Analyzing the Determinants of Public Transit Ridership

Jordan Winter and Alex Mcleod

Abstract

In this paper we examine various determinants that explain public transit demand. This paper follows in step with past papers written on the subject and focuses specifically on data pertaining to Ontario, Canada. OLS and 2SLS regressions are used to estimate demand for public transit. Municipal government expenditure on infrastructure construction is used as an instrument to represent supply. Particular attention is paid to the price of gasoline, the supply variable, and economic condition variables. Their effects are interpreted in accordance with the theory on the topic.

Introduction

Public transit is a crucial part of any modern society's public infrastructure. Patrons of public transit who live in high density areas and who may not be able to afford more expensive forms of transportation depend on it to get to their chosen destination (work, residence, etc.). Estimating the demand for public transit has many applications for the public policy of local communities and higher bodies of government. Policy becomes extremely important for a province such as Ontario, which has many public transit systems, because of its large number of densely populated cities and towns. In Southern Ontario, the most densely populated area in Canada, it may be more viable and cost effective for an individual to use public transit as opposed to other forms of transportation. For example, Toronto has many residents who depend on public transit as a means of transportation. In recent years, traffic congestion in the city has become a huge problem. Highway 401, which spans the city, has the distinction of being the busiest highway in North America.[1] The increasing number of automobiles and lack of new roads have created a situation in which alternative modes of transit are now needed. If the factors that are significant in determining transit ridership per month were known, it would be possible to see what effect a change in each of these factors would have on the demand for transit in a given system. The implication is that if transit authorities knew what factors determined ridership, then they would be more able to respond to demand with an efficient level of supply.

Basic demand theory still holds for the consumption of transportation. A person will consume a good when the utility gained is greater than the disutility of its cost. Taylor et al. (2008) stated that the cost of public transit includes three components: time (wait time, travel time), riding fare, and the level of uncertainty associated with transit ridership (schedule, safety, etc.). According to Horowitz, Koppelman, and Lerman (1986), the most important of these costs is travel time. The various wait times associated with travel can be seen as one way to approximate the level of service; the faster a person can get to

[1] *Annual Average Daily Traffic.* Ministry of Transportation of Ontario. 2004.

a desired destination, the higher the quality of transportation service. In this paper, in order to measure the level of service, the variable "vehicle distance run" is used as opposed to variables capturing a particular wait time. This will be discussed later in the sections describing the data and the estimation model.

Peng et al. (1997) and Taylor et al. (2008) pointed out that a potential problem in estimating travel demand is the simultaneous effect that demand and supply have on each other. The solution put forth in this paper was to use a 2SLS specification to account for the endogeneity of the variable representing supply that was being represented in the demand equation. Not accounting for the simultaneity bias would be tantamount to including supply in the demand equation. Such an error could overinflate the significance of the endogenous variable and produce misleading results. This fact makes demand estimation awkward. In an OLS regression, it is wrong to interpret a statistic which shows increased public transit ridership as a pure increase in demand. That is, it might be the case that the supply variable is having the unobservable effect of creating an upward bias. This feedback effect between demand and supply leads to the conclusion that the OLS framework that could be adopted for estimating the level of transit demand may not be sufficient for this type of analysis, as it will cause a simultaneous equations bias. A better method would be to use 2SLS specification, which accounts for the two simultaneous equations and their feedback effects. This method was adopted for the estimation used in this paper and is discussed in further detail in the Empirical Methods section.

The goal of this paper is to estimate the level of transit demand in Ontario with fair accuracy while taking into account the inherent endogeneity problem. Transit demand is defined as the level of ridership per month in Ontario. It is important to identify what variables are significantly involved in this estimation. Questions being posed at the outset are how does gasoline price affect the level of public transit in Ontario, and what are the other factors associated with explaining variations in the dependent variable.

The remainder of the paper will proceed as follows. The next section will review the relevant theory that relates to public transit demand and will lay down basic assumptions which will be used throughout the paper. The third section discusses the data used in the model. The fourth section will discuss the model itself and justifications for the different variables included. The fifth section will describe the empirical methods used and their justifications. The sixth section will display the results from the estimation and discuss relevant findings and statistics. The final section will conclude the paper by discussing the implications of our results and identifying areas for possible further research.

Relevant Theory and Assumptions

The relevant theory used for this paper concerns the causal analysis of factors that contribute to public transit patronage. In a causal analysis approach to transit demand analysis there is a wide range of variables to consider. These variables can be divided into two categories: external and internal. External factors are outside the control of the transit authorities; these are variables such as population, unemployment rate, and the price of gasoline. Internal factors are variables that are within the transit authorities' control; these are variables such as fare, length of route, the number of buses per route, and the various wait times associated with a trip. Papers that are focused on causal analysis use both internal and external characteristics to help explain the determinants of transit demand. Demand for public transit is a derived demand, meaning that it is consumed in conjunction with other goods or activities. It is derived from the demand to take trips. Trips are motivated by the desire to go to work, to go shopping, to participate in recreation, etc. Keeping this in mind, travel demand can be broken down into a four dimensional model. Small (1997) came up with a framework for transit theory. The theory encompasses all modes of transit but it can be adapted to modeling public transit ridership on its own. The dimensions of transit demand he describes as relevant for consideration when carrying out transportation research are trip generation (the total number of trips taken from or to a certain destination), trip distribution (the area in which these trips are taking place), modal choice (the different modes of travel, for example automobile, bus, bike, walking, etc.), and finally, trip assignment (the precise route that is used). These four dimensions are assumed to be simultaneously determined and based on social and economic factors such as employment characteristics, residential location, and income.

A causal analysis, such as the one used in this paper, has advantages over studies that are more descriptive in nature. Taylor et al. (2008) stated that a study looking at a large number of systems has the ability to produce results that are robust and generalizable, but provides little opportunity for the conceptual development of models that can be used for further research. The estimation in this paper uses data collected over a large number of systems in Ontario, making the results generalizable to other transit systems in this area.

The standard approach to travel demand models (and the one adopted for this paper) is an aggregation approach: regressing total ridership in Ontario on a number of independent variables that have an effect across all of Ontario. This modeling technique can take different forms. Studies such as Gordon and Wilson (1984) take data from a wide range of urban areas and run simple regression analyses in order to get their results. Other studies use a form called *direct demand modeling*, which compiles a detailed data set on a specific city or area and then models demand based on it. Gaudry (1975) takes this direct demand-modeling approach by gathering a very specific data set for Montreal and then using it to determine the level of transit ridership for the city. Models of this type have been criticized because they lack the ability to be generalized. Aggregate data studies also have the problem of possible collinearity. Gaudry (1975) argues, however, that with quality data sets and econometric techniques, this should not be a problem.

An important aspect of transit demand has to do with measuring price elasticity and various cross-elasticities associated with the different variables that affect the demand for transit. Such analysis is imperative to transit operators for setting prices and determining the level of service provided. Kemp (1973) stated that an accepted standard price elasticity for a North American transit system was around -0.33. Although the standard value of -0.33 was once widely accepted, its meaningfulness has diminished in recent years and estimates are now usually determined based on transit areas' individual characteristics. Measurement of elasticities becomes relevant in the Results section where

they are used to show how the price of gasoline affects the level of demand. It is expected that the price of gasoline will have a significant effect on transit demand.
In this paper we make certain assumptions regarding the demand for transit based on the nature of the available data. The model takes the level of transit ridership in Ontario as the dependent variable with various exogenous factors explaining it. The estimation is taken at the system level, meaning it considers the transit system as a whole. Other papers have estimated route level models. Peng et al. (1997) estimated a route level model for transit demand in Oregon. They took into account the relationship between routes and transfers from one route to another. A route level model assumes that a change in one route's level of demand will have either a positive or a negative effect on associated routes. The model in this paper does not take this relationship into account, and thus implicitly assumes that routes do not have an effect on each other. All routes are assumed to be identical and a change in the demand for one route has the same effect on any route in the system. This assumption is made due to the limited availability of route level data.

Several other assumptions are made in this paper. The first assumption is that the demand for transit is the same for every hour of the day, every day of the week. This is, however, certainly not true. The level of ridership differs during a weekend as opposed to a weekday because weekend travelling is typically for leisure activities, whereas weekday travelling is more often for work-related activities. This discrepancy arises for different times of day as well. During rush hours (times of day in which people usually travel to and from their jobs) it is expected that transit usage will be higher than usual. Small (1997) acknowledged that travelers using transit during this time will have a more inelastic level of demand than those who do not. This makes sense, because people travelling for work-related purposes are travelling in order to earn income. Their demand for public transit will be more inelastic compared to people who are travelling for a less significant reason, such as recreation or leisure. This assumption was made because the data collected on ridership are monthly and do not allow for analysis of demand in smaller time frames. Similarly, it is also assumed that all transit travelers have identical preferences and therefore face identical cost curves. This assumption is also untrue because people from different age groups and different situations will differ with respect to their preferences and their associated costs of travel. For instance, children and senior citizens receive discounted fares. Their transit demand elasticities are also likely to be different because the typical adult riders are often travelling to work, while children and seniors are travelling for leisure. This assumption is implicit in the aggregate nature of the data being used for this paper. The aggregate data are taken over the system as a whole and cannot capture individual preferences and costs.

Data

The data used for this study are collected from Statistics Canada via CANSIM e-stat. All variables were collected from this public source through the various tables available. The data most important to the questions being posed are taken from The Large Urban Transit Survey, which was administered from January 1975 to December 1994. Data was collected monthly over this period. The number of transit systems reporting over this time period has a minimum of 15 and a maximum of 41, with the number of systems reporting increasing steadily over the 19 year period.

The CANSIM table used includes the variable *passengers carried*, which is a measure of the level of ridership per month in Ontario. This variable is the dependent variable in our model, and will be discussed in greater detail in The Model section. The table also includes vehicle distance run (a proxy for the level of service in all transit systems). Vehicle distance run can also be thought of as a measure of the level of supply of public transit, as it is the sum of all the kilometers that a bus runs in a month summed over every bus in the transit system. It is thought that this variable is endogenous in the demand equation and therefore causes a bias. This bias is the reason that OLS gives improper results and 2SLS is therefore adopted.

The time period of study was chosen because it contains the longest range of data (15 years of monthly data), while also maintaining many of the variables of interest. The remaining variables used in the final model were also collected from similar tables falling in the same time range.

The variables used in this paper are aggregates for all of Ontario. This means that an observation of labour force, for example, measures the total number of people in the labour force for that month in Ontario. Doing this creates a disconnection between our dependent variable and our independent variables because transit systems operate only in certain areas, while a variable like labour force includes areas without transit systems. This problem has to do with the limitations of aggregate data. Ideally, disaggregate data would be used to capture the behavior of people using public transit. Data of this sort would be able to seperate the impact of the external variable from the internal variable for each particular city's transit system. Although desirable, data of this kind are not available.

When using aggregate data taken at the system level, measuring transit demand leaves out certain nuances of the transit system. For example, if a person makes a trip by bus and has to transfer to four separate buses to complete the trip, the number of passengers carried will go up by four according to the model. A model of transit taken at the route level, however, can distinguish between passengers who board the bus due to a transfer and passengers who are starting their trip on that route. Taking data at the system level creates the problem of double counting individuals who have to take multiple buses in order to complete a single trip. Retrieving route level data is one possible way to disaggregate the model being used and obtain more detailed results from the transit system as a whole. System level data, however, does serve the purpose of this paper's estimation and still allows for significant results with respect to different transit systems.

The Model

In the model, time series data is used in order to estimate the amount of transit ridership for each month in Ontario. The model includes *passengers carried* (the dependent variable) regressed on various explanatory variables and their interaction terms. Due to the nature of the simultaneous equation bias mentioned above, a 2SLS specification is adopted to solve this problem. This model takes the form of supply and demand equations as follows:

$$D_i = \beta_0 + B_1 X_{ki} + \beta_2 Y_i + \mu_i$$
$$Y_i = \Phi_0 + \Phi_1 E_{mi} + \Phi_2 D_i + \upsilon_i$$

Where D_i and Y_i are the levels of demand and supply respectively for each transit system i. X and E are both vectors of different exogenous factors that affect both demand and supply respectively. μ and υ are the error terms for each equation. It can be seen that both supply and demand have a feedback effect on each other. This is the problem described throughout the paper up to this point. Three steps will be taken to correct for this problem. The first step is to regress Y_i on an instrumental variable E_1 , which is a variable from the set of variables contained in E and thought to affect only the level of supply and not demand. The second step is to acquire the estimated value of Y_i , denoted by $\hat{Y}_i$. The third step is then to regress D_i on $\hat{Y}_i$ and all other exogenous variables. Thus, the model takes the form

$$D_i = \beta_0 + \beta_1 X_i + \beta_2 \hat{Y} + \mu_i \quad \text{where} \quad \hat{Y} = \alpha_0 + \alpha_1 E_{1i} + \varepsilon_i.$$

All the data for these variables are aggregates for Ontario and are taken on a monthly basis from January 1976 to December 1994, as mentioned above.[2] Variables based on data collected from the CPI all have a base year of 1997.

The variables used in the final model are:

Log(passcarr) = The log of passengers carried = The log level of transit ridership per month in Ontario.
Log(gas) = The log of the price of gas = The log price of gasoline per month in Ontario.
VDR = Vehicle Distance run = The number of kilometers travelled by a bus during a given month summed over all the buses in a given system.
Labourforce = The number of people in the labour force during a given month in Ontario.
Unemploymentrate = The percentage of labour force participants that are unemployed in Ontario during a given month.
Avgearnings = Average earnings for a full-year, full-time worker in Ontario for a given month.
Winter = A dummy variable that takes the value 0 from April to October and 1 from November to March in adjacent years.
Along with all the explanatory variables, there are interaction variables for each. The interaction variables are all the exogenous variables multiplied the dummy variable *winter*.

[2] A table of the various variables used in the model and their summary statistics is presented in Appendix A.

Passengers Carried is the chosen measure of transit demand in the model, as it captures the idea that transit demand is derived from the demand for trips to various locations, as mentioned above. The exogenous variables used in the demand equation were chosen to cover the different aspects that are thought to affect demand.

The variable price of gasoline uses the Canadian CPI values rather than the Ontario CPI values. This is because data for the Ontario CPI is not available for the full time period covered by the study. Running a regression of the Canadian CPI values on the Ontario CPI values over the period 1978 to 1994 obtained an R^2 of 0.9886. In light of this high correlation, the Canadian CPI values were used instead of Ontario's.[3] Price of gasoline is meant to capture the costs of alternative modes of transportation. In this paper, the main alternative to using public transit is assumed to be driving. This is not the only alternative available (for example, people can walk, bike, etc.), but it was the alternative for which costs could be measured most easily. The major costs of walking or biking are the time and discomfort associated with each, both of which are difficult to measure. The major costs associated with driving are buying a vehicle and purchasing gasoline. In a preliminary regression, the variable Gasoline is found to be statistically significant in determining passengers carried, whereas variables for the price of purchasing a car and the number of cars registered in Ontario are not. This might indicate that fewer people are on the margin when the price of purchasing a car increases than when the price of gasoline does. This implies that the surplus from driving can be outweighed by an increase in the price of gasoline for some trips, but the surplus derived from owning a car is calculated over the life of the car and is therefore harder to offset.

The variable *Labour force* attempts to capture the different trip activities in which individuals engage in an economy. A person in the labour force will have to travel to work or to look for work and, it is believed, will demand more trips. It also captures the level of economic prosperity, and because it is capturing two important aspects of the environment, more economic variables are included to separate the effects of each aspect. *Unemployment rate* is an economic variable meant to capture information on the level of economic prosperity in a similar vein as *labour force*.
Avgearnings is an economic variable that is also meant to capture information on economic prosperity. In times of economic expansion, it is expected that *avgearnings* will increase and in times of recession, it will decrease. Including various economic variables helps to capture different aspect of economic conditions so that the forces affecting transit demand can be separated.

The dummy variable *winter* was included in the regression in order to account for seasonality, which affects both the number of passengers carried per month as well as some of the other independent variables. For example, from the graph of labour force and passengers carried over time, it can be seen that there is a definite dip in labour force during winter months (possibly because of seasonal jobs and the fact that students who were working during the summer are back in school). Conversely, passengers carried sees

[3] Regression of Ontario CPI Values on Canadian CPI values for price of gasoline is presented in appendix B.

a spike during the winter months (possibly due to the fact that during times of cold weather it is more costly for people to walk, bike, etc.). These countervailing effects could distort our results if not controlled for.

Along with all the exogenous variables mentioned above, interaction terms are also included in order to better capture how each of the independent variables changes in the winter months. Having information on seasonality and the effects it has on the assorted exogenous variables allows transit operators to know how demand changes within a given year and how different types of weather affect ridership.

The supply variable in this model is *vehicle distance run*. It is the number of kilometers travelled by each bus in Ontario summed over all buses in a month. It is included in the regression in order to control for the fact that a transit system that has added more routes or lengthened an existing route will see an increase in passengers carried that was not caused by increases in the demand for transit. In this way, it can be thought of as a proxy for the level of service for a given transit system. It will have the effect of reducing wait times and congestion, removing excess demand, attracting new consumers, and possibly improving the connections between routes. However, because of the aforementioned problem with *vehicle distance run*, it is not logical to include it in the demand equation as is. An instrumental variable (IV) is used in place of *vehicle distance run* to account for its inherent endogeneity.

The IV used in place of vehicle distance run is the amount of public capital expenditure spent on infrastructure construction at the municipal level, lagged for five years. This variable can be thought of as the amount of spending a municipal government puts into the construction of public infrastructure. The variable is measured at the municipal level for Ontario and falls within the appropriate time period. The intuitive justification is as follows: funds for infrastructure construction can be spent on a transit system directly, or on building roads, bridges etc. The effect on transit supply through spending directly on the transit system is obvious. The effect on transit supply through spending on other forms of infrastructure is more complicated. If a city is constructing roads, bridges, etc., then it is expanding its city limits to new areas. These areas will become populated and will need to be serviced by the transit system. They will not be as densely populated, however, as the more established areas of the city, and the increase in supply will be met with a very weak increase in demand. In choosing the number of lags on the IV, it seemed natural to go with five years. Construction was assumed to take a few years, and then it would take another couple of years before the transit system could fully expand to into this new area. The IV was found to be highly correlated with the supply proxy *vehicle distance run* and uncorrelated with the demand variable *log of passengers carried*.[4] This gives empirical justification to for the use of this particular IV.

[4] Regressions of the IV on both the measures of supply and demand are displayed in Appendix C.

Empirical Methods

The method we use in this paper for estimating transit demand is a 2SLS estimation applied to the time series data collected. This method is in the same vein as in Taylor et al. (2008), who regressed a similar measure of transit ridership on a number of exogenous factors taken to represent different aspects known to affect demand. Taylor et al. (2008) was also the source of the idea that *vehicle distance run* is endogenous in the demand equation. A 2SLS estimation accounts for the fact that both demand and supply are determined simultaneously and have feedback effects. As both demand and supply show up as endogenous variables in the opposite equation, it is possible that they are correlated with the error terms in both equations. An increase in demand will also increase supply and vice versa.

Working with time series data, we asked whether any of the independent variables contained a unit root. Newbold and Granger (1974) stated that the existence of a unit root provides the possibility of performing a spurious regression with misleading results. After conducting a Dickey-Fuller test on the variables of interest, it was found that *logprice of gasoline*, *labour force*, *unemployment rate*, and *average earnings* all failed the test, and therefore contain a unit root. None of the variables had a MacKinnon approximate p-value for Z(t) less than 0.16. A test then was conducted to see if the linear combination of these variables has a unit root. By regressing the non-stationary variables on each other and saving the residual, it is possible to see whether the linear combination of the suspected variables contains a unit root. The MacKinnon approximate p-value for Z(t) was zero, rejecting the hypothesis of the variables jointly having a unit root and verifying the variables' joint trend as having a stationary process.[5]

[5] The results of the linear combination Dickey-Fuller test are in Appendix E.

Results

Table 1

	OLS	2SLS
Variable	**Coefficient (Standard Error)**	**Coefficient (Standard Error)**
loggas	.3891629* (.0537349)	.5750427* (.0801052
VDR	3.30e-08* (2.61e-09)	1.53e-08 (2.32e-08)
labourforce	-.0004505* (.000057)	-.0005666* (.0000876)
unemploymentrate	-.0245045* (.0025922)	-.0268401* (.0056082)
avgearnings	.0000573* (9.96e-06)	.000088* (.0000197)
winter	2.468813* (.6296336)	5.503552* (.9791525)
Inter-gas	-.269628* (.0930405)	-.6654893* (.1471693)
Inter-VDR	-7.32e-09*** (3.93e-09)	-7.14e-08** (3.18e-08)
Inter-LF	.0004671* (.0001126)	.0010492* (.0001781)
Inter-unemply	.0109428** (.0048094)	.0379671* (.0094594)
Inter-avgearnings	-.000074* (.0000182)	-.0001298* (.0000303)
	R2 = 0.8816 Obs = 169	R2 = 0.5479 Obs = 169

*Significant at 1%

**Significant at 5%

***Significant at 10%

The results of the 2SLS regression compared to the OLS regression are displayed above. All regressors used in the 2SLS model are found to be significant at the 5% level except the instrumented level of *vehicle distance run (VDR)*. In comparing the two results, it is clear that the IV for *VDR* is not significant at all, whereas in the OLS model, *VDR* is the best predictor of public transit demand. This result was expected because of how *VDR* is defined. Including *VDR* in the OLS regression is tantamount to including supply in the demand equation. *VDRhat* is inevitably a worse predictor because the strong feedback connection is broken. Thinking of *VDR* in this light, it makes sense that it is an extremely strong predictor of ridership, given that a great way to predict transit ridership would be to include the service level which the transit authority supplies to meet transit demand.

The variable *labourforce* has the opposite sign from what was expected. This could be because members of the labour force choose to travel largely by driving.

AvgEarning was expected to be negative because it was thought that public transit is an inferior good, but it could be that this measure does not take into account how income is distributed through the sample population. While average income may be growing, the distribution of income may be getting more unequal and not picking up the behaviour of certain income brackets. Another explanation is that as average income rises in the economy, people increase their demand for trips motivated by shopping and recreation.

The coefficient on *log(gas)* can be interpreted as a cross-price elasticity because the dependent variable is logged. The coefficient for the model using 2SLS is 0.575, indicating that a 1% change in the price of gasoline causes a 0.575% change in the number of passengers carried. The results show that the price of gasoline has an inelastic effect on the number of passengers carried. This result was expected because of the high utility derived from driving a car compared to taking public transit. Compared to the OLS elasticity of 0.389, the 2SLS value is relatively more elastic, but not by a large amount.

A winter dummy variable and interaction terms are included in the model in order to account for the behaviour of different exogenous variables during the winter months. The seasonal effect on transit is obvious from the results. The winter variable has a strong positive effect on the level of transit ridership. This was to be expected, since, as mentioned above, during the winter months the cost of other modes of transit (biking, driving, etc.) increases. The discomfort of the cold coupled with snow and ice make driving and other modes of transportation more troublesome. This is reflected in the results of the data, where *winter* is seen to have a strongly positive effect on transit ridership. Figure 1 below displays this seasonality effect graphically, with both the dependent variable, *log(passengers carried)*, and *labour force* being graphed across time. We include labour force in the graph in order to show how seasonality has a countercyclical effect with respect to the level of transit ridership.

It is also interesting to note that all the coefficients of the explanatory variables and their interaction terms have the opposite sign to one another, indicating that transit ridership is very sensitive to changes in weather. When attempting to predict ridership levels, one would have to examine the exogenous factors differently depending on the season.

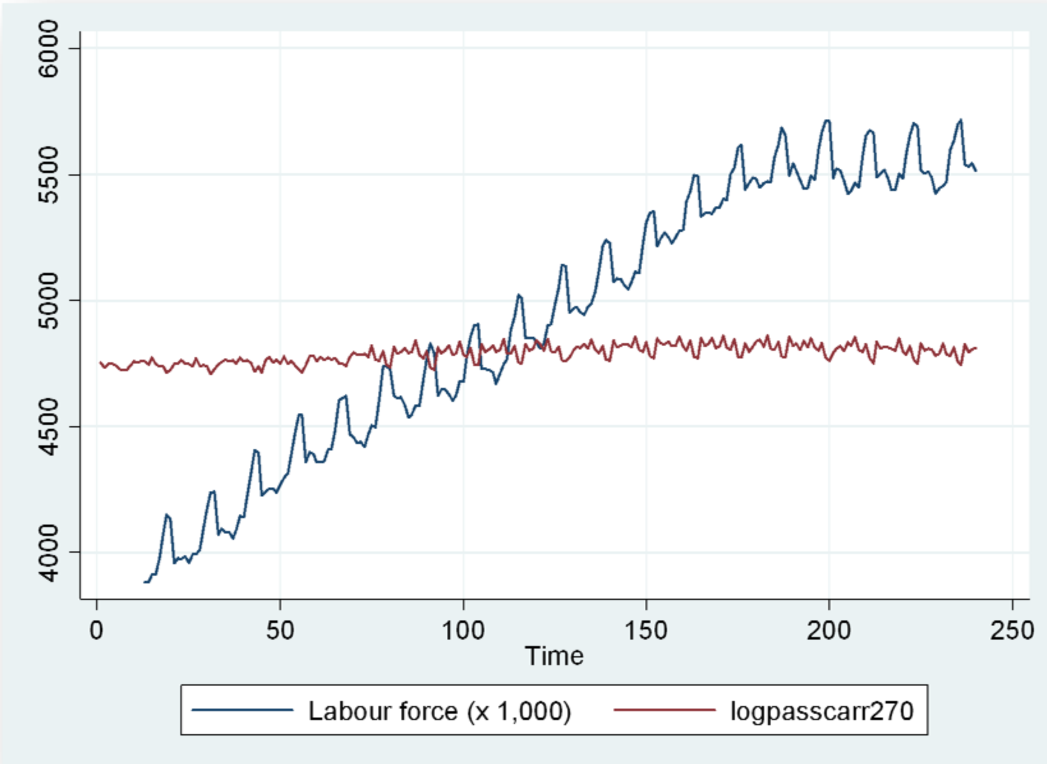

Figure 1

Table 2 shows that a large portion of the variation in transit ridership can be explained by the two variables*: log(gas)* and *winter* (the log of the price of gas and a seasonal dummy variable, respectively). This is an interesting result, which indicates that transit ridership depends heavily on the cost associated with driving an automobile (which can be thought of as a substitute for public transit). As the cost of driving increases, people substitute away from driving and toward using public transit. The argument for *winter*'s significance is the same as above. During the winter season the discomfort of other modes of transit carries with it a higher associated cost, causing people to switch to public transit.

Table 2

Variable	Coefficient (Standard Error)
LogGas	.2118203* (.0122403)
Winter	0876408* (.0110326)
	$R2 = .6035$ Obs = 240

*Significant at 1%

During the decision process for the model, there were variables that were considered to be significant to transit demand but in the end did not make the final regression. Most notable of these variables are the price of bus fare and the number of passenger automobiles registered in Ontario.[6] The price of bus is a direct cost to using public transit, and thus it is strange that it is not significant in the final model. Passenger automobile was found also to be insignificant in the final model, having the opposite sign to what was expected.

This 2SLS model is fairly robust in its application to predict the number of passengers carried in Ontario. This is because the t-values are very high, meaning that the dependent variable was very securely represented by the coefficients. This allows authorities to predict future demand fairly accurately based on the explanatory variables suggested in this paper. These explanatory variables have strong predictive power, showing transit operators what to look for when estimating an individual system's demand. These results are equivalent to what would be the weighted averages of regressions for each particular transit system in Ontario, and thus the model can be applied to any transit system in Ontario, assuming there are no large fixed effects. Transit authorities will be able to respond more efficiently to a change in demand with a corresponding change in supply. The method can also be applied to other provinces or states for predicting levels of transit ridership.

[6] Appendix D contains other regressions that are not included in the Results section

Conclusion

The questions raised at the beginning of the paper were: how do exogenous variables (especially price of gasoline) and vehicle distance run affect the level of transit ridership? The final model shows that vehicle distance run is a very strong predictor of transit patronage, signifying that transit authorities are very adept at setting a level of supply in a given system that is highly correlated with demand. In fact, the correlation between supply and demand is based on more than just good prediction, because demand for public transit is observed at the equilibrium point of demand and supply. That is, people can only consume an amount of public transit equal to or less than the amount supplied. The data on passengers carried will not pick up if the true level of demand is not being met by supply. Vehicle distance run will then be determined simultaneously with demand. Hence, there is a need for a 2SLS model to account for the relationship between demand and supply.

It was also found that the cross-price elasticity between the price of gasoline and the number of passengers carried is positive and inelastic. This was to be expected; as mentioned earlier, the price of gasoline is one of the costs of driving (a substitute for public transit). As it increases, more people would be expected to make the switch from one mode of travel to the other.

The findings on *VDR* and *VDRhat* have two important policy implications. The first arises from the fact that *VDRhat* was statistically insignificant and its coefficient was very small. This indicates that transit authorities cannot use supply to bring about an increase in the level of ridership. The second implication comes from the fact that all the external variables were much more significant than the internal variables. This indicates that transit authorities do not have the power on their own to affect the level of ridership (for example, increasing the length of a route or adding more buses to an existing route will not guarantee an increase in ridership). The results do give transit operators an idea of what exogenous factors contribute to transit demand, and thus the ability to forecast changes in the level of ridership by observing changes in the demographic and other economic variables shown in this study. Transit authorities will not be able to forecast with 100% accuracy how demand will change given changes in certain exogenous factors, but knowing how a factor affects demand allows for more efficient allocation of resources when authorities want to respond to a forecasted change in demand.

The final results with regard to the seasonal variables are revealing about how transit ridership reacts to changes in weather. They show that in seasons with snow and cold temperatures, the variables which are expected to affect demand will behave differently than in seasons with “good” weather. This seasonal effect may seem rather obvious, but the results in the previous section strongly reinforce the idea that seasonal effects are a strong predictor of the behaviour of transit riders.

Although the results presented in this paper are strong, there are always ways to improve the research. An area for future research could be answering a similar question with data disaggregated to the route or even the stop level, instead of to the system level. Another form of disaggregation would be to obtain a detailed data set for an individual city. Although the results in this paper can be generalized to all transit systems in Ontario, there are nuances to any city that may be missing if these results are applied. In the introduction, Toronto was mentioned as having congestion problems. A detailed data set of this nature would give further insight into the behaviour of public transit patrons, and provide transit operators a more solid tool for estimating the level of demand required for them to meet.

References

Gaudry, Marc. 1975. An aggregate time-series analysis of urban transit demand: The Montreal case. *Transportation Research* 9 (4): 249-258.

Gordon, Peter, and Richard Wilson. 1984. The determinants of light-rail transit demand: an international cross-sectional comparison. *Transportation Research Part A: Policy and Practice* 18 (2): 135-140.

Horowitz, J. L., F. S. Koppelman, and S. R. Lerman. 1986. A self-instructing course in disaggregate mode choice modeling. US Department of Transportation, Washington, D.C. (No. Report DOT-T-93-18).

Kemp, Michael A. 1973. Some evidence of transit demand elasticities. *Transportation* 2 (1): 25-52.

Newbold, P., and C. W. J. Granger. 1974. Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society, Series A* 137, Part 2:131-165.

Peng, Zhong-Rhen, Kenneth J. Dueker, James Strathman, and Janet Hopper. 1997. A simultaneous route-level transit patronage model: demand, supply, and inter-route relationship. *Transportation* 24 (2):159-181.

Small, Kenneth A. Urban transportation economics. In *Regional and urban economics, part I*. ed. Richard Arnott, 251-336. London: Routledge.

Taylor, Brian D., Douglas Miller, Hiroyuki Iseki, and Camille Fink. 2008. 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.

Appendix A: Summary Statistics for explanatory Variables
Table 3

Variable	Obs	Mean	Std. Dev.	Min	Max
logpasscarr	240	17.73806	.1332613	17.44017	18.00894
LogGas	240	4.402469	.4452925	3.411148	4.957234
LogBus	240	4.425325	.4598072	3.48124	5.096201
VDR	180	2.74e+07	2767647	2.30e+07	3.27e+07
VDRhat	199	2.78e+07	1989577	2.58e+07	3.20e+07
Pop	228	7307.016	687.0503	6136.3	8467.3
LF	228	4936.327	536.4474	3882.3	5717.1
Urate	228	7.784649	2.04547	4.3	12.9
AvgEarnings	217	46684.56	1708.846	43600	50700
under25	217	44.81587	2.007781	40.7	48.7
PassAuto	229	4159232	548738.6	3225243	5069383
winter	240	.4166667	.494037	0	1

Appendix B: Ontario and Canadian CPI values for gasoline
Table 4

gasCAN	Coef.	Std. err.	T	P>t
gasONT	1.582807	.0121849	129.90	0.000
cons	-1.976095	.8075198	-2.45	0.015
N= 196 R^2= 0.98886				

Appendix C: correlation of IV with demand and supply

Table 5

Logpasscarr	**Coef.**	**Std. err.**	**t**	**P>t**
capitexpcons_60	.0000349	.0000202	1.73	0.086
Cons	17.74925	.0241101	736.18	0.000
$R^2 = 0.0165$ N = 180				

Table 6

VDR	**Coef.**	**Std. err.**	**t**	**p>t**
capitexpcons_60	3797.087	425.2155	8.93	0.000
Cons	2.31e+07	507392.2	45.51	0.000
$R^2 = 0.3094$ N = 217				

Appendix D: Tables of other regressions

Table 7: OLS regression with economic condition variables

Variable	Coefficient (Standard Error)
Population(1000s)	.0004649* (.0000613)
LF	-.0003467* (.0000693)
Unemploymentrate	-.0168648* (.0061315)
Under25	-.0034004 (.006833)
AvgEarnings	-.0000367* (7.05e-06)
	R2 = .5032 Obs = 217

*Significant at 1%

Table 8
OLS regression including variables thought to be significant but didn't end up in final model

Variable	Coefficient (Standard Error)
LogGas	.2494602* (.0559487)
LogBus	-.156042 (.1144997)
PassAuto	1.11e-07 (6.94e-08)
AvgEarnings	1.22e-07 (6.23e-06)
Winter	.0887828* (.0116161)
	R2 = .5994 Obs = 217

*Significant at 1%

Appendix E: The Linear Combination Dickey-Fuller Test

```
Dickey-Fuller test for unit root                   Number of obs   =       216

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -6.108            -3.471            -2.882            -2.572
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
```