency because transit agencies can anticipate vehicle ownership levels within the communities they serve and will size themselves accordingly to minimize costs. This creates endogeneity within the model. To determine an influential instrumental variable to use in the 2sls regression, a correlation matrix was generated to determine the levels of correlation between variables of interest. The matrix showed a 0.99 correlation value between drivers and population, and relatively low correlation values between population and other dependent variables within the model, so the population size of transit service areas were selected as the instrumental variable to correct for the endogeneity caused by the drivers variable. This is consistent with the study by Diab (2020) which also used population as an instrumental variable in their 2sls regression. Model 2 uses gas prices, average mortgage payments exceeding 25% of household expenditures, a dummy variable indicating whether or not a transit agency operates within a rural area, the fuel consumption of transit agencies, and transit operating costs (fleet maintenance in compliance with government regulations) to explain the variation in the dependent cost per trip variable. The rural dummy variable was generated according to a definition from Statistics Canada (2016) which delineates rural areas as having a population density of fewer than 400 individuals per square kilometer, and urban areas as having more than 400 people per square kilometer. Model 2 is based on 144 observations and can account for 32.4% of the variation in the cost-per-trip variable.

```
 Instruments: gasprice2log avgmort2log rural operfuel2 opertrans2 pkmlog
              poplog

. ivreg costptlog gasprice2log avgmort2log rural operfuel2 opertrans2 (driverslog = p
> plog)

Instrumental variables 2SLS regression

      Source |       SS           df       MS      Number of obs   =       144
-------------+----------------------------------   F(6, 137)       =     11.32
       Model |  16.6799113         6  2.77998522   Prob > F        =    0.0000
    Residual |  34.7992863       137  .254009389   R-squared       =    0.3240
-------------+----------------------------------   Adj R-squared   =    0.2944
       Total |  51.4791977       143  .359994389   Root MSE        =    .50399

------------------------------------------------------------------------------
   costptlog | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
  driverslog |  -.1353289   .0442817    -3.06   0.003    -.2228928   -.0477649
gasprice2log |   .7657819   .3379507     2.27   0.025     .0975077    1.434056
 avgmort2log |   1.182008   .2219826     5.32   0.000     .7430524    1.620963
       rural |  -.0895522   .1244675    -0.72   0.473     -.335678    .1565736
   operfuel2 |  -3.17e-08   1.04e-08    -3.06   0.003    -5.22e-08   -1.12e-08
  opertrans2 |   4.09e-09   1.83e-09     2.24   0.027     4.80e-10    7.70e-09
       _cons |  -9.482493   2.356295    -4.02   0.000     -14.1419   -4.823081
------------------------------------------------------------------------------
Instrumented: driverslog
 Instruments: gasprice2log avgmort2log rural operfuel2 opertrans2 poplog

.
```

Table 4: Model 2 Regression Output

5.2.1 Vehicle Ownership

Vehicle ownership is negatively related to the expected per-trip subsidy in Model 2. For a 10% increase in vehicle ownership, the expected cost per trip will decrease by 1.35% holding all other variables constant. A p-value of 0.003 is associated with this coefficient, so vehicle ownership is statistically relevant to this regression model at all levels of significance. The impact on cost per trip depicted in this model is greater than the impact shown in Model 1, but it is important to note that a slight negative relationship between vehicle ownership and expected cost per trip is consistent between both regressions. This is due to the mixed effects vehicle ownership has on transit demand and transit supply. Increased personal vehicle ownership is expected to decrease transit demand in an area which decreases the operating revenue a transit agency will receive; however, the decreased demand for transit allows transit agencies to operate with fewer routes and lower vehicle costs, (Hayes, 2022). Model 2 shows that for an increase in vehicle ownership, the operating expenditures of transit agencies are expected to decrease by more than the operating revenues fall, and so the expected required subsidy per trip will also decrease.

5.2.2 Gas Prices

Gas prices are positively correlated with the expected per-trip cost in the Model 2 specification. A 10% increase in gas prices is expected to increase the required subsidy by 7.65% holding all other variables within the model constant. A p-value of 0.025 is associated with the coefficient for the gas price variable, and at the 5% level of significance, its impact on cost per trip is statistically significant to the regression. Model 2 predicts higher required subsidies for transit agencies operating in areas with higher gas prices, and this is fairly consistent with expectations about transit demand. Higher gas prices increase the cost of personal vehicle use which raises the demand for public transit as an alternative method of transportation (APTA 2008). With higher demand for transit, agencies can generate more revenue by providing service to more transit users. However, in addition to the increase in revenue transit agencies are expected to receive as a result of higher transit ridership, transit agencies are also expected to face higher operating costs. This is because higher gas prices will also increase the operating costs of transit agencies by forcing transit services to pay more to fuel their transit vehicles. Additionally, to match the increased levels of ridership, transit agencies are often forced to increase the service they provide along existing routes which further inflates their operating costs (APTA 2008). It is important to note that this model does not contain data collected during the COVID-19 pandemic. During the pandemic, public transit demand was low despite gas prices increasing to record levels due to health concerns surrounding public transit use (Bakx, 2022). Future research should take this into account when explaining the relationship between cost per trip and gas prices.

5.2.3 Average Mortgage Payments

Similar to the Model 1 specification, the average mortgage payments exceeding 25% of the monthly household expenditures variable is shown to have a positive correlation with the expected per-trip subsidy required by transit service. The specification of Model 2 shows that for a 10% increase in the avgmort2log variable, the per-trip cost is expected to increase by 11.8% holding all other variables in the regression constant. A p-value of 0.000 indicates that this variable is statistically significant at all levels of significance within the regression. While the impact observed in Model 1 was larger (a 21% increase in per-trip cost for a 10% increase in avgmort2), both specifications depict similar findings that are consistent with existing research.

5.2.4 Rural

Model 2 includes a dummy variable indicating whether or not the transit agency operates within a rural or urban location. According to the Model 2 specification, public transit in rural areas is expected to have 8.9% lower expected costs per trip than public transit operating in urban areas. The p-value associated with this coefficient is 0.473 which means that this variable may be statistically insignificant in the model. The negative relationship between rural service areas and required subsidy is consistent with existing literature which finds that distance to transit stops has a major impact on transit demand, El-Geneidy et al. (2013), Taylor et al. (2009). Transit services that have larger distances between stops and commuter dwellings experience less demand for transit. Because transit agencies operating in population-dense areas are expected to have smaller distances between stops and commuters' residences, they receive high levels of ridership compared to transit areas in rural areas. The higher ridership levels associated with urban areas force transit services to increase service along their routes to accommodate the high level of transit demand, and this causes operating expenditures to increase at a faster rate than operating revenues. Because of this, transit agencies operating in densely populated urban areas require larger per-trip subsidies to operate.

5.2.5 Fuel Consumption and Transit Operations

The fuel consumption and nominal value of transit operations exhibit slightly negative and slightly positive relationships with the expected cost per trip. While both have statistically significant p-values associated with their regression coefficients (0.003 and 0.027 respectively), the coefficients themselves show that the variables have very small impacts on the required subsidy and transit agency is expected to have. A 10% change in either variable will cause approximately a 0% change in expected cost per trip according to the -3.17e-08 coefficient value associated with fuel consumption, and the 4.09e-09 coefficient value associated with transit operations. Although these variables appear to have a low level of significance within the model, it is important to include them in the specification as controls which allow the regression to provide better estimates for the coefficients of other variables in the model.

6 Conclusion

In general, it was found that transit agencies operating in areas with higher levels of transit ridership required larger operating subsidies. This is because public transit experiences diseconomies of scale when dealing with high levels of demand. While transit services experience increasing returns to scale when operating on a small scale by becoming more efficient, when dealing with high levels of demand they become inefficient as operating expenditures grow faster than operating revenues (Coulombel, Nicolas, et al. (2019)). Certain variables such as population density, car ownership, and gas prices highlight the diseconomies of scale experienced by transit agencies, by showing increasing per-trip costs corresponding to increased demand for transit.

This research created two separate models which can predict the expected per-trip subsidy a transit agency requires based on the social, demographic, and economic characteristics of the populations they serve. These models can predict the expected costs of new transit projects, or currently operating transit systems so that funds can be allocated more effectively to transit agencies. By using data collected from transit agencies across Canada over a 35-year period of time, these models can be implemented anywhere in Canada and can account for regional development over time. This is particularly relevant for estimating the per-trip cost of a new transit development when a new investment is being proposed. By using the values of the independent variables for the region in which the new transit system is being developed, the expected per-trip subsidy of the new transit system can be calculated. Using this predicted value, along with additional capital subsidies, the total opportunity cost of the transit project can be estimated which can provide important information to policymakers when conducting cost-benefit analyses of the project.

Average mortgage payments were shown to have the greatest impact of the independent variables on the expected cost per trip. Specification 1 showed a 21% increase in cost per trip for a 10%

increase in households with average mortgage payments exceeding 25% of the monthly household expenditures, and specification 2 showed an 11.% increase in cost per trip when all other variables were held constant in the regression.

The availability of data limited the number of observations included in each model and reduced the possible combinations of variables that could be included in the same models. Future research should attempt to improve upon this by using data taken from a smaller timeframe with more consistent observations between years and variables. Although the research in this paper used data collected from 1981 to 2016 to account for regional changes and developments over 35 years, this limited the quality of the data sets available for use.

This research can also benefit future studies into models of transit demand by providing insight into which variables may cause collinearity issues if used together in models. Table 1 illustrates the collinearity estimates between the variables collected for use in this study, and highlights the variables showing high degrees of correlation that should not be used to predict the same regressions.

The goal of this study was to research the determinants of transit demand, understand their impacts on transit costs, and predict the expected per-trip subsidies required by transit agencies. Models 1 and 2 (as shown in Tables 3 and 4) are able to account for 86.6% and 32.4% of the variation in per-trip costs of the 104 transit agencies currently operating in Canada. The specifications used in this investigation can be used by policymakers looking to plan future transit investments or to predict the per-trip subsidies that current public transit systems will require as population demographics change.

References

Alfa, Attahiru Sule, and Diana Clayton. “An Improved Procedure for Allocating Public Transit Operating Subsidies ...” An Improved Procedure for Allocating Public Transit Operating Subsidies in Canada, Wiley Online Library, 1986, https://onlinelibrary.wiley.com/doi/abs/10.1002/atr.5670200304.

Bakx, Kyle. “People Are Driving and Travelling Again - so Why Hasn't Public Transit Bounced Back? | CBC News.” CBCnews, CBC/Radio Canada, 31 May 2022, https://www.cbc.ca/news/business/bakx-transit-bus-train-commute-gas-fuel-1.6471060.

Berechman, Joseph, and Genevieve Giuliano. “Economies of Scale in Bus Transit: A Review of Concepts and Evidence - Transportation.” SpringerLink, Kluwer Academic Publishers, 1985, https://link.springer.com/article/10.1007/BF00165470.

Borjesson, Maria. “How Rural Is Too Rural for Transit? Optimal Transit Subsidies and Supply in Rural Areas.” Journal of Transport Geography, Pergamon, 12 Sept. 2020, https://www.sciencedirect.com/science/article/pii/S0966692320302398.

Buchanan, Mark. “The Benefits of Public Transport.” Nature News, Nature Publishing Group, 3 Sept. 2019, https://www.nature.com/articles/s41567-019-0656-8#:~:text=New%20research%20tells%20a%20very,of%20people%20in%20any%20city.

Cooke, S., and R. Behrens. “Correlation or Cause? the Limitations of Population Density as an Indicator for Public Transport Viability in the Context of a Rapidly Growing Developing City.” Transportation Research Procedia, Elsevier, 8 June 2017, https://www.sciencedirect.com/science/article/pii/S2352146517305343.

Correlation, Stata Annotated Output. UCLA: Statistical Consulting Group. from https://stats. oarc.ucla.edu/sas/modules/introduction-to-the-features-of-sas/ (accessed February, 2023)

Coulombel, Nicolas, & Monchambert, Guillaume. (2019). Diseconomies of scale and subsidies in urban public transportation. Retrieved from (https://hal.science/hal-0237 3768v1/document#:~:text=We%20find%20that%20urban%20public,cost%2C%20implyi g%20diseconomies%20of%20scale)

CUTA. (2016). Canadian Transit Fact Book.

Diab, Ehab, et al. “The Rise and Fall of Transit Ridership across Canada: Understanding the Determinants.” Transport Policy, Elsevier, 8 July 2020, https://www.sciencedirect.com/science/article/pii/S0967070X19307930.

El-Geneidy, A., Grimsrud, M., Wasfi, R., Tétreault, P., & Surprenant43 Legault, J. (2014). New evidence on walking distances to transit stops: Identifying redundancies and gaps 44 using variable service areas. Transportation, 41(1), 193-210.

Frankena, Mark. “Income Distributional Effects of Urban Transit Subsidies - Jstor.org.” JSTOR, Journal of Transport Economies and Policy, Sept. 1973,https://www.jstor.org/stable/pdf/ 20052329.pdf.

Guerra, Erick, and Robert Cervero. “Cost of a Ride.” *Taylor & Francis*, Journal of the American Planning Association, 27 June 2011, https://www.tandfonline.com/doi/abs/10. 1080/01944363.2011.589767.

“GO Transit: Regional Public Transit Service for the GTHA.” GO Transit | Regional Public Transit

Service for the GTHA, https://www.gotransit.com/en/.

Gomez-Ibanez, Jose A. “Big-City Transit Rider Snip, Deficits, and Politics: Avoiding Reality in Boston.” Taylor & Francis, Journal of the American Planning Association, 1996, https://www.tandfonline.com/doi/abs/10.1080/01944369608975669.

Hannan, Larry. “Public Transit Not a Profitable Enterprise as Exemplified by First Coast Systems.” The Florida Times-Union, Florida Times-Union, 18 May 2012, https://www.jacksonville.com/story/business/transportation/2012/05/18/public-transit-not-profitable-enterprise-exemplified-first-coast-systems/15866296007/.

Hayes, Adam. “How Substitutes Work.” Investopedia, Investopedia, 8 July 2022, https://www.investopedia.com/terms/s/substitute.asp#:~:text=A%20substitute%2C%20or%20substitutable%20good,used%20in%20place%20of%20another.

Hidalgo, Darío. “To Subsidize or Not to Subsidize Public Transport: That Is the Question: .” TheCityFix, World Resources Institute, 4 Dec. 2018, https://thecityfix.com/blog/subsidize-public-transport-that-is-the-question-dario-hidalgo/#:~:text=Subsidies%20to%20Public%20Transport%20Make%20Sense&text=According%20to%20new%20research%20from,the%20%E2%80%9CMohoring%20Effect%E2%80%9D.

Jones, Ryan Patrick. “Trudeau Pledges Billions in Permanent Funding for Public Transit | CBC News.” CBCnews, CBC/Radio Canada, 10 Feb. 2021, https://www.cbc.ca/news/politics/trudeau-transit-fund-1.5908346.

Library, Okanagan College. “Libguides: Census of Canada: 2011 Census Controversy.” 2011 Census Controversy - Census of Canada - LibGuides at Okanagan College, https://libguides.okanagan.bc.ca/c.php?g=592660&p=4098551.

Min, Hokey, "Assessing the Efficiency of Mass Transit Systems in the United States" Mineta Transportation Institute Publications (2017).

Ong, S. T. (2021). Understanding cost differences of transit services between urban and rural Canada. https://scholar.acadiau.ca/islandora/object/theses:3616

Parry, Ian W. H., and Kenneth A. Small. “Should Urban Transit Subsidies Be Reduced?” American Economic Association, American Economic Review, 3 June 2009, https://www.aeaweb.org/articles?id=10.1257%2Faer.99.3.700.

Perez, Yarilet. “Diseconomies of Scale Definition: Causes and Types Explained.” Investopedia, Investopedia, 13 Sept. 2022, https://www.investopedia.com/terms/d/diseconomies ofscale.asp.

“Rising Fuel Costs: Impacts on Transit Ridership and Agency Operations.” American Public Transportation Association, Sept. 2008.

Rivers, Nicholas, and Bora Plumptre. “The Effectiveness of Public Transit Tax Credits on Commuting Behaviour and the Environment: Evidence from Canada.” Case Studies on Transport Policy, Elsevier, 11 Aug. 2018, https://www.sciencedirect.com/science/article/pii/S2213624X18300853.

“Rural Transit Solutions Fund.” Infrastructure Canada, Gouvernement Du Canada, 20 Jan. 2023, https://www.infrastructure.gc.ca/rural-trans-rural/index-eng.html.

Sanchez, Thomas W. “The Connection between Public Transit and Employment.” Taylor & Francis, Journal of the American Planning Association, 1999, https://www.tandfonline.com/doi/abs/10.1080/01944369908976058.

Shalaby, Amar, et al. “Canadian Transit Ridership Trends Study Final Report.” Canadian Urban Transit Association, Oct. 2018.

Taylor, B. D., and C. N. Fink. “Explaining Transit Ridership: What Has the Evidence Shown?” Taylor and Francis Online, 12 Nov. 2013, https://www.tandfonline.com/doi/full/10.1179/1942786712Z.0000000003.

Thompson, Gregory L., and Jeffrey R. Brown. “Explaining Variation in Transit Ridership in U.S. Metropolitan Areas ...” Sage Journals, Transportation Research Record, 1986, https://journals.sagepub.com/doi/10.1177/0361198106198600121.

Understanding the OLS Method for Simple Linear Regression, https://towardsdatascience.com/understanding-the-ols-method-for-simple-linear-regression-e0a4e8f692cc?source=post_internal_links---------0----------------------------.

Viton, P.A. (1981). "A translog cost function for urban bus transit", Journal of Industrial Economics 29 (3): 287-304.

Wang, Kyungsoon, and Myungje Woo. “The Relationship between Transit Rich Neighborhoods and Transit Ridership: Evidence from the Decentralization of Poverty.” *Applied Geography*, Pergamon, 27 July 2017, https://www.sciencedirect.com/science/article/pii/S0143622817307166.

Xinyu, Cao, and Jill A. Hough. “Hedonic Value of Transit Accessibility: An Empirical Analysis in a Small Urban Area.” Journal of the Transportation Research File, 2008.