

MACALESTER COLLEGE
ECONOMICS HONORS THESIS

**How do changes in gasoline prices
affect bus ridership in the Twin
Cities?**

Author:

Kim Eng KY

Supervisor:

Prof. Sarah WEST

Readers:

Prof. Gary KRUEGER

Mr. Joel HUTING

November 8, 2016

Abstract

What determines the likelihood a resident will take public transit? To understand the relative effects of various determinants of ridership, this study focuses on the effects of gasoline prices, controlling for other determinants such as number of workdays in a given month, traffic, unemployment rate, and service quality. As gasoline prices increase, it becomes more expensive to drive a car. Thus, customers would likely shy away from driving and substitute towards the alternatives; one of which is public transportation. This study aims to find the relationship between gasoline prices and bus ridership at a disaggregated level (route type level) in the Twin Cities using Ordinary Least Squares and Autoregressive Integrated Moving Average estimators. The cross-price elasticities of local and express bus ridership with respect to gasoline prices are 0.139 and 0.220, respectively.

Contents

1	Introduction	2
2	Literature Review	3
3	Theory	7
4	Data and Summary Statistics	13
4.1	Data	13
4.2	Summary Statistics	16
5	Main Analysis	18
5.1	Ordinary Least Squares	18
5.1.1	Local Routes	18
5.1.2	Express Routes	21
5.2	Autoregressive Integrated Moving Average	24
5.2.1	Introduction to ARIMA with Exogenous Regressors	24
5.2.2	Local routes	25
5.2.3	Express Routes	27
6	Discussion	28
7	Conclusion	30
	Acknowledgements	31
	Figures and Tables	35

1 Introduction

In order to reduce gasoline consumption, traffic congestion, and pollution, policymakers may implement policies that increase gasoline prices. Policymakers hope that by doing so, they will induce households to switch from gasoline-intensive transportation modes into cleaner public transit. This thesis examines the relationship between gasoline prices and bus ridership in the Twin Cities.

Within the past decade in the Twin Cities, gasoline prices have fallen to less than \$2 per gallon towards the end of 2008 and the beginning of 2009 (see figure 1 for the plot of gasoline prices in the Twin Cities over time). Gasoline prices then start increasing to be between \$3 and \$4 and fluctuate on a relatively flat trend from 2011 until mid 2014, which is when there is a significant fall in gasoline prices. Along with the fall in gasoline prices, we have also noticed the fall in transit ridership. This seems to suggest a positive relationship between gasoline prices and transit ridership. The change in consumers' behavior towards public transportation when gasoline prices change is not unique to the Twin Cities. Morath (2016) writes on the Wall Street Journal blog about how the fall in gasoline prices is associated with the fall in bus ridership nation wide. For policy implications, policymakers are interested in whether or not changes in gasoline prices affect transit ridership.

As one of the two light rail lines in the Twin Cities is relatively new, I will focus on bus ridership. For a more fine-grained analysis, I will divide bus ridership into two categories: local and express. This thesis contributes to the literature by providing a more recent empirical analysis on the relationship between gasoline prices and bus ridership at a disaggregated level. The question I am investigating is: How do Changes in Gasoline Prices Affect Bus Ridership in the Twin Cities? In order to answer this research question, I will also control for other factors that may affect bus ridership in addition to gasoline prices such as number

of workdays in a given month, unemployment rate, traffic, fare change, and service quality.

The following section discusses findings in the literature and how my thesis contributes to the study of the relationship between gasoline prices and bus ridership. I will build the theoretical framework for answering the question in section 3. In section 4, I discuss how I select variables to be included, possible issues with the data, and summary statistics. Section 5 presents the empirical analyses and includes discussion of potential estimation issues, and the main regression results. Section 6 provides discussion of the results. I conclude in section 7 and discuss directions for future research.

2 Literature Review

The impact of gasoline price changes on transit ridership has been a subject of interest among economists for the past several decades. This relationship is frequently measured by cross-price elasticities of demand for public transit ridership with respect to gasoline prices. This term, cross-price elasticities, is defined as the percentage change in transit ridership caused by a one percentage change in gasoline prices. Public transit can be divided into different modes: buses, light rail, heavy rail, and commuter rail. Though it is not explicitly specified, most studies that distinguish between transit modes find different elasticities for each mode (Litman 2004).

Table 1 summarizes cross-price elasticities of ridership with respect to gasoline prices found by previous literature. We can see that cross-price elasticities range from -0.17 to 0.88. This range is considerably large. This might be because of the diversity in locations of data and different time spans used in these studies, meaning that people from different time periods and/or different places react to changes in gasoline prices differently. Cities that are more auto-dependent tend to be more sensitive towards changes in gasoline prices than more

transit-oriented communities (Lane 2010).

Negative cross-price elasticities imply that as gasoline prices increase, transit demand decreases. This contradicts many studies' finding that a rise in gasoline prices would cause auto trips to be more expensive, causing consumers to substitute driving with other modes of transportation, one of which is public transportation. Lane (2010) finds the cross-price elasticities to be negative for Boston, Los Angeles, and San Francisco, for the time period between 2002 and 2008 controlling for service changes, seasonality and trends. He does not control for the lagged effects that changes in gasoline prices may have on transit ridership. Lane (2012) aims to answer the same question as Lane (2010) but also includes lagged effects of gasoline prices. He finds the cross-price elasticities to be negative for 13 cities but they correspond to different time lags of gasoline prices. Nine of the 13 cities have overall positive cross-price elasticities, when taking into account elasticities at other lagged times. Lane (2012) explains that this negative impact of lagged gasoline prices on bus ridership is probably because bus riders tend to be lower-income and more transit-dependent.

The time periods studied might contribute to cross-price elasticities varying considerably across studies. The values found in studies whose time period is during recession could be very different from those found in other studies as there might be other unobservable factors that affect transit ridership. During recession, people tend to have less disposable income so they likely move toward cheaper modes of transportation that are readily available; one of which is public transit. This, in effect, leads to an increase in transit ridership.

Although the studies have the same overall purpose, which is to estimate the impact of gasoline price changes on transit demand, each paper uses different independent variables or measures them differently. Some studies do not control for other variables when they study the relationship between gasoline prices and transit ridership (Jong et al. 1999; Haire and

Machemehl 2007; Currie and Phung 2007; Brand 2009; Maley and Weinberger 2009), while some others do. Estimating the effect of gasoline prices on bus ridership without controlling for other determinants may lead to an omitted variable bias. In the case where an omitted variable has an effect on bus ridership and is positively correlated with gasoline prices, we will overestimate the coefficient on gasoline prices.

Common covariates used in past studies are transit fares, service quantity and quality, trends, lagged variables, season dummies, and other economic variables (for example, unemployment rates, consumer price index, income level, etc.). Table 2 presents a summary of the independent variables used in previous studies. Among those variables that are under transit agencies' control, service quality has proven most effective in incentivizing customers to take public transit (Taylor and Fink 2003). Transit fare directly affects transit ridership, yet it does not change often. Litman (2004) notes that the effects of fare increase are different from those of a fare decrease. Customers tend to be more sensitive to price increases than to price decreases.

There are other factors affecting transit ridership that are not under agencies' control. Maley and Weinberger (2009) and Lane (2010) suggest that seasonal variation affects transit ridership. Summer ridership tends to be lower compared to the overall average (Doi and Allen 1986) because for close distances, people tend to walk and go on more vacations outside the cities.

The effect of gasoline price changes on transit ridership differs in the short-run compared to the long-run. According to previous literature, short-run is generally considered to be within two years or less; longer than two years is considered long-run. In the short-run, passengers tend to be less responsive (i.e. demand for public transit is less elastic) than in the long-run (Litman 2004) because it takes time to switch from one transportation mode to

another. For example, person A usually commutes via public transportation to work in the city and does not own a car. Even though gasoline prices fall significantly, she needs time and resources to switch from public transportation to driving. Contrary to Litman, Jong et al. (1999) concluded that, in general, short-run elasticities tend to be higher. Litman derives cross elasticities based on auto operating costs whereas Jong et al.'s cross elasticities are based on fuel prices. Litman (2004) notes that the decline in elasticity values in Jong et al.'s (1999) study is unique to fuel prices as motorists move towards more fuel-efficient vehicles when fuel becomes more expensive. Jong et al. does not include other explanatory variables. This may explain why their finding contradicts expectations. In addition to the differences in elasticities over time and across transit modes, size and population of the city also affect elasticity values. Demand for public transit is less elastic in large cities (bigger size and/or denser cities) due to a higher number of transit-dependent riders (Lane 2010; Litman 2004). Hence, it is more accurate to estimate the elasticities at a disaggregated level instead of a nationwide estimate or the average across cities.

Even though there might be immediate effects on transit ridership when gasoline prices change, evidence suggests that demand for public transit takes time to respond to the shock (Goodwin 1992; Chen et al. 2011; Lane 2012). This suggests that we should not leave out lagged gasoline prices in our study as it might affect the accuracy of our results; we would most likely underestimate the effects of changes in gasoline prices on transit demand. This also helps explain the wide range of cross-price elasticities found in various studies, because although most studies include lagged gasoline prices, they are not of the same time lags. Furthermore, Maley and Weinberger (2009), Nowak and Savage (2013) and Kennedy (2013) suggest a non-linear response of transit ridership to gasoline price changes; elasticities are larger at higher gasoline price range. Iseki and Ali (2014) find a higher cross-price elasticity when gasoline prices are higher than \$3 per gallon than when lower than \$3, suggesting the salience of price matters.

The literature reviewed above measures the relationship between gasoline prices and transit ridership using monthly observations and aggregate ridership at the mode level. Though answering a similar research question, this thesis attempts to contribute to the literature by evaluating the cross-price elasticities of bus ridership with respect to gasoline prices in the Twin Cities using monthly observations broken down into local and express rides. By separating local rides from express rides, it allows me to investigate these two types of bus service in the Twin Cities at a more fine-grained level. This would likely give me a better estimate of the relationship between gasoline prices and bus ridership. Local and express riders face different service quantity and service quality. Express buses are mostly operated during rush hours, tend to have fewer stops and longer route lengths, and are more likely to be on-time compared to local buses. Thus, riders of these two bus route types are expected to react to changes in gasoline prices differently.

3 Theory

In order to understand demand for public transit, I consider the choice among different modes of transportation. Consumers choose mode depending on their individual characteristics as well as mode-specific characteristics. They will maximize their utility by choosing the mode that gives them the highest utility.

Following Train (1986) qualitative choice model (also known as a discrete choice model), the description above can be written as:

$$n \text{ choose } i \in J_n \text{ if and only if } U(x_{in}, r_n) > U(x_{jn}, r_n), \text{ for all } j \in J_n, j \neq i \quad (1)$$

where n : the decision-maker,
 J_n : the choice set faced by individual n ,
 $U(x, r)$: the utility function,
 x_{in}, x_{jn} : all relevant characteristics of alternative i, j
 faced by individual n ,
 r_n : all relevant characteristics of individual n .

Equation (1) tells us that a decision-maker chooses the alternative that gives him the highest utility among all the alternatives he faces. Utility function $U(x_{in}, r_n)$ can be decomposed into two sub-functions: the deterministic portion that is observed, and another portion that represents all factors and aspects of utility that are unknown. We can write our utility function as:

$$U_{in} = U(x_{in}, r_n) = V(z_{in}, s_n, \beta) + \epsilon_{in} \quad (2)$$

where z_{in} : the observed characteristics of alternative i
 faced by individual n ,
 s_n : the observed characteristics of individual n ,
 β : a vector of parameters of the observed factors,
 $V(z_{in}, s_n, \beta)$: the observed portion of the utility function,
 ϵ_{in} : the unknown portion of the utility function.

We can see from equation (2) that two individuals who face the same choice set and the same observed factors might choose different modes depending on how small or how large their unknown portion, ϵ_{in} , is. In order to predict which mode a decision-maker chooses, we will look at the choice probabilities for each alternative. The probability that an individual n chooses alternative i is the limit of the proportion of times, as the number of times increases

without bound, that a decision-maker who faces the same alternatives as individual n , and with the same values of observed utility for each alternative, to choose alternative i . The choice probabilities can be written as:

$$P_{in} = \text{Prob}(U_{in} > U_{jn}, \text{ for all } j \in J_n \text{ and } j \neq i) \quad (3)$$

where P_{in} is the probability that individual n chooses alternative i .

Substituting (2) in (3), we get:

$$P_{in} = \text{Prob}(V_{in} + \epsilon_{in} > V_{jn} + \epsilon_{jn}, \text{ for all } j \in J_n \text{ and } j \neq i)$$

Rearrange:

$$P_{in} = \text{Prob}(\epsilon_{jn} - \epsilon_{in} < V_{jn} - V_{in}, \text{ for all } j \in J_n \text{ and } j \neq i) \quad (4)$$

where $V_{in} = V(z_{in}, s_n, \beta)$.

In equation (4), the difference $V_{jn} - V_{in}$ is observed but $\epsilon_{jn} - \epsilon_{in}$ is unknown, varying across individuals with the same observed components of utility. Since both ϵ_{jn} and ϵ_{in} are random variables, the difference $\epsilon_{jn} - \epsilon_{in}$ is also random. Thus, the right-hand side of equation (4) is a cumulative distribution: the probability that the random variable $\epsilon_{jn} - \epsilon_{in}$ is below the known value $V_{jn} - V_{in}$. This leads us to choosing between probit and logit probabilities in order to investigate this relationship.

In this theory section, I use the logit probabilities because they are readily interpretable. In this functional form, the random variable, ϵ_{in} , for all $i \in J$, is assumed to be extreme-value (or Gumbel) distributed, identically and independently distributed across alternatives and across individuals. Therefore, the probability that individual n chooses alternative i can be written as:

$$P_{in} = \frac{e^{V_{in}}}{\sum_{j \in J_n} e^{V_{jn}}}, \text{ for all } i \in J_n \quad (5)$$

The logit probability model ensures that each of the choice probabilities is between 0 and 1, probabilities for all alternatives sum to 1, and the relation of the choice probabilities is a sigmoid, or S-shaped. In addition, it also has the independence from irrelevant alternatives (IIA) property. This means that the ratio of two probabilities does not depend on any other alternatives other than the two alternatives being compared.

Assuming that the representative utility is linear in parameters, the deterministic portion of our utility function can be rewritten as:

$$V_{in} = \beta w(z_{in}, s_n) \quad (6)$$

β is a vector of parameters and w is a vector-valued function of the observed data.

We can then rewrite equation (5) as:

$$P_{in} = \frac{e^{\beta w_{in}}}{\sum_{j \in J_n} e^{\beta w_{jn}}}, \text{ for all } i \in J_n \quad (7)$$

where $w_{jn} = w(z_{in}, s_n)$.

The main question asked in this paper is how changes in gasoline prices affect transit ridership. The changes in gasoline prices will affect an individual's decision whether or not to drive, and thus may affect transit ridership. A fundamental property of logit models is that only differences in representative utility affect the choice probabilities, not their absolute levels. Thus, in binary choice situation like this, the logit probabilities can be simplified to:

$$P_T = \frac{e^{\beta w_T}}{e^{\beta w_T} + e^{\beta w_D}} = \frac{e^{\beta w_T}}{e^{\beta w_T} + e^{\beta w_D}} \times \frac{e^{-\beta w_T}}{e^{-\beta w_T}} = \frac{1}{1 + e^{\beta w_D - \beta w_T}} \quad (8)$$

where the subscripts T and D denote transit and driving, respectively. This property also implies that variables and changes in variables that do not vary over alternatives cannot affect the choice probabilities.

In addition to understanding how each observed variable affects choice probabilities, we could also study the extent to which these probabilities change in response to a change in some observed factor, i.e. elasticities. Elasticities are defined as the percentage change in one variable caused by one percentage change in another variable. Direct-elasticities, meaning the changes in choice probabilities are caused by their own observed factors, can be written as:

$$E_{iy_i} = \frac{\partial P_{in}}{\partial y_{in}} \times \frac{y_{in}}{P_{in}} \quad (9)$$

where E_{iy_i} : the direct-elasticity,

y_{in} : the observed factors of alternative i .

We have:

$$\frac{\partial P_{in}}{\partial y_{in}} = \frac{\partial(e^{V_{in}})(\sum_{j \in J_n} e^{V_{jn}})^{-1}}{\partial y_{in}} = \frac{\partial V_{in}}{\partial y_{in}} \times P_{in}(1 - P_{in}) = \beta_y P_{in}(1 - P_{in}) \quad (10)$$

Substituting (10) in (9):

$$E_{iy_i} = \beta_y P_{in}(1 - P_{in}) \times \frac{y_{in}}{P_{in}} = \beta_y y_{in}(1 - P_{in}) \quad (11)$$

The direction of E_{iy_i} depends on constant β_y . If the factor is desirable, β_y will be positive, and negative otherwise. In this paper, we are more interested in cross-elasticities of the probability of choosing transit with respect to gasoline prices and other observed factors.

This cross-elasticity can be written as:

$$E_{iy_j} = \frac{\partial P_{in}}{\partial y_{jn}} \times \frac{y_{jn}}{P_{in}} \quad (12)$$

where E_{iy_j} : the cross-elasticity,
 y_{jn} : the observed factor.

We have:

$$\frac{\partial P_{in}}{\partial y_{jn}} = \frac{\partial(e^{V_{in}})(\sum_{j \in J_n} e^{V_{jn}})^{-1}}{\partial y_{jn}} = -\frac{\partial V_{jn}}{\partial y_{jn}} \times P_{in} \times P_{jn} = -\beta_y P_{in} P_{jn} \quad (13)$$

Substituting (13) in (12):

$$E_{iy_j} = -\beta_y P_{in} P_{jn} \times \frac{y_{jn}}{P_{in}} = -\beta_y P_{jn} y_{jn} \quad (14)$$

Assume that the only factor that affects a decision-maker n whether or not to drive is gasoline prices, and the only other mode available is public transit. If gasoline prices increase, the decision-maker should have less incentive to drive, meaning that gasoline prices are not a desirable attribute. Thus, β_y should be negative. Because P_{jn} and y_{jn} are always positive, E_{iy_j} is also positive. This means that an increase in gasoline prices has a positive effect on transit ridership.

In order to understand how express and local bus riders might react to changes in gasoline prices differently, let's assume that a decision-maker faces two choices: driving and local buses, or driving and express buses. The magnitude of the cross-elasticities for the two scenarios will be different if at least one of β_y , P_{jn} , and y_{jn} is different. Local and express bus riders likely take buses for different purposes. Express riders are more likely to take buses to commute to work compared to local riders. This suggests that the probability of choosing public transit over driving (P_{jn}) will be different for express riders compared to local riders. Thus, the cross-elasticities are very likely different for the two types of riders, keeping everything else constant.

4 Data and Summary Statistics

4.1 Data

This study will be based on bus ridership data provided by Metro Transit, the largest transit agency in Minnesota. Large transit agencies like Metro Transit usually serve different neighborhoods, i.e. urban and suburban areas. Different neighborhoods most likely have different service frequencies and different groups of customers, i.e. commuters versus urban riders. Commuters are those who regularly use transit for commuting to work. Thus, they are more likely consistent riders. Urban riders tend to be less consistent in taking transit as they take shorter trips and for purposes other than commuting to work. I will separate these two groups of riders and perform different analyses for each of them. In order to isolate the effects of gasoline prices on bus ridership from one-time events that significantly affect bus ridership such as the opening of the new light rail at Metro Transit in June 2014 and the ongoing construction on Nicollet avenue in downtown Minneapolis, I will not include ridership from routes that are heavily affected by these events.

There are different types of gasoline, each of which has a different price. I use the monthly average of retail gasoline prices for my analyses. I convert these nominal prices to real prices using the semi-annual consumer price index for all urban consumers in Minneapolis-St. Paul. I obtain gasoline prices from the U.S. Energy Information Administration website (<http://www.eia.gov>) and the consumer price index for all urban customers from the U.S. Bureau of Labor Statistics (<http://www.bls.gov>).

Without services provided by the transit agency, riders would not be able to use public transportation. Other things being equal, the more service provided, the higher the probability that one takes transit. One way to measure service quantity at Metro Transit is the

number of in-service hours. While this is most of the time accurate, it can be biased. An increase in in-service hours can either mean Metro Transit adds more service to a particular route, or buses get delayed or detoured.

In addition to the biasedness in the variable itself, the in-service hours variable is also an endogenous predictor. When riders notice the increase in in-service hours which is most likely equivalent to higher frequencies, they might start riding and thus ridership increases. At the same time, if Metro Transit notices an increase in ridership for any particular route, it would likely increase service for that route. One way to solve this endogeneity caused by the simultaneity between ridership and service levels is to estimate using an instrumental variable. In my case, an instrumental variable is any variable that is highly correlated with in-service hours, and uncorrelated with ridership. Due to limited time and resources, I could not find an appropriate variable that fulfills this requirement. A possible proxy for service quantity is service frequency. At Metro Transit, in addition to each route having a different frequency, the service frequency for the same route is also different at different times of day. They usually provide higher service frequencies during rush hours. So, it is very difficult to quantify service frequency. Thus, I will use a different proxy, the number of workdays in a given month. A workday is the weekday (Monday through Friday) that is not a holiday. At Metro Transit, more service is provided on workdays compared to non-workdays. A portion of routes operated by Metro Transit do not offer service at all on non-workdays, and most of them provide limited service at a lower frequency on weekends. This suggests that more workdays within a given month is more or less equivalent to more service provided by Metro Transit. Thus, the number of workdays in a given month serves as a proxy for service levels.

In addition to service quantity, service quality is also a factor that can affect ridership. If buses are never on time or always show up at different times every day, the utility of taking transit will decrease and riders tend to shy away from taking buses. I will quantify service quality using the percent of the time buses are on time, i.e. on-time performance.

At Metro Transit, a bus is on time if it is between one minute early and five minutes late compared to the schedules.

Increase in transit fare should negatively affect ridership. Within the past decade, Metro Transit has changed its fare system only once in October 2008. Thus, instead of including the actual fare levels, I will create a dummy variable that identifies periods before and after the fare changes.

Commuters, i.e. those who take transit to work, tend to choose transit over other transportation modes. Most of express trips take the same amount of travel time as driving if not less. With a relatively high parking cost and/or low parking availability, customers who commute to work are better off taking transit than driving. This reflects in the level of employment in the metropolitan areas. Thus, employment level within transit-accessible areas will positively affect ridership. In other words, the unemployment level will negatively affect ridership, which is quantified by the unemployment rates within Minneapolis-St. Paul-Bloomington areas. The unemployment rates are available on the U.S. Bureau of Labor Statistics website (<http://www.bls.gov>).

For customers who can choose either to drive or take transit, traffic congestion will affect their decision. When traffic is bad, driving or taking transit will not be significantly different in terms of travel time. Sometimes buses, especially express routes, take less travel time because they have access to lanes that regular vehicles cannot use. It is difficult to quantify traffic congestion. So, I will use the monthly sum of the daily average number of cars passing automatic traffic recording stations placed in the Twin Cities as a proxy. Only the 56 stations that are active during the whole study period will be included to preserve consistency. I obtain the traffic counts from the reports on Automatic Traffic Recorders published on the Minnesota Department of Transportation website (<http://www.dot.state.mn>).

Customers behave differently at different times of the year. As Doi and Allen (1986) note, ridership tends to be lower in the summer. I will include month dummies to account for this seasonality.

There are a few other variables that I would like to include in my model that are not available or accessible. Car sharing service has been introduced to the Twin Cities for several years and the demand tends to increase with time, but the data are private. The closest proxy I find is the count of Google searches for Uber. Bike usage is another variable that would add to my model but I could not find a good source for the data in the Twin Cities. Since these two variables have an upward sloping trend over time, the time trend will be able to capture their effects, if any, on bus ridership. The time trend will also capture the effects of other variables that only vary annually such as population, income level, and passenger vehicles registered in the Twin Cities.

For car drivers, parking availability/cost will affect their decision-making, but the data on parking are not available to be collected. As a nature of this variable, it does not vary much as parking spaces do not change unless there is a big event that adds or eliminates parking spaces, and the costs remain relatively stable over time. So, it is most likely uncorrelated with other explanatory variables, which means its absence in my model will not introduce biasedness to my estimates.

4.2 Summary Statistics

The time period studied in this paper is from January 2008 to September 2015, which spans the period for which data are available. Table 3 summarizes the data including their means, standard deviations, minima, and maxima. Note that fare dummy variable is not included in this table. The fare dummy variable identifies the periods before and after the fare change

at Metro Transit in October 2008. All these variables have monthly frequency.

The monthly average of gasoline prices is 2.94 dollars per gallon with a standard deviation of 0.51. Gasoline prices have a minimum of 1.68 dollars per gallon in December 2008 and a maximum of 3.94 dollar per gallon in June 2008. This tells us that the average gasoline prices dramatically drop in the second half of 2008 from its six-year maximum to its six-year minimum (see figure 2). Gasoline prices from the literature seem to cover a larger range of values , from 1\$ per gallon to more than 4\$ per gallon, as some of the studies cover the periods where gasoline prices fluctuate dramatically (e.g. the Great Recession) or use nominal prices.

Monthly express bus ridership averages at 578.1 thousand rides with a standard deviation of 47.2 thousand. Monthly local bus ridership, which is about five times of that of express buses, averages at 2.7 million rides with a standard deviation of 0.18 million. Both time series are very seasonal (see figure 3), with higher values in the fall and lower in the summer. My ridership values are smaller than most of the literature as they tend to use city-wide or nation-wide values. Moreover, these data are very area- and date-range specific.

The number of workdays ranges from 18 to 23 days in a given month, which is similar to Bates' (1981) values. More often than not, there are 21 workdays in a given month (see figure 4). Unemployment rates are as low as 3% in October 2014 and as high as 8.3% in March 2010, with an average of 5.63% and a standard deviation of 1.45%. There is a significant increase in unemployment rate during the Great Recession from about 5% to 8%, and it has been following a downward sloping trend since then (see figure 5). The unemployment rates used in Nowak and Savage (2013) range from about 3.5% to 12% which spans a larger range. These unemployment rates are for Chicago, whereas my values are for the Twin Cities. As unemployment rates are area- and date-range specific, they are expected to have different values.

Traffic counts average at about 2.0 million vehicles passing the 56 automatic traffic recording stations within a given month with a standard deviation of 0.1 million. This time series is very seasonal (see figure 6) and tends to have the lowest values in the winter. On average, buses are on-time 87.9% of the time, with the worst performance at 79.8% and the best performance at 91.6% (see figure 7). These two variables, traffic count and on-time performance, do not appear in other studies so I cannot compare them. The difference in values and variables used in the estimates will likely lead to a slightly different value of the cross-price elasticity of bus ridership with respect to gasoline prices.

5 Main Analysis

5.1 Ordinary Least Squares

I begin by estimating the relationship between bus ridership and gasoline prices with an Ordinary Least Squares estimator while controlling for other factors including traffic, unemployment rates, bus fare change, on-time performance, number of workdays in month, and seasonality. I then decide on the lag structure of gasoline prices. The literature does not agree on what the lag structure of gasoline prices should be. So, in order to identify the lag structure I start by over-specifying it up to 11 month lags, and then eliminate those that are statistically insignificant.

5.1.1 Local Routes

Since my data are time series, unit roots might be a problem. Thus, before regressing, I check whether my dependent variable has unit roots using Augmented Dickey-Fuller (ADF) test. I fail to reject the null hypothesis of the ADF test that the time series has unit roots, i.e. it is nonstationary. I check for nonstationarity for all of quantitative explanatory variables. Gasoline prices and unemployment rates are nonstationary. I then run the regression with my

dependent variable and all the explanatory variables, and check whether the residuals of the regression are also nonstationary. The residuals are stationary, which suggests that my time series, which in this case is local ridership, is Engle Granger cointegrated with gasoline prices and unemployment rates. Hence, I can estimate the relationship between gasoline prices and local bus ridership in levels. The regression equation that I end up with for local bus routes is:

$$\begin{aligned} \log(\text{Local Rides})_t &= \beta_0 + \beta_1 \log(\text{Gas Prices})_t + \beta_2 \text{L.log}(\text{Gas Prices})_t \\ &+ \beta_3 \log(\text{Traffic Count})_t + \beta_4 \text{Number of Workdays}_t \\ &+ \beta_5 \text{Unemployment Rates}_t + \beta_6 \text{Fare Dummy}_t \\ &+ \beta_7 \text{On-time Performance}_t + \sum_{i=1}^{12} \beta_{i+7} (\text{Month}==i) + \epsilon_t \end{aligned}$$

For an OLS estimator to be the “Best Linear Unbiased Estimator,” there cannot be multicollinearity among explanatory variables. I check for multicollinearity by calculating Variance Inflation Factors (VIFs). VIFs calculate the extent to which a given explanatory variable can be explained by all other explanatory variables. My cutoff is VIF equals to five; if VIF is higher than five, then the variables are highly correlated. As expected, gasoline prices variable is very multicollinear with its lag. To keep both current effect and one month lagged effect of changes in gasoline prices, I generate a new variable which is the moving average of gasoline prices with two window periods. Thus, my regression equation becomes:

$$\begin{aligned} \log(\text{Local Rides})_t &= \beta_0 + \beta_1 \log(\text{Gas Prices}) \text{ 2-month Moving Average}_t \\ &+ \beta_2 \log(\text{Traffic Count})_t + \beta_3 \text{Number of Workdays}_t \\ &+ \beta_4 \text{Unemployment Rates}_t + \beta_5 \text{Fare Dummy}_t \\ &+ \beta_6 \text{On-time Performance}_t + \sum_{i=1}^{12} \beta_{i+6} (\text{Month}==i) + \epsilon_t \end{aligned}$$

Table 4 shows the OLS regressing results for local ridership. The two-month moving average of gasoline prices significantly affects local bus ridership with a coefficient of 0.137, and this is within the range of cross elasticities found in the literature (-0.17 to 0.88). In other words, ten percentage increase in gasoline prices this month is associated with 1.37 percent increase in local bus ridership in total for this month and next month.

One percentage point increase in unemployment rates is associated with 1.7% decrease in local bus ridership. This agrees with my hypothesis that unemployment rates are inversely related with bus ridership. One extra workday within a given month is associated with 1.7% increase in local bus ridership. Metro Transit usually provides more frequent service on workdays compared to non-workdays. So, a higher number of workdays means more service is provided, and thus more riders.

Traffic count does not have a statistically significant effect on local bus ridership. I expect a positive relationship between the two because if traffic is bad, it might take essentially the same amount of travel time to get from one place to another, whether one drives or takes transit, and transit is much cheaper. Local buses do not usually operate on the freeways and many auto traffic recorders are on the free ways. This might explain the insignificant relationship between traffic count and local bus ridership. Fare dummy also appears statistically insignificant. Metro Transit changed its fare system in October 2008, early in my study period. So, there are many more periods after than before fare change.

On-time performance does not statistically significantly affect local bus ridership. I expect a positive and significant relationship between the two. This insignificant relationship might be because most of the high ridership local routes provide high frequency service, which means it runs at least every 15 minutes during rush hours. The on-time performance might not matter that much because the expected waiting time would not dramatically

change when buses are late. It could also be because this on-time performance measurement is calculated agency-wide.

5.1.2 Express Routes

I check for unit roots for express bus ridership using Augmented Dickey Fuller test. The time series does not have unit roots. I estimate the relationship between gasoline prices and express bus ridership in levels. After eliminating the statistically insignificant lags of gasoline prices, the regression equation that I end up with for express bus routes is:

$$\begin{aligned} \log(\text{Express Rides})_t = & \beta_0 + \beta_1 \log(\text{Gas Prices})_t + \beta_2 L.\log(\text{Gas Prices})_t \\ & + \beta_3 L2.\log(\text{Gas Prices})_t + \beta_4 L3.\log(\text{Gas Prices})_t \\ & + \beta_5 L4.\log(\text{Gas Prices})_t + \beta_6 \log(\text{Traffic Count})_t \\ & + \beta_7 \text{Number of Workdays}_t + \beta_8 \text{Unemployment Rates}_t \\ & + \beta_9 \text{Fare Dummy}_t + \beta_{10} \text{On-time Performance}_t \\ & + \sum_{i=1}^{12} \beta_{i+10} (\text{Month}==i) + \epsilon_t \end{aligned}$$

I check for multicollinearity using variance inflation factors. Again, contemporaneous and lagged gasoline prices are highly correlated. Hence, I generate a new variable which is the moving average of gasoline prices with five window periods. My regression equation can, then, be rewritten as:

$$\begin{aligned} \log(\text{Express Rides})_t = & \beta_0 + \beta_1 \log(\text{Gas Prices}) \text{ 5-month Moving Average}_t \\ & + \beta_2 \log(\text{Traffic Count})_t + \beta_3 \text{Number of Workdays}_t \\ & + \beta_4 \text{Unemployment Rates}_t + \beta_5 \text{Fare Dummy}_t \\ & + \beta_6 \text{On-time Performance}_t + \sum_{i=1}^{12} \beta_{i+6} (\text{Month}==i) + \epsilon_t \end{aligned}$$

Table 5 shows the OLS regression results for express routes. Overall, a ten percent increase in gasoline prices is associated with a 2.20 percent increase in express bus ridership. In other words, the cross-price elasticity of express bus ridership with respect to gasoline prices is 0.220. This falls in the range of cross-price elasticities found in the literature (-0.17 to 0.88). It makes sense that the effect of changes in gasoline prices on express riders is somewhat larger than on local riders. Express riders are often commuters who take transit to get from a suburban area to work in the urban area, which usually is a longer trip than that taken by local riders. Moreover, in the Twin Cities, express riders tend to be less transit-dependent compared to local riders as they are more likely to have access to other mode of transportation, i.e. driving, and are associated with higher income. Thus, express riders are more sensitive to changes in gasoline prices than local riders (Mattson 2008).

Unemployment rates have a statistically significant effect on express bus ridership. A one percentage point increase in unemployment rates is associated with 2.4 percent decrease in express bus ridership. This inverse relationship agrees with my hypothesis. Number of workdays also significantly affects express bus ridership. An additional workday in a given month is associated with 4.8 percent increase in express bus ridership. This relative high coefficient compared to that of local ridership makes sense. Most of the express routes provide highly reduced service on non-workdays, if any. So, most of the ridership is from workdays. Thus, an additional workday relatively highly affects the total monthly ridership.

The fare change dummy is statistically significant and positive, which does not conform with my hypothesis. It is very unlikely that the increase in fare attracted more riders so this might be some factors that are not explained in my model. Moreover, it might also be because of the lack of periods before the fare change. Traffic count does not significantly affect express ridership. Again, this might be because the measurement is not specific to the particular streets express buses operate on.

To validate the models for both local and express ridership, I look at their residuals. Figures 8 and 9 show the partial autocorrelation plot of residuals for local and express routes, respectively. We can clearly see in figure 8 that the residuals for local ridership model are autocorrelated at lag 1, and slightly at lag 4. To confirm there is autocorrelation in my model for local bus routes, I test using the Durbin Watson test. I reject the null hypothesis of no autocorrelation. This suggests that OLS does not do a good job in estimating given the nature of my data. For express routes (figure 9), most of the partial autocorrelations are within the 95% confidence intervals, but there is some autocorrelation at lag 4. I check for the autocorrelation using the Durbin Watson test, and the result is inconclusive. The inconclusive result of the Durbin Watson test and the partial autocorrelation plot of residuals does not suggest a clear indication of whether or not there is autocorrelation. To ensure this does not affect my estimates I will still take into account the possible autocorrelation in my model for express ridership.

One of the main assumptions of an ordinary least squares estimator to be the best linear unbiased estimator is that the residuals are normally distributed with mean zero and there is no autocorrelation. If the residuals are autocorrelated and the autocorrelation is ignored, the standard errors will be underestimated, which leads to unreliable conclusion on coefficients' statistical significance. In the presence of autocorrelation, OLS is not efficient as it no longer has the lowest variance among linear estimators. To deal with the autocor-

relations, I can either include autoregressive process of the dependent variable in my OLS models or use a different estimator that can deal with this problem. In this paper, I will use Autoregressive Integrated Moving Average estimator, which estimates the coefficients via maximum likelihood instead of least squares as in the case of ordinary least squares estimator.

5.2 Autoregressive Integrated Moving Average

5.2.1 Introduction to ARIMA with Exogenous Regressors

To introduce ARIMAX models, I will base my definition on Hyndman's online texts which can be found at <https://www.otexts.org/fpp>.

Autoregressive Models

An autoregressive (AR) model is a regression with the dependent variable's past values as its regressors. An AR model with order p , i.e. $AR(p)$, can be written as:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + e_t$$

where c is constant, p non-negative integer, and e_t white noise.

Moving Average Models

A moving average (MA) model regresses on the past errors terms instead of the values themselves. A MA model with order q , i.e. $MA(q)$, can be written as:

$$y_t = c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q}$$

ARIMA models

ARIMA stands for autoregressive integrated moving average. ARIMA model incorporates both autoregressive and moving average models, and the “integrated” refers to differencing in case of non-stationary data, i.e. there is trend. An ARIMA(p, d, q) model, where p is the order of autoregressive, d the degree of first differencing, and q the order of moving average, can be written as:

$$(1 - \phi_1 L - \dots - \phi_p L^p)(1 - L)^d y_t = c + (1 + \theta_1 L + \dots + \theta_q L^q) e_t$$

where L is a lag operator such that $Ly_t = y_{t-1}$. The equation above clearly separates the autoregressive, first differencing, and moving average parts of the model. If $p = 0$, then $(1 - \phi_1 L - \dots - \phi_p L^p) = 1$ corresponding to the model ARIMA($0, d, q$). The same is applied to the first differencing and moving average parts of the model.

ARIMA models with Exogenous Regressors

The models I have discussed so far are univariate time series models. The ARIMAX model, (also known as dynamic regression model) is an ARIMA model that allows inclusion of other information, i.e. covariates as the “X” refers to, in order to predict or estimate the response variable. An ARIMAX model with p orders of autoregressions, d orders of first differences, q orders of moving average, and n predictors can be written as:

$$(1 - \phi_1 L - \dots - \phi_p L^p)(1 - L)^d y_t = c + \sum_{i=1}^n \beta_i X_i + (1 + \theta_1 L + \dots + \theta_q L^q) e_t$$

5.2.2 Local routes

To identify an appropriate ARIMA model, I start by looking at the partial autocorrelation plot of my dependent variable. Figure 10 shows a partial autocorrelation plot for local

ridership. We can see that after lag 5, most of the partial autocorrelations at larger lags are essentially zero. This suggests that the order of the autoregressive process is five. I fit the same specification as in the case of OLS but with autoregressive processes. I simplify my model by dropping the autoregressive processes that are not statistically significant. The model specification I end up with is ARIMA(4,0,0).

$$\begin{aligned}
 (1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \phi_4 L^4) \log(\text{Local Rides})_t = c \\
 + \beta_1 \log(\text{Gas Prices}) \text{ 2-month Moving Average}_t \\
 + \beta_2 \log(\text{Traffic Count})_t + \beta_3 \text{Unemployment Rates}_t \\
 + \beta_4 \text{Number of Workdays}_t + \beta_5 \text{On-Time Performance}_t \\
 + \beta_6 \text{Fare Dummy}_t + \sum_{j=1}^{12} \beta_{6+j} (\text{Month}==j) + e_t
 \end{aligned}$$

Before presenting the regression result from this model specification for local ridership, I validate the model by checking whether there is autocorrelation in the residuals. Figure 11 shows the partial autocorrelation plot of residuals from the ARIMA model for local ridership. We can see that autocorrelation is no longer an issue; up to lag 15, all partial autocorrelations are within the 95% confidence interval box.

Table 6 shows the regression result from this model specification for local ridership. The results are very similar to the OLS regression results I present earlier in terms of magnitude and statistical significance. Overall, ten percent increase in gasoline prices is associated with 1.39 percent increase in local ridership (significant at 1% significance level). In other words, the cross-price elasticity of local bus ridership with respect to gasoline prices is 0.139 compared to 0.137 found using an OLS estimator. Again, this aligns with the cross elasticities found in the literature which range from -0.17 to 0.88.

Unemployment rates and number of workdays also statistically significantly affect local bus ridership. One percentage point increase in unemployment rates is associated with 1.7 percent decrease in local bus ridership, which is the same as the OLS estimate. An additional workday in a given month is associated 1.7 percent increase in local bus ridership in that month, which is also exactly the same as that found with an OLS estimator.

Traffic count, fare dummy, and on-time performance are not statistically significant which is the same as estimating with an OLS estimator. Overall, the ARIMA regression results in terms of magnitudes and significance of coefficients are very similar to those found with the OLS estimator. The contribution of an ARIMA estimator in terms of giving a more robust estimate is that the residuals are no longer autocorrelated. Moreover, from the autoregressive processes coefficients, we can see that current ridership is significantly affected by its previous month values and four-month lagged values. Current month local ridership is positively correlated with its previous month ridership, and negatively correlated with its value from four months ago. The OLS estimator does not take this into account.

5.2.3 Express Routes

Although there is not a strong indication of autocorrelation, I will check to see if an autoregressive model would improve my model. I follow the same procedure to identify an appropriate ARIMA model for express ridership. Figure 12 shows the partial autocorrelation plot for express bus ridership. The plot does not provide a clear drop in partial autocorrelations, but it seems lag 6 is where the drop is even though some partial autocorrelations are outside the 95% confidence intervals at larger lags. Thus, the order I am starting with is six for the autoregressive process. I fit the same specification as in the case of OLS but with autoregressive processes. I simplify my model by dropping the autoregressive processes that are not statistically significant. The model specification I end up with is ARIMA(0,0,0). This is essentially the same as ordinary least squares regression. Without autoregressive and/or

moving average process, ARIMA estimates are identical to OLS estimates even though one is a maximum likelihood estimator and the other is a least squares estimator.

6 Discussion

From my regression results, there is enough evidence to suggest a significant relationship between gasoline prices and bus ridership for both express and local bus routes. The cross elasticities with respect to gasoline prices of local and express bus ridership in the Twin Cities are 0.139 and 0.220, respectively. This aligns with my hypothesis as well as the literature. Both cross elasticities found are within the range of values found by other literature, which is from -0.17 to 0.88. Express riders are more sensitive to changes in gasoline prices as they tend to be less transit-dependent and are associated with higher income.

A policy implication derived from this study is to make it expensive to drive, which is to keep gasoline prices at a relatively high level or increase gasoline taxes in order to incentivize customers to switch to more environmentally friendly options, one of which is public transportation.

Unemployment rates and number of workdays in a given month are two other factors that significantly affect both local and express bus ridership. There is an inverse relationship between unemployment rates and bus ridership, and a positive relationship between number of workdays and bus ridership. These two findings align with my hypotheses as well as the literature. The magnitude of the coefficients on both unemployment rates and number of workdays are higher for express than local bus ridership. This makes sense because express riders are more likely commuters who take public transit daily to work in downtown, so a higher portion of express rides would be affected by the change in unemployment rates compared to local bus rides which customers likely take for other purposes. Express bus

routes are mainly operated on workdays. So, the ratio of service hours during workdays to non-workdays is much higher than that of local bus routes. Thus, the higher effect of an additional workday on express compared to local routes is expected.

Traffic counts do not have a statistically significant effect on bus ridership. This variable, which is the proxy for traffic congestion, is very station specific. So, it is not surprising that it did not show statistical significance across route types as the counts from those stations may not reflect the actual traffic that both route types face.

The fare dummy variable, used in this study to identify the periods before and after the fare change in October 2008 at Metro Transit, does not show any statistically significant impact on local bus ridership but there is some impact on express bus ridership. The sign of the coefficient is positive which does not agree with my hypothesis. In addition to the lack of periods before the fare change, a dummy variable may not reflect what actually happens and there may be other unobservable factors that contribute to the ridership. These unobservable factors might be from those variables that are omitted from the model (e.g. car ownership, population growth). Even though they unlikely affect the coefficient estimate of my main variable of interest which is gasoline prices, they might affect other explanatory variables' coefficients.

I expected on-time performance to significantly affect all types of ridership as it has proven most effective in terms of policy implications (Taylor and Fink 2003). But, it does not significantly affect bus ridership at all. On-time performance is measured agency-wide, and is not specific to either of the route type in this study. Moreover, ridership from many routes that Metro Transit operates are excluded because they are either seasonal, or significantly affected by one time event such as the Green Line opening and the ongoing construction on Nicollet avenue in downtown Minneapolis.

7 Conclusion

Overall, both contemporaneous and lagged gasoline prices have significant effect on local and express bus ridership as reflected in the coefficients of its moving average values. Unemployment rates negatively affect bus ridership, and number of workdays positively affect bus ridership. The fare dummy variable does not provide any insight into the relationship between the fare change and bus ridership within the time period studied. Traffic counts and on-time performance do not provide any statistically significant effect on bus ridership.

These findings are valuable in terms of policy implications for Metro Transit. With limited resources, they can make better adjustments to changes in gasoline prices. The significant coefficient on number of workdays provides evidence that an increase in service levels is associated with higher ridership. So, one way to increase ridership is to increase service provided. If the economy hurts employment level which causes unemployment rates to be high, Metro Transit should consider decrease service for routes that mostly serve commuters which are likely express bus routes.

For future research, a more route-type-specific proxy for traffic counts and on-time performance, instead of the sum or average of all, would help improve this model. We should also consider the possible non-linear relationship between changes in gasoline prices and bus ridership as suggested by Maley and Weinberger (2009), Nowak and Savage (2013), and Kennedy (2013). It could also be more robust by including all the variables that are hypothesized to contribute to the changes in bus ridership such as parking availability/cost, population growth, income levels, and car ownership.

Acknowledgements

The author would like to thank Sarah West, Joel Huting, and Gary Krueger for the invaluable advice and guidance. This research would not have been possible without the data from Metro Transit as well as the help from the Strategic Initiatives team with collecting them. Appreciation is also due to family and friends who were very supportive during the whole experience.

References

- [1] Agthe, Donald E., and R. Bruce Billings. "The impact of gasoline prices on urban bus ridership." *The Annals of Regional Science* 12, no. 1 (1978): 90-96.
- [2] American Public Transportation Association. "Rising Fuel Costs: Impacts on Transit Ridership and Agency Operations Survey Results." (2008).
- [3] Bates, John W. "A study of demand for transit use." (1981).
- [4] Blanchard, Christopher. "The Impact of Rising Gasoline Prices on US Public Transit Ridership." PhD diss., Duke University, 2009.
- [5] Brand, Daniel "Impacts of Higher Fuel Costs" *U.S. Department of Transportation*. (2009). Web. <http://www.fhwa.dot.gov/policy/ootp/innovation/issue1/impacts.hth>. Feb 16. 2015.
- [6] Chen, Cynthia, Don Varley, and Jason Chen. "What affects transit ridership? A dynamic analysis involving multiple factors, lags and asymmetric behaviour." *Urban Studies* 48, no. 9 (2011): 1893-1908.
- [7] Currie, Graham, and Justin Phung. "Exploring the impacts of fuel price increases on public transport use in Melbourne." In *29th Australasian Transport Research Forum, Gold Coast*. 2006.
- [8] Currie, Graham, and Justin Phung. "Transit ridership, auto gas prices, and world events: new drivers of change?." *Transportation Research Record: Journal of the Transportation Research Board* 1992, no. 1 (2007): 3-10.
- [9] Currie, Graham, and Justin Phung. "Understanding links between transit ridership and gasoline prices: evidence from the United States and Australia." *Transportation Research Record: Journal of the Transportation Research Board* 2063, no. 1 (2008): 133-142.
- [10] Doi, Masayuki, and W. Bruce Allen. "A time series analysis of monthly ridership for an urban rail rapid transit line." *Transportation* 13, no. 3 (1986): 257-269.
- [11] Goodwin, Phil B. "A review of new demand elasticities with special reference to short and long run effects of price changes." *Journal of transport economics and policy* (1992): 155-169.

- [12] Haire, Ashley R., and Randy B. Machemehl. "Impact of rising fuel prices on US transit ridership." *Transportation Research Record: Journal of the Transportation Research Board* 1992, no. 1 (2007): 11-19.
- [13] Haire, Ashley R., and Randy B. Machemehl. "Regional and modal variability in effects of gasoline prices on US transit ridership." *Transportation Research Record: Journal of the Transportation Research Board* 2144, no. 1 (2010): 20-27.
- [14] Hyndman, Rob J. "Forecasting: Principles and Practice." *Forecasting: Principles and Practice*. Web. 13 Feb. 2016. <https://www.otexts.org/fpp>.
- [15] Iseki, H., and Ali, R. "Net Effects of Gasoline Price Changes on Transit Ridership in US Urban Areas." (2014).
- [16] Jong, GC de, L. Biggiero, and P. Coppola. "Elasticity Handbook: Elasticities for prototypical contexts (Deliverable 5)." *TRACE, Costs of private road travel and their effects on demand including short and long term elasticities*. Europese Commissie, Brussel (1999).
- [17] Kemp, Michael A. *Bus service in San Diego: a study of patronage growth in the mid-1970s*. Urban Institute, 1981.
- [18] Kennedy, David. "Econometric models for public transport forecasting." (2013).
- [19] Lane, Bradley W. "The relationship between recent gasoline price fluctuations and transit ridership in major US cities." *Journal of Transport Geography* 18, no. 2 (2010): 214-225.
- [20] Lane, Bradley W. "A time-series analysis of gasoline prices and public transportation in US metropolitan areas." *Journal of Transport Geography* 22 (2012): 221-235.
- [21] Lee, Jaimin, Sangyong Han, and Chang-Woon Lee. "Oil Price and Travel Demand." (2009): 1-88.
- [22] Litman, Todd. "Transit price elasticities and cross-elasticities." *Journal of Public Transportation* 7 (2004): 37-58.
- [23] Maley, Donald W., and Rachel Weinberger. "Rising gas price and transit ridership." *Transportation Research Record: Journal of the Transportation Research Board* 2139, no. 1 (2009): 183-188.

- [24] Mattson, Jeremy Wade. *Effects of rising gas prices on bus ridership for small urban and rural transit systems*. Upper Great Plains Transportation Institute, North Dakota State University, 2008.
- [25] Morath, Eric. "Low Gas Prices Drove Down Transit Use, So Why Can't You Find a Seat on the Train?" WSJ. N.p., 6 Apr. 2016. Web. 10 Apr. 2016. <http://blogs.wsj.com/economics/2016/04/06/low-gas-prices-drove-down-transit-use-so-why-cant-you-find-a-seat-on-the-train/>.
- [26] Nowak, William P., and Ian Savage. "The cross elasticity between gasoline prices and transit use: Evidence from Chicago." *Transport policy* 29 (2013): 38-45.
- [27] Stover, V. W., and C-H. Christine Bae. "The impact of gasoline prices on transit ridership in Washington State. CD-ROM." *Transportation Research Board of the National Academies* (2011).
- [28] Taylor, Brian D., and Camille NY Fink. "The factors influencing transit ridership: A review and analysis of the ridership literature." *University of California Transportation Center* (2003).
- [29] Train, Kenneth. *Qualitative choice analysis: Theory, econometrics, and an application to automobile demand*. Vol. 10. MIT press, 1986.
- [30] Wang, George HK, and David Skinner. "The impact of fare and gasoline price changes on monthly transit ridership: empirical evidence from seven US transit authorities." *Transportation Research Part B: Methodological* 18, no. 1 (1984): 29-41.
- [31] Yanmaz-Tuzel, Ozlem, and Kaan Ozbay. "Impacts of gasoline prices on New Jersey transit ridership." *Transportation Research Record: Journal of the Transportation Research Board* 2144, no. 1 (2010): 52-61.

Figures and Tables

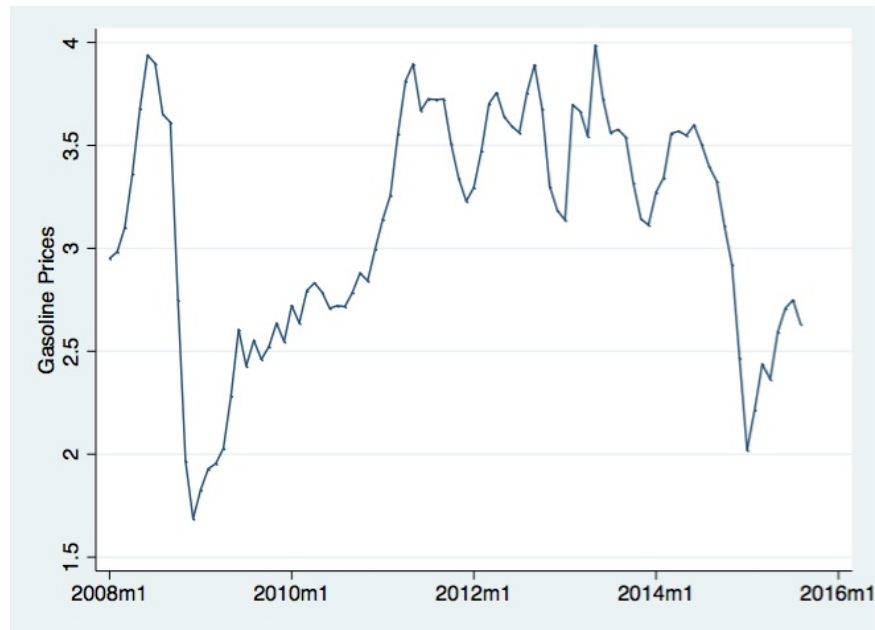


Figure 1: Time Series Plot of Nominal Gasoline Price
U.S. Energy Information Administration

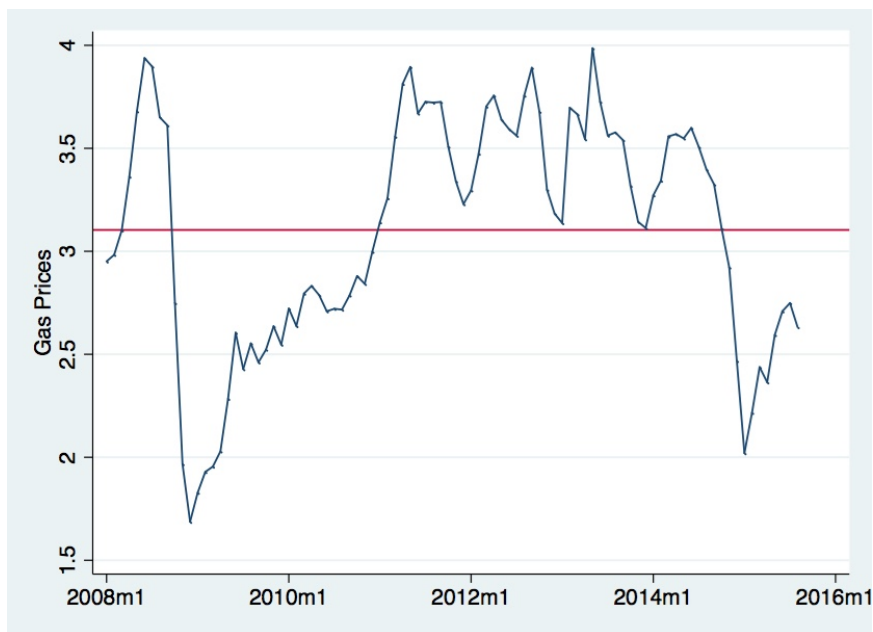


Figure 2: Time Series Plot of Real Gasoline Price
U.S. Energy Information Administration and U.S. Bureau of Labor Statistics

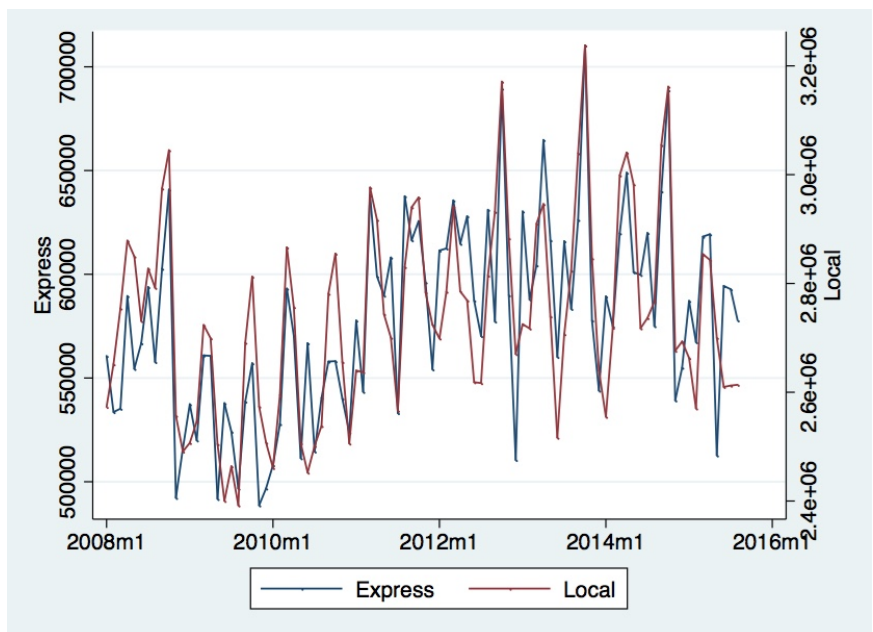


Figure 3: Partial Autocorrelation Plot of Residuals - Local Routes
Metro Transit, Minnesota

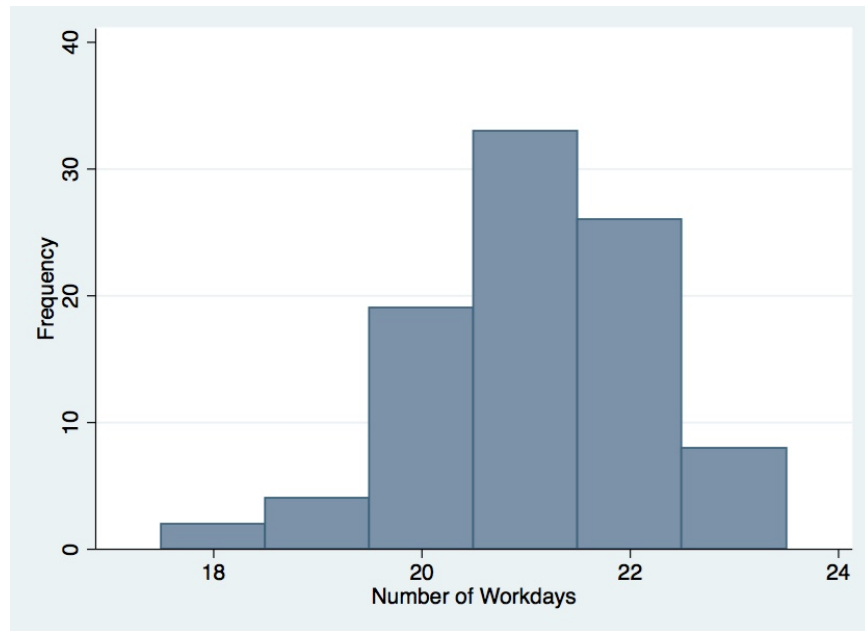


Figure 4: Partial Autocorrelation Plot of Residuals - Local Routes
Metro Transit, Minnesota

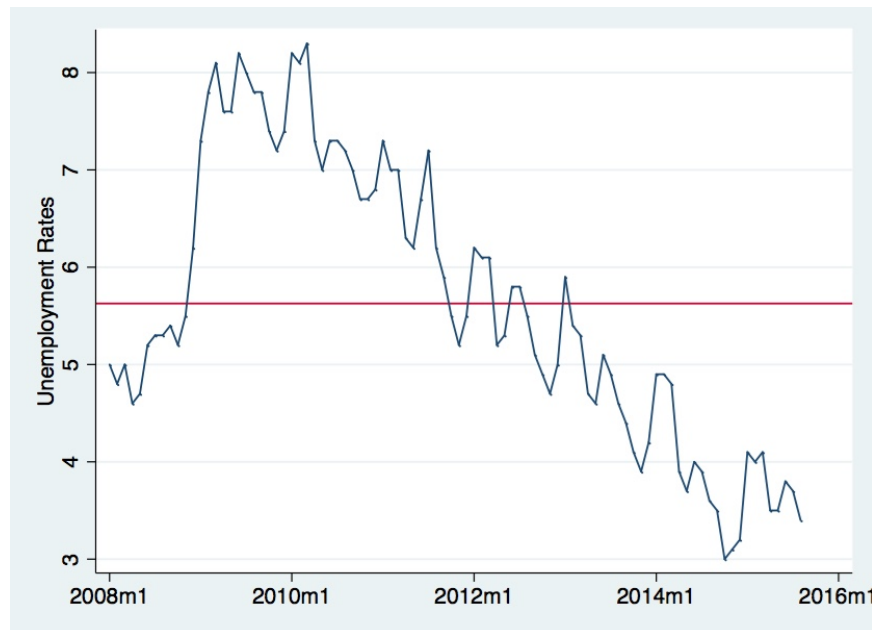


Figure 5: Partial Autocorrelation Plot of Residuals - Local Routes
U.S. Bureau of Labor Statistics website

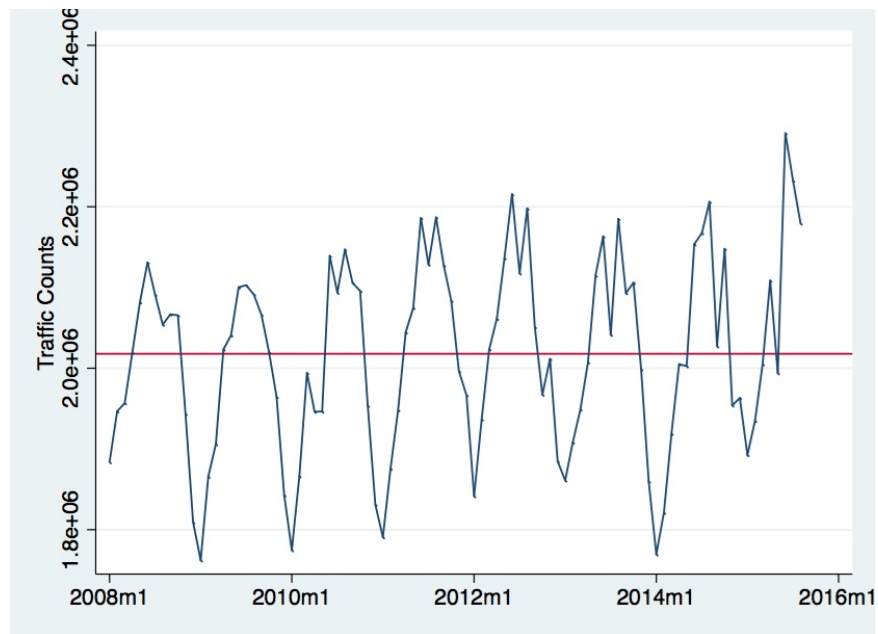


Figure 6: Partial Autocorrelation Plot of Residuals - Local Routes
Minnesota Department of Transportation

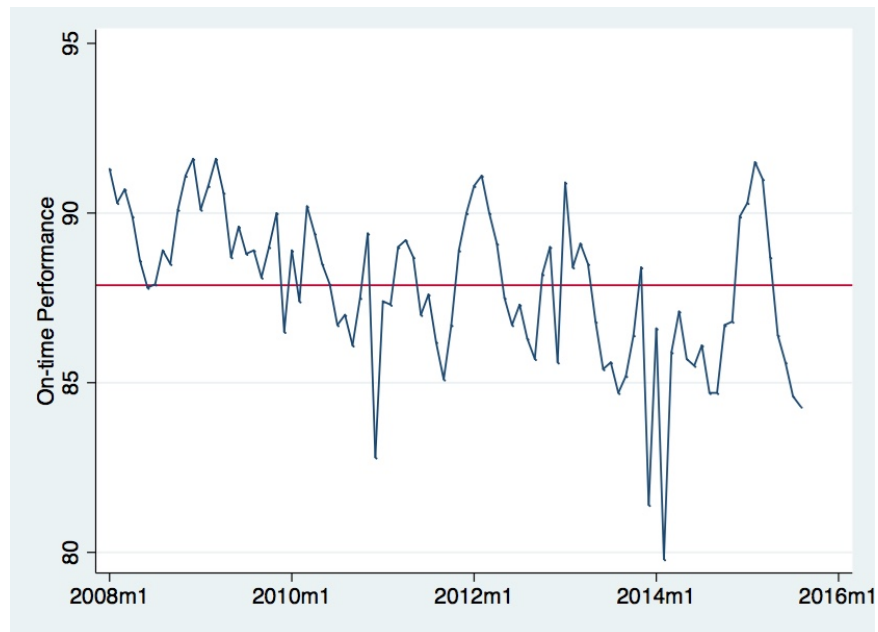


Figure 7: Partial Autocorrelation Plot of Residuals - Local Routes
Metro Transit, Minnesota

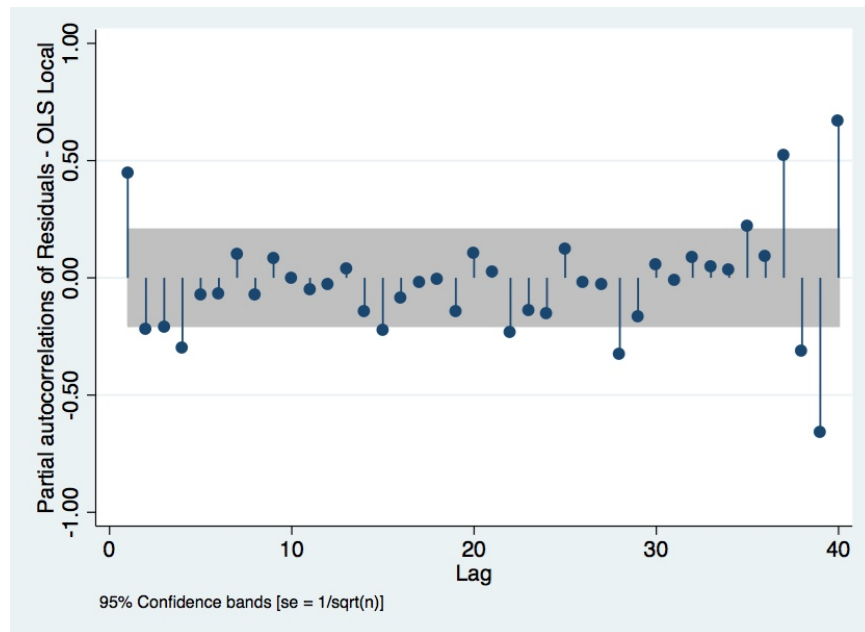


Figure 8: Partial Autocorrelation Plot of Residuals - Local Routes

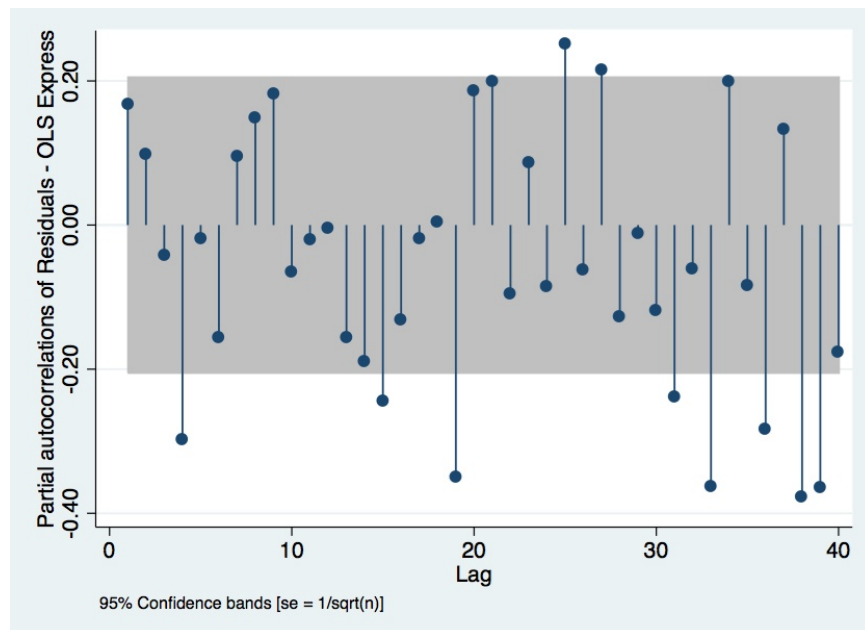


Figure 9: Partial Autocorrelation Plot of Residuals - Express Routes

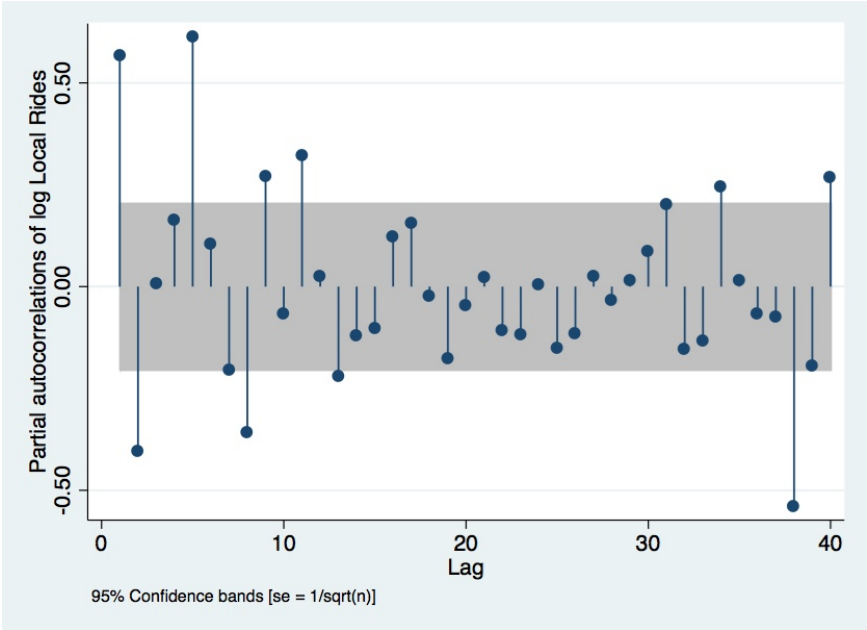


Figure 10: Partial Autocorrelation Plot of Local Ridership

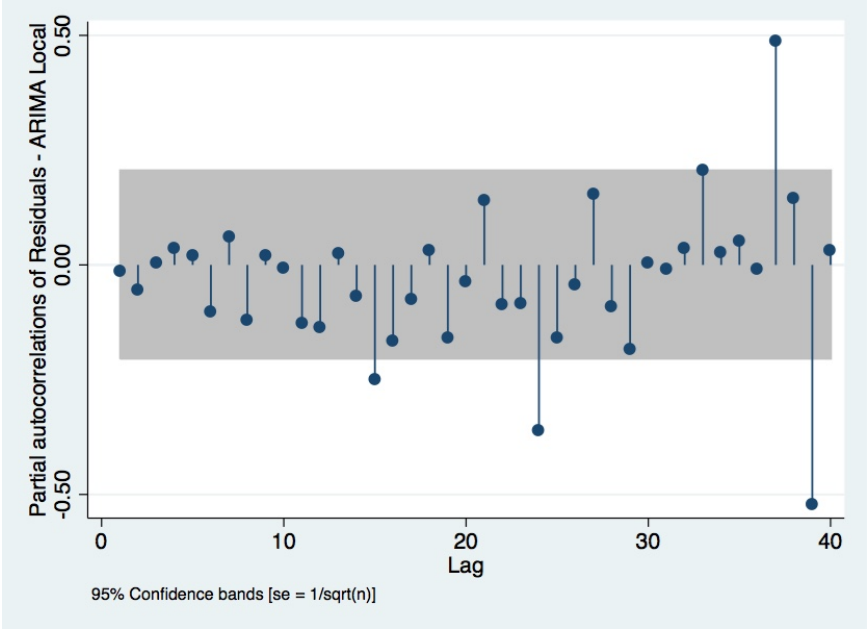


Figure 11: Partial Autocorrelation Plot of ARIMA Residuals - Local

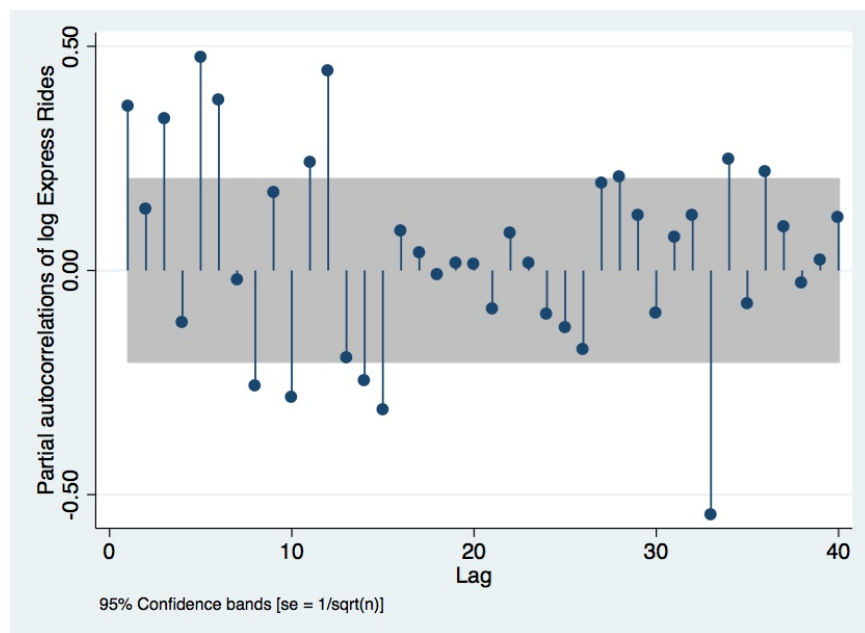


Figure 12: Partial Autocorrelation Plot of Express Ridership

Table 1: Buses' Cross Elasticities w.r.t Gasoline Prices from Previous Literature

Paper	Location	Date Range	Elasticities
Agathe & Billings (1978)	Tucson	1973-1976	0.42 (NS)
Bates (1981)	Atlanta	1973-1976	0.23 (NS)
Kemp (1981)	San Diego	1972-1975	0.29 (NS)
Wang & Skinner (1984)	7 U.S. cities	1970-1980	0.08-0.8 (SR)
Jong et al. (1999)	Europe	N/A	0.12 (LR)-0.16 (SR)
Currie & Phung (2006)	Melbourne, AU	2002-2005	0.35 (SR)
Currie & Phung (2007)	U.S.	1998-2005	0.04 (NS)
Currie & Phung (2008)	Melbourne, AU	2002-2005	0 (NS)
	Adelaide, AU	2004-2006	0.21 (NS)
	Brisbane, AU	2002-2006	0.25 (NS)
Mattson (2008)	Upper Midwest	1999-2006	0.08-0.50 (LR)
Blanchard (2009)	218 U.S. Transits	2002-2008	0.05-0.12 (NS)
Brand (2009)	U.S.	2007-2008	0.13 (NS)
Maley & Weinberger (2009)	Philadelphia	2001-2008	0.15-0.23 (NS)
Lane (2010)	9 U.S. Cities	2002-2008	-0.17-0.40 (NS)
Yanmaz-Tuzel & Ozbay (2010)	New York City	1998-2008	0.12-0.22 (NS)
Stover & Bae (2011)	Washington	2004-2008	0.17 (NS)
Lane (2012)	32 U.S. Cities	2002-2009	-0.09-0.88 (NS)
Nowak & Savage (2013)	Chicago	1999-2010	0.06-0.28 (NS)
Iseki and Ali (2014)	10 U.S. Cities	2002-2011	0.06 (SR)-0.08 (LR)

LR indicates Long-run; SR indicates Short-run; NS indicates Not Specified

Table 2: Independent Variables Used in Previous Literature

Paper	Independent Variables
Agathe and Billings (1978)	Vehicle miles, Energy crisis dummy, school in session dummy
Bates (1981)	Vehicle miles, Working days in month, Time trend, School weeks in month
Kemp (1981)	Fare, Working days in month, Time trend, Average speed, Waiting time, Service duration, Stop spacing, Route length, School days, Route dummies, Oil shortage dummy
Currie and Phung (2008)	Interest rates
Mattson (2008)	Monthly price variability, Service quantity, Trends
Blanchard (2009)	Level of service, Time trends, Monthly fixed effects
Haire and Machemehl (2010)	Fare, Vehicle revenue hours, Vehicle operated in maximum service, Consumer price index, Weekdays in month
Lane (2010)	Gas price standard deviation, vehicle revenue miles mode, Vehicle operated in maximum service mode, Time, Seasons
Chen et al. (2011)	Fare, Consumer price index
Lane (2012)	Monthly gas price variability, Service provision, Trend proxies
Nowak and Savage (2013)	Fare, Unemployment rates

Note: These variables are additional to Gasoline Prices

Table 3: Summary Statistics

Variables	Mean	St. Dev.	Min	Max
Real Gasoline Price (\$/gallon)	2.94	0.51	1.68	3.94
Express Ridership	578,114	47,166	488,624	710,333
Local Ridership	2,738,463	181,589	2,392,063	3,237,668
Number of Workdays per month	21.1	1.1	18.0	23.0
Traffic Counts per month	2,017,887	119,436	1,762,109	2,291,018
Unemployment Rate (%)	5.63	1.45	3.00	8.30
On-time Performance (%)	87.9	2.3	79.8	91.6

Table 4: OLS Regression Results for Local Ridership

VARIABLES	log(Local Rides)
log(Gas Prices) 2-month Moving Average	0.137*** (0.017)
Unemployment Rates	-0.017*** (0.002)
Number of Workdays	0.017*** (0.004)
log(Traffic Counts)	-0.043 (0.132)
Fare Dummy	0.012 (0.011)
On-time Performance	0.001 (0.002)
Constant	14.919*** (1.843)
Month Fixed Effects	YES
Observations	91
R-squared	0.885

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 5: OLS Regression Results for Express Ridership

VARIABLES	log(Express Rides)
log(Gas Prices) 5-month Moving Average	0.220*** (0.019)
Unemployment Rates	-0.024*** (0.003)
Number of Workdays	0.048*** (0.004)
log(Traffic Counts)	0.025 (0.137)
Fare Dummy	0.077*** (0.011)
On-time Performance	0.001 (0.002)
Constant	11.614*** (1.914)
Month Fixed Effects	YES
Observations	91
R-squared	0.920

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 6: ARIMA Regression Results for Local Ridership

VARIABLES	log(Local Rides)
log(Gas Prices) 2-month Moving Average	0.139*** (0.015)
Unemployment Rates	-0.017*** (0.002)
Number of Workdays	0.017*** (0.003)
log(Traffic Counts)	-0.020 (0.110)
Fare Dummy	-0.006 (0.012)
On-time Performance	-0.000 (0.001)
L.ar	0.451*** (0.120)
L2.ar	-0.169 (0.146)
L3.ar	-0.088 (0.180)
L4.ar	-0.316** (0.144)
Constant	14.668*** (1.500)
Sigma	0.018*** (0.002)
Month Fixed Effects	YES
Log Likelihood	237.571
Observations	91

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1