Economics H195B Senior Honors Thesis

Understanding Public Transit Ridership through Gasoline Demand: Case Study in San Francisco Bay Area, CA

Advised by: Professor Michael Anderson

Hansen Ouquan Sun

Department of Economics
University of California, Berkeley

April 2016

Abstract

The long trends of fuel price increase along with increasing transit ridership over the past few decades has been established by past transit research. However, in 2008, there has been a drastic drop in the prices of oil, which has had significant effects on consumption patterns of gasoline. The objectives of this paper are two folds; firstly to ascertain the nature of the gasoline-transit ridership relationship and secondly, to establish the empirical figures of cross demand gas price elasticity. The locality of San Francisco Bay Area was used to carry out such an analysis, with the BART, MUNI Light Rail and MUNI Bus services being analyzed. This study showed that gas prices indeed affect transit ridership in all forms, with elasticity values ranging between 0.0581-0.147, which are generally consistent with the aggregate national average of 0.12 for all modes of transport. Another key finding of the paper, contrary to past research, is that the elasticity values were higher for Bus and Light Rail compared to Heavy Rail and Light Rail being greater in value than Bus. Different regression models were also identified to be best suited for each transportation types. This paper also suggests that current transit authorities in the Bay Area do not adjust its service schedules in response to fluctuations of gasoline prices, and this could be important means of improving declining service standards. This paper also attempts to distinguish the substitution effect of public transit by analyzing the highway speed data in the same time period in the Bay Area.

Contents

1 Introduction

1.1 Research Question

This study aims to revisit past studies to ascertain and quantify the relationship between gasoline prices and its impacts on public transit ridership. The price of gasoline affects the total cost of travel incurred by consumers. Based on market observation and standard economic intuition, different travel options will become relatively more attractive at different gas prices. The working hypothesis for this paper, as is consistent with past papers, would be that an increase in gas prices increases the public transit ridership and conversely a decrease in gas prices decreases public transit ridership. In determining this relationship, the paper also estimates a cross demand elasticity of gasoline.

1.2 Motivation

Public transportation in the United States has become a significant part to solving economic, energy and environmental challenges that the US faces. The many benefits to public transportation such as fuel and monetary savings, congestion reduction and lessening carbon footprints are just the main benefits that the American Public Transportation Association (APTA) advertises.

According to APTA Statistics, over the 18 year period from 1995 to 2013, average public transportation ridership grew 37.2%, when the average population growth was 20.3%, which is a long term trend showing that more Americans are using public transportation. The highest recorded total public transit ridership, as of 2014, was also logged in 2013 (APTA). On average, public transit has risen across US over the years (Figure 1), with the highest number of public transport trips being made in 2013. This trend is also observed in this paper's area of study in the San Francisco Bay area, as shown in the ridership data charts below (Figure 2).

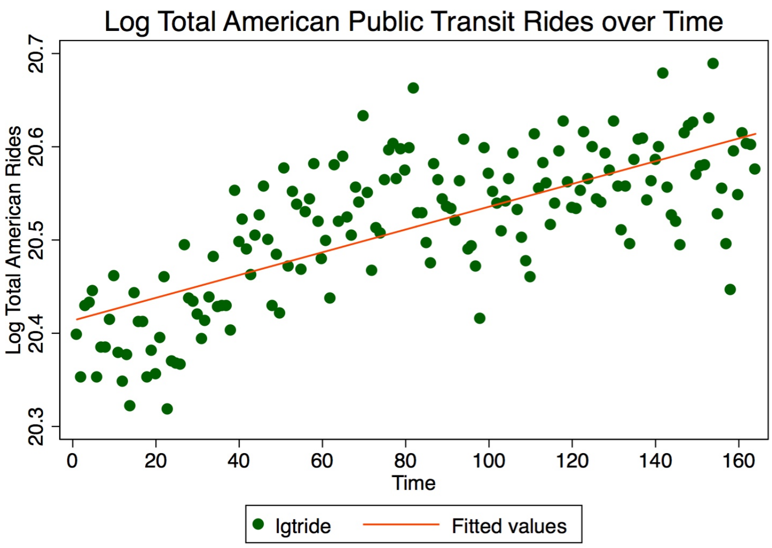

Figure 1: Trend of Total American Public Transit Ridership Over Period of Investigation

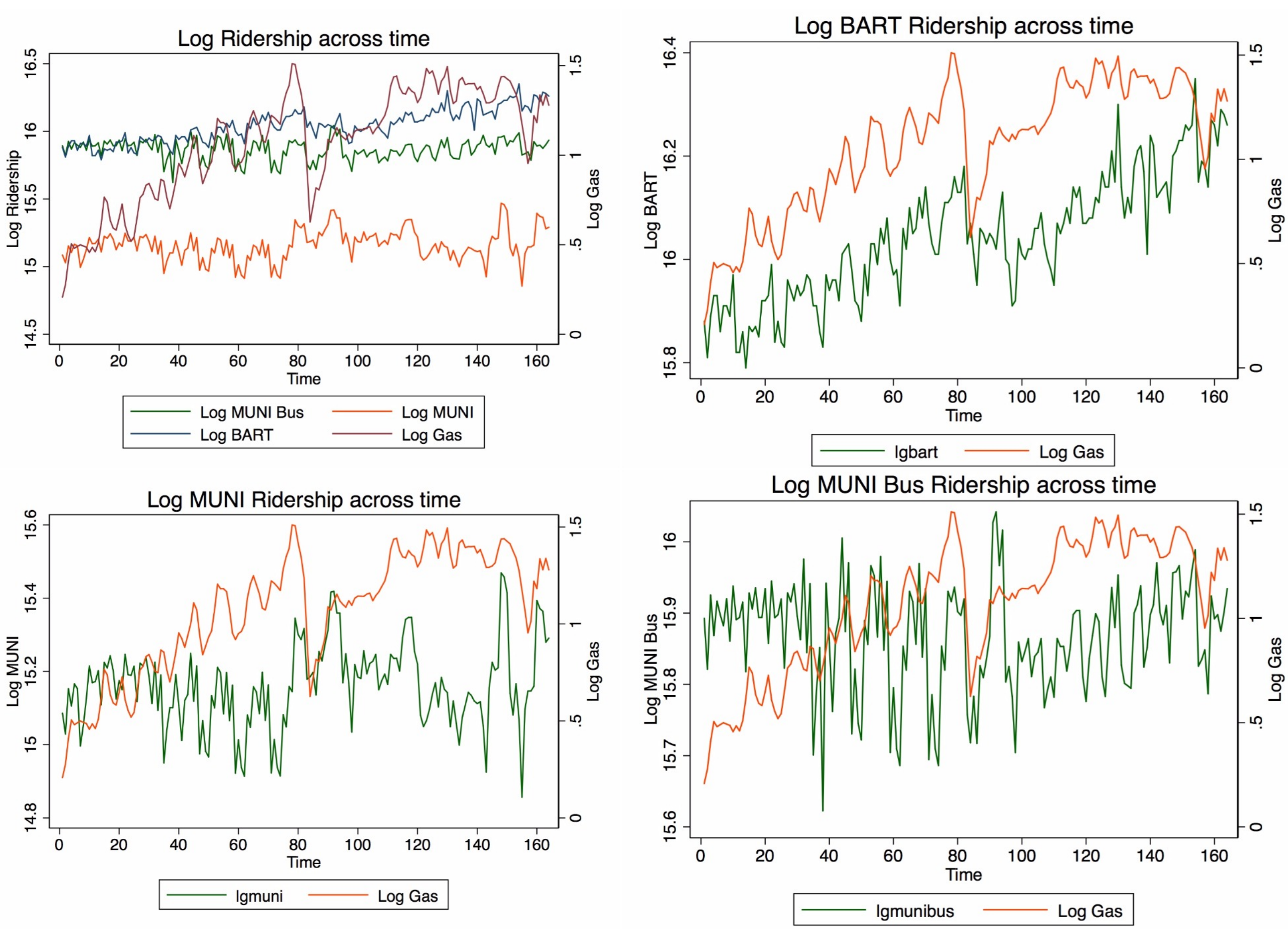

Figure 2: Trend of San Francisco Public Transit Ridership Over Period of Investigation

The significance of this research lies in isolating key factors of public transit ridership that is potentially useful for policy makers to consider on improving overall ridership for other bigger aims.

In particular, it would be particularly useful to determine what is the optimal level of subsidies and taxation to determine the amount of resources that are dedicated to the development of public transportation.

Other papers have also looked into factors that affect public transit ridership, such as external factors outside the control of the systems itself, including highway systems characteristics, regional geography, metropolitan economy and population characteristics(Taylor et al,2008). However, fundamentally, standard economic intuition would point us to the most basic component of all transportation systems, the price of gasoline.

Furthermore, since around 2008, there has been a steep drop in oil prices, with a culmination of many factors such as the onset of increased supply of oil domestically from sources like oil fracking, the decision of OPEC not to manipulate oil prices and a global weaker demand for oil. With such changes in the macroeconomic environment, past papers that investigated the trend at the height of gas prices around 2008 might be time outdated.

1.3 Choosing the Research Location

California is known to have higher gas prices than most other states in the US, with approximately 30 cents extra per gallon on average (Fueling California, 2016) The many unique features of California such as it being a 'fuel island' and not being linked directly to petroleum or crude oil supplies decrease the available supply compared to other states. Furthermore, there are many California regulations on the blends of petroleum that also cause the prices to increase. With the higher prices, these make the study of the gasoline price impacts more significant in California.

In determining the specific location within California to be investigated, a key selection criteria was the quality of service that is provided for by the public transportation network. Regardless of the city size, a transit system that does not provide adequate service to prove a viable alternative to driving cannot expect travelers to change modes, no matter the expense savings, as this disrupts their mobility (Haire & Machemehl, 2007). The converse is also true, where driving is no longer a necessity in localities like New York City, where the population is too heavily dependent on public transit. They are also demand inelastic with regards to changing gas prices because of the lack of viable alternatives available (Yanmaz-Tuzel & Ozbay (2010). Hence, as proposed in Yanmaz-Tuzel and Ozbay (2010), metropolitan areas with established transit systems and large numbers of automobile commuters are the best locations to test such a hypothesis of the subsitutability of public transport. Indeed, this is the main rationale for choosing the San Francisco Bay Area, where there possibly could be a substitution effect being observed due to the price changes of gasoline.

Figure 3 demonstrates the high percentage of residents residing in the San Francisco Bay Area that commute by public transit and Figure 4 demonstrate the low percentage of residents that commute to work by driving.

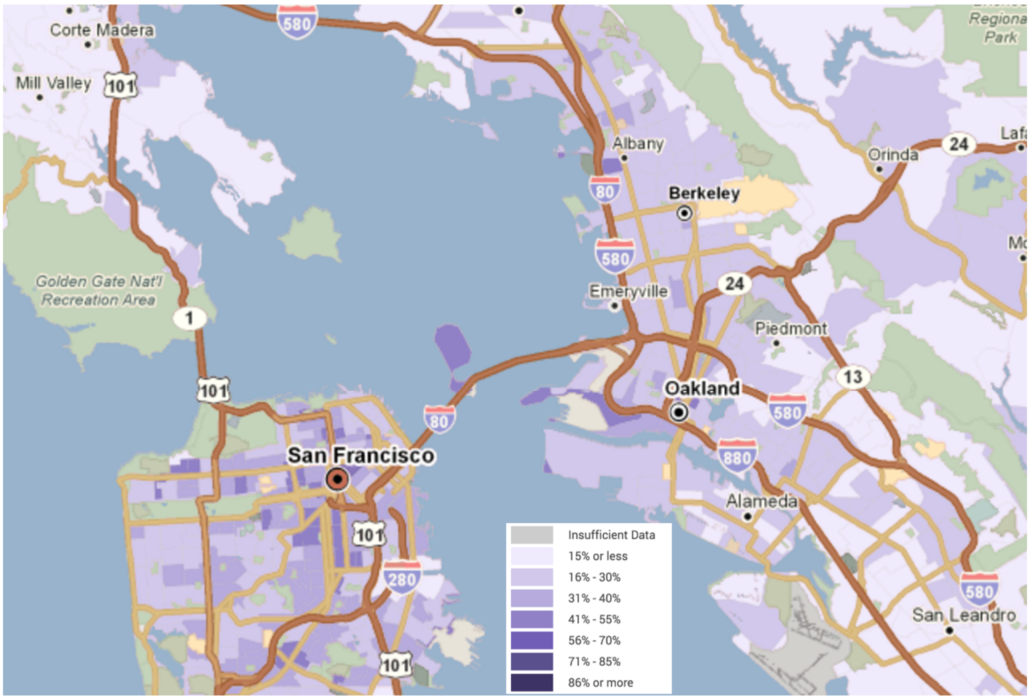

Figure 3: Map depicting % of population taking public transit in San Francisco Bay Area

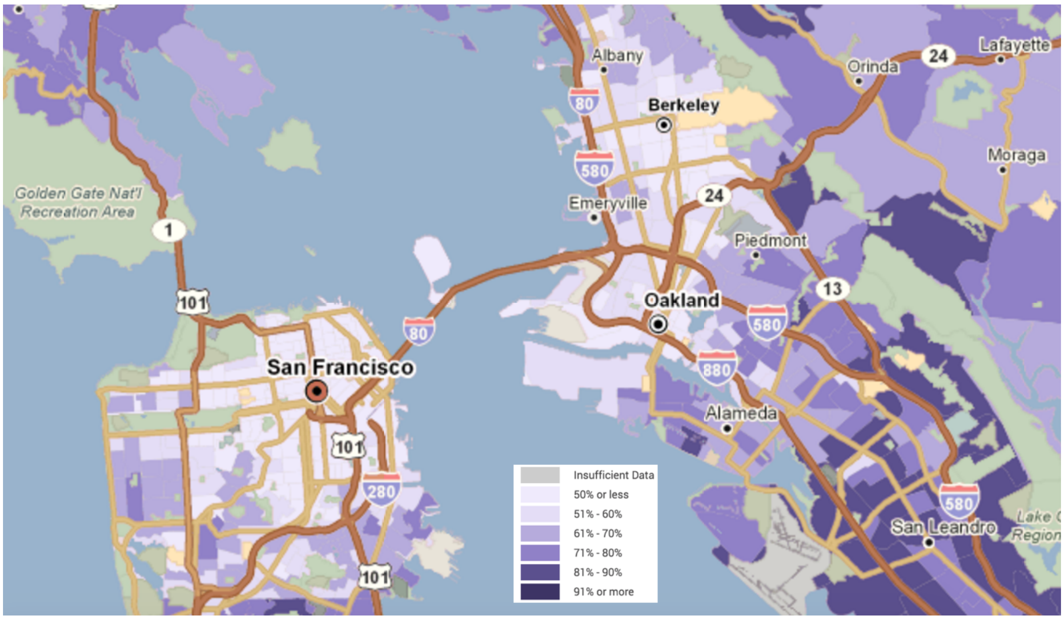

Figure 4: Map depicting % of population driving to work in San Francisco Bay Area

With a significantly high proportion of car ownership in the periphery of the regions, through the above map visualizations from PolicyMaps[1], we can see that there is a significant possibility of substitution effect happening within the San Francisco Bay Area.

[1] `www.policymaps.com`

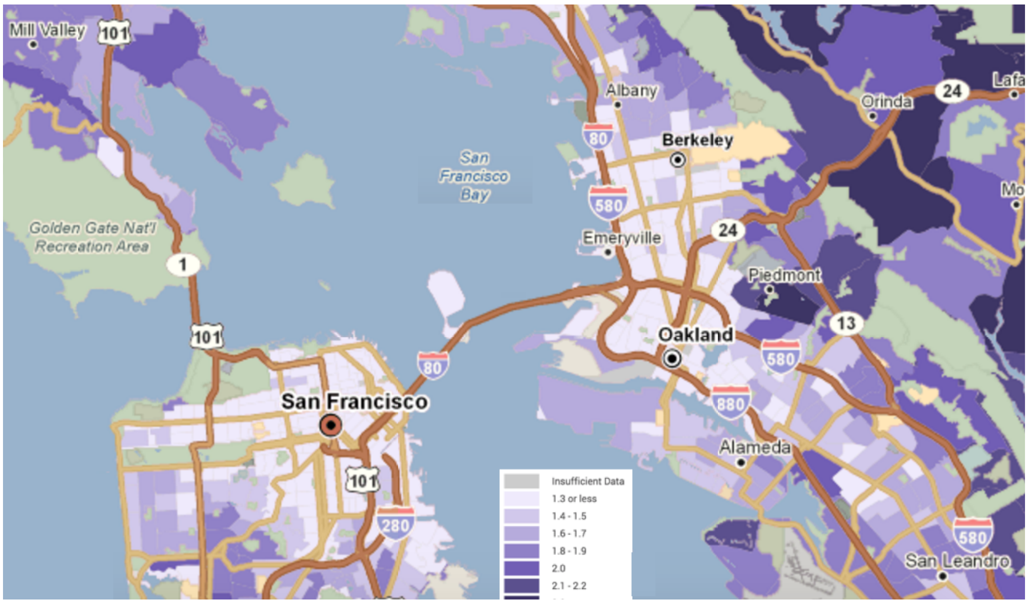

Figure 5: Map depicting average number of cars owned in the San Francisco Bay Area

In comparison, the geographical distributions seen in New York City (both Manhattan and Brooklyn) show substantial difference as compared to the San Francisco Maps. The regions are marked significantly lighter in Figure 6 depicts the low percentage of driving and the dark colors in Figure 7 depicting high proportion of population taking public transit.

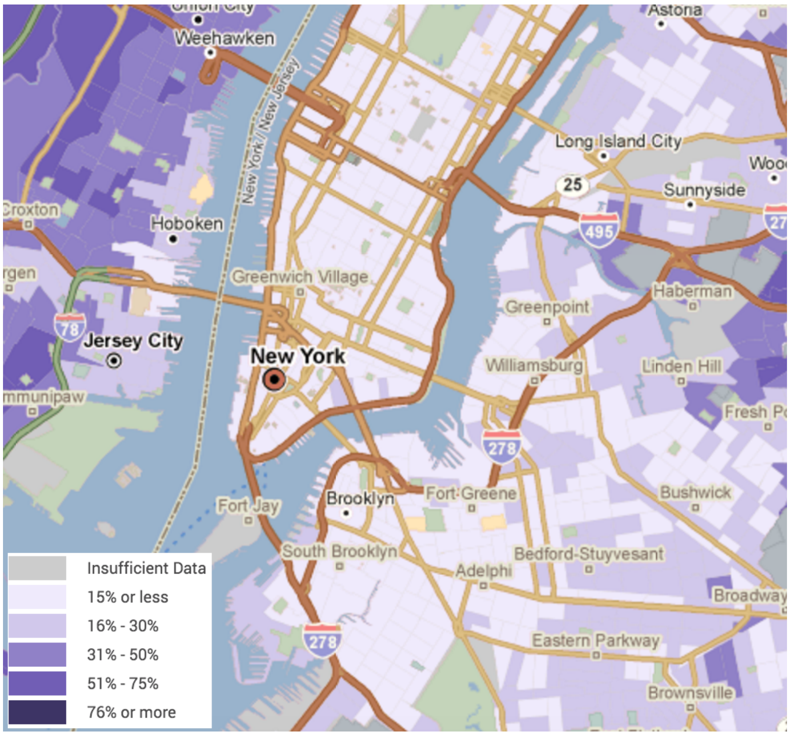

Figure 6: Map depicting % of population driving to work in New York City Area

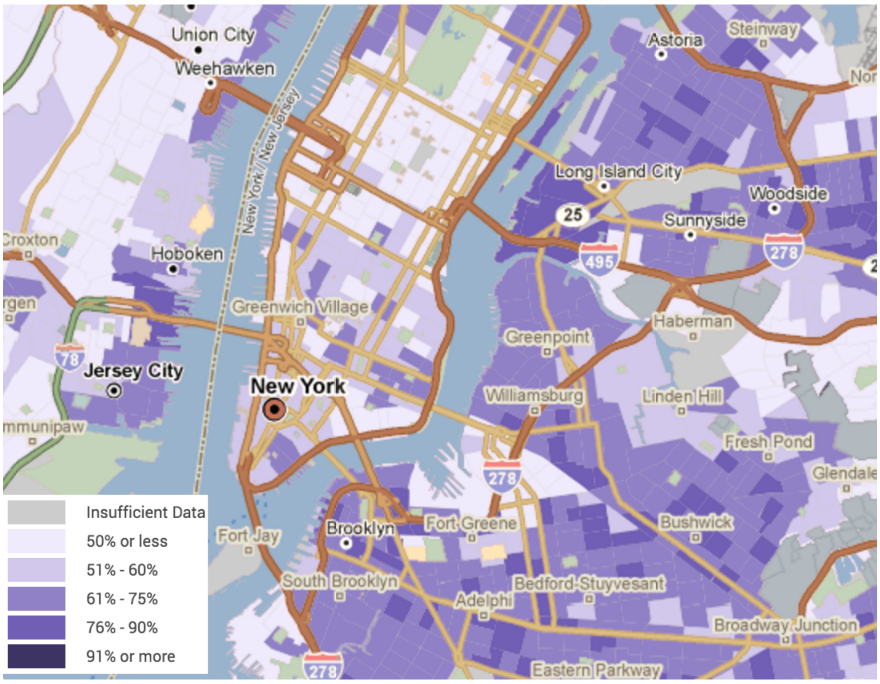

Figure 7: Map depicting % of population taking public transit in New York City Area

1.4 Organization of the paper

There will be three segments to this paper. Firstly, in Section 4, the paper will ascertain the significance of gasoline prices in determining the transit ridership. Secondly, in Section 5, the paper will test if current transport operators have been moderating or adjusting their services according to the gasoline price fluctuations. Finally, in Section 6, the paper will seek to provide evidence of the presence of the substitution effect between public transit and car driving that forms the underlying premise of this research investigation.

1.4.1 The Significance of Gas Prices

Following the regression models of past papers, this paper will run similar regressions aligned with the models used in previous studies, using a time-series dataset that spans monthly data from January 2002 to August 2015, through the following different models that utilize the Ordinary Least Squares estimates. More will be discussed in the Data and Methodology section of the paper.

- Constant elasticity, where we assume a constant elasticity throughout the period of study.
- Event elasticity, where we assume that elasticity values as events such as fare price increase and structural upgrades occur within the transport system.
- Instrumental variables analysis, where we account for a possible endogeneity within the model, and use an instrument variable to correctly predict the effect.

- Lagged models, in which we assume that consumers respond to the gas prices with a lag time of 6 months.

Furthermore, the fixed effects control variables that are being used within the regression methods are more comprehensive than past studies. To account for possible omitted variable biases, control variables such as the yearly and monthly fixed effects, system upgrades, fare increases and unemployment are included. Details and the reasoning for inclusion of the variables are elaborated in the Data and Methodology components of the paper.

1.4.2 Agency Responses to Gasoline Price changes

After determining the significant relationship between gasoline prices and ridership, Transit/Vehicle Revenue Hours is being tested for significant relationship with gasoline prices. Using the most appropriate regression model decided for each transit mode in the analysis above, we change the dependent variable from Log(Ridership) to that of Log(Revenue Hours) in this segment of analysis.

1.4.3 Providing evidence of the substitution effect

To justify the elasticity, or inelasticity of the values obtained of public transit, evidence of the substitution is being investigated along critical highway segments of the San Francisco Bay Area, where service such as the BART and MUNI operate, such as the CA-24E, I-80W and US-101 highways.

Data of average speeds of the highway at the peak hours (5-7pm) at these freeway segments was collected. Using the most appropriate regression model determined in the first segment of analysis, the highway speed is used as the dependent variable to predict and test for a corresponding significant relationship with the independent variables. The underlying foundation of such analysis is that a higher gasoline price would deter driving, and in doing so with less drivers on the roads will lead to an increase in average travelling speeds which are picked up by these sensors along the freeway. Similarly, the contrary, whereby gasoline prices are lower would encourage more people to drive and hence result in a lower average travelling speeds because of the traffic congestion caused.

1.5 Key Contributions

Past papers have provided this paper with the fundamental models of regression that will be covered in detail in the next section. The understanding of the methodology behind the papers have elucidated three key areas of renewal that this paper hopes to achieve.

Firstly, the time relevance of past research is to be questioned, particularly after the significant decrease in crude oil prices in 2008. All of the research mentioned above were conducted during the time where prices of oil have been on a steady upward trend. Given the significant drop of prices from the high of $\approx$ USD100/barrel to the current $\approx$ USD45/barrel, this update of macroeconomic circumstances will be an important contribution to the literature that follows beyond this paper, to ascertain if the relationship between gas prices and public transit ridership is still statistically significant and time relevant. This is particularly pertinent as many research reports, including one from the National Public Radio (2015) that point to the likely possibility of an environment with consistently lower gasoline prices than before. Under such an environment, policymakers will have to perhaps revise their judgement and policydecisions.

Secondly, upon careful inspection of the econometric methods utilized in past studies such as Currie and Phung (2007), it occurs that the regressions in those studies have not accounted nor controlled fully for other important variables that are determinants of public transit ridership. This paper posits that there could be phenomenon of Omitted Variable Bias (OVB) present. In the past papers, the only control variables utilized in the models were the month fixed effects, which this paper also takes into account, but other control variables such as yearly and monthly fixed effects, Population Unemployment Rate, Fare Price Increases (FI) and System Upgrades were not included. This paper gathered all of these important variables and will include them in the regressions as determinants as to whether a person would take public transport. Hence, this paper will show the value of adding these control variables into the regression and test for their significance.

Finally, previous studies have used aggregated data from many cities differing in size and density across the entire United States, which has limited policy implications given the decentralized political nature of the various transit agencies. Other studies that covered on one locality were on other metropolitan areas such as New Jersey, Philadelphia etc. The focus on San Francisco Bay Area is novel and serves a local purpose to shed insights for the San Francisco Transportation planning agencies and public transit operators.

1.6 Key Findings

The ridership trends were evaluated and compared with the California gas prices for each mode of public transit in the San Francisco Bay Area. Based on the study's regression, there seems to be an overall positive effect of gas prices on public transit ridership, statistically significant in all transportation modes investigated, of the Heavy Rail, Light Rail and Bus services. The elasticity values obtained from the regressions range from 0.0581 to 0.147. It was also discovered that different regression models best suited different transportation modes, with the Constant Model applicable for both the MUNI Bus and MUNI Light Rail, whereas the Behavioral Model applying for the BART. Using the Vehicle Revenue Hours to relate to the same set of independent and control variables, it appears that the transit agencies do not currently take into account fluctuations in gasoline prices to factor for the transit service schedules. Upon further investigation, there are preliminary relationships between the gasoline prices and highway speed at selected key highway segments within the Bay Area, near the operation of the BART, MUNI Light Rail and MUNI Bus modes. The following sections will elaborate these key findings in more detail.

2 Literature Review

2.1 General Transportation Research

The public transportation sector is an area of huge public sector spending that generates controversy and lots of debate. Although the amount of public spending on transportation has increased over time, the share that it takes from GDP maintains at around 3%. Most spendings on transportation are still on highway construction, as compared to mass transit development (CBO, 2015) As a result, there have been many studies that look into the determinants of public transit ridership, from both economics and transportation studies perspectives. The analysis offered have interesting policy implications and good intuition, but they do not all offer a direct explanation to this paper's question of finding the causal relationship between gasoline prices and public transit ridership.

In Taylor et al (2005), they regressed external and control variables thought to influence transit usage, looking at dozens of variables spanning transit system characteristics, highway system characteristics, regional geography, metropolitan economy and population characteristics. These variables, while interesting, do not answer this paper's question of gasoline prices' effect on public transport ridership. These variables, such as the number of highway lanes and population density, while controlling for the differences across geographical regions, are not as essential in an analysis that focuses on a single geographical locality. In a study of one geographical location, we can assume these factors to be relatively constant over time.

There are also papers that deal with the price elasticity of demand of gasoline, but do not delve into the effect on public transit ridership directly. Small and Van Dender (2007) looked into the price elasticity of gasoline but measured responses underlying the elasticity with changes in amount of driving and changes in fuel intensity, under the assumption that an increase in gasoline prices will reduce travel on private driving. Goodwin (1992) also briefly mentioned on an elasticity of public transit but pointed out mostly on the relationships between petrol price on traffic congestion levels and petrol consumption. The more relevant question to this research would have been the direct relationship between gasoline price change to public transportation ridership. Their research had great policy implications, as they explored energy policy and climate change impacts, and can be interesting add-ons to consider.

2.2 Relevant Transportation Research

Amongst the few papers that have investigated the relationships between gasoline prices and public transit ridership, the consensus from these papers seems that there is a small but significant amount of the variability in transit ridership that is attributable to fluctuations in gasoline prices (Lane, 2010). This is indeed the relationship that this paper wishes to ascertain given the changing gasoline price environment.

2.2.1 The Study of Time and Seasonal Effects

The time and seasonal effects in studying public transportation behavior is highly vital and accounted for in most relevant literature.

Haire and Machemehl (2007) showed that the seasonal average of trips within the US from 1992-2001 differed greatly from seasons, with the Spring months (March-May) and Fall (Sept-Oct) showing much more trips being made. Furthermore, controlling for seasons also yielded higher t-statistics on the gas price coefficient, giving evidence that this is an important variable to be controlled for.

Similarly, in Currie and Phung (2007) and Yanmaz-Tuzel and Ozbay (2010), monthly dummy variables were included to account for these cross month differences.

2.2.2 Lagged Temporal Choices

As for behavioral choices, Goodwin (1992) found that behavioral response to cost changes in transportation is a response that takes place over time. Hence, people take time to adapt to the change in prices. This was similarly talked about in Yanmaz-Tuzel & Ozbay (2010) in the New Jersey context, where there was evidence that several months elapse before travelers respond to gasoline price changes. Currie and Phung (2007) also agrees on the same point by specifying evidence of 7-8 months time lag between changes in gas price and public transit demand. This gives the theoretical underpinning of the Behavioral Lag model that will be elaborated further in the Data and Methodology segment of the paper.

Miley and Weinberger (2007) ends off their paper with future considerations with a question that this paper hopes to find some information on. Towards the end of their study period in Summer 2008, gas prices had decreased substantially in Philadelphia. However, ridership had increased, which contradicts standard economic theory, or our intuition, under conditions of ceteris paribus. Could habit formation play a role in determining public transport ridership? With renewed data, it is possible to find more evidence for the above hypotheses, and this paper institutes a six-month behavioral lag gas price variable to test this hypothesis.

2.2.3 New Areas of Consideration

The contributions of this paper are threefold. Firstly, this paper seeks to establish and confirm the time relevance of the gas-transit relationship. Secondly, this paper utilizes a more robust regression method in determining this relationship by expanding on past regression models through including more control variables. Thirdly, this paper also carries out the analysis in the San Francisco locality, providing novel insights in an area that has not been previously investigated.

This paper serves to establish and determine this relationship following this macroeconomic environmental shift, such that there is a quantifiable effect that is of value to both policy makers and public transit operators to better effectively manage fleet and respond to consumer demands.

3 Empirical Strategy and Data Methodology

3.1 Transport Choice Modelling

As a foundation to frame the public transportation choice issue into an economics problem, this paper refers to the standard model[2] which represents the costs structure and considerations from the perspective of an urban dwelling individual. This model brings into perspective on how fuel costs enter the decision making process, and we make a few assumptions to explicate the model, namely

- The individual has three options to choose from: driving alone, riding a bus or riding a train (both heavy and light rail). This essentially reflects a choice between private and public transportation.
- The individual seeks to cost minimize, as represented in the equation below:

$$Trip\ Cost = m + T_a d_a + T_v d_v \tag{1}$$

Full cost of the trip involves both monetary and time costs, where time costs are split into both access time(T_a) and travel time(T_v). O'Sullivan showed with empirical studies that the marginal disutility of access time is larger than marginal disutility of in-vehicle time. The model hence shows the implications of fluctuating gas price can have significant impacts on "m" that will affect commuter decisions in deciding to use private or public transportation.

3.2 Description Of Data

3.2.1 Data Collection and Sources

Sources Of Data: Given the novel nature of this study, the dataset was self constructed using various sources. The two most important datasets used to construct this dataset were the independent variable of gasoline prices and the dependent variable of public transit ridership information.

[2] Model used by Arthur O'Sullivan in *Urban Economics*, Chapter 11

Ridership Information was obtained from the American Public Transport Administration (APTA) [3]'s Ridership Report section, where information about the Average Monthly Ridership for various transport systems around America are documented since January 2002. It specifically also provides breakdown into the various modes of transport within each of the transport system, allowing users to focus on the subset of data required. Specifically for the San Francisco geographic subset, there are 164 data points in the latest monthly dataset that spans January 2002 to August 2015 data on the Unlinked Passenger Trips (UPT) figures for the Bus, Light rail and Heavy rail services that this paper analyses, in the forms of Bay Area Rapid Transit (BART) and the San Francisco Municipal Railway Bus (MUNI Bus) and Light Rail (MUNI). This dataset is updated quarterly to include the latest figures, which presents opportunities for the study to update the regressions as more data become available.

The gasoline prices were obtained from the US Energy Information Administration (EIA), which provides a weekly California All Grades All Formulations Retail Gasoline Prices (Dollars per Gallon). The dataset extends back into 2000, offering approximately 180 monthly data points. This paper hence constructs the dataset from the time period of overlap between January 2002 and August 2015. In using the Behavioral Model of analysis, where the paper accounts for a delayed response in public transit to gas prices, we shift the scope of the gas prices scope from July 2001 to February 2015, which is available in the dataset provided by the EIA. This dataset is updated weekly, and also presents opportunities for the study to update the regressions as more data become available.

Control Variables: To ensure regression accuracy, this paper constructs many fixed effects that are included in the regression to reduce the omitted variable bias effects.

Month and Year dummy variables were created manually based on the information obtained from the above two datasets to account for time and seasonal effects. These monthly dummies will help to account for the inherent seasonal differences experienced in travel patterns between different time periods of the year. 11 dummy variables were created to dictate January to November, with December (dropped) being the base value. Similarly, Years 2003 to 2015 were created, with 2002 (dropped) being the base value to prevent perfect multicollinearity.

The unemployment rate gives us an overview of the demographic profile of the population we are studying, and possibly explain for any sudden shocks when there are macroeconomic restructuring that affects consumption patterns. Yanmaz-Tuzel & Ozbay(2010) also mentioned the significant positive impact of economic growth on transit ridership, hence including the unemployment rates in the regression helps to control for the differences in economy state. Unemployment rates in the Bay Area was gathered from datasets obtained from the Bureau of Labor Statistics (BLS), whereby the

[3] The APTA, based in Washington DC, is America's leading public transportation agency that provides news and information about the industry and brings together members to strengthen and improve public transportation.

economic situation of San Francisco-Oakland-Fremont was listed, and unemployment rates of each month going back to 1990 was available. The intuition is that when unemployment rate increases, we might possibly see higher public transit as the people choose public transit to reduce cost incurred of travel. Also, at time periods of higher unemployment, it is likely that work-related commute will be reduced, resulting in decreased total trips of travel.

Fare Increases for each mode of transport was carefully constructed in the following ways. For the MUNI Bus and Light Rail systems, they follow a consistent and fixed fare structure, and given that the MUNI website did not specifically present information on fare increases, the author collated the fare increase data through searching for Press Releases and News Archives that documented the fare price increases from the dollar value of $1 in 2002 to $2.25 today. Fare Increases may cause intuitively a decrease in ridership as the cost of riding public transit could possibly become more similar to that of private transportation.

For the BART, as the fare prices were calculated dependent on the stops one travels, an average Fare Increase percentage was obtained from the BART Management[4] itself, and the average percentage increase of the fare was considered for this variable.

These system upgrades listed were sourced from the key openings within the systems that would predictably have an increased ridership influence during the period of study. For the BART, there were three system upgrades considered in this research. Firstly, in June 2003, the San Francisco International Airport (SFO) Extension, with South San Francisco, San Bruno and Milbrae stations were connected. Secondly, in February 2011, the West Dublin/Pleasanton station was opened. Thirdly, and finally, the Oakland International Airport (OAK) Connector service was opened in November 2014.

For the MUNI, there was only one significant line extension which is the opening of the Third Street (T line) in April 2007 was marked as a significant upgrade within the MUNI system.

For the MUNI Bus, it is much harder to model using discrete events such as the above given the numerous lines that are within the system that have been expanded or collapsed during this period of investigation. Hence, this variable is not included for the MUNI Bus analysis.

Finally, the vehicle revenue hours helps to control for the variance experienced from month to month in terms of operating hours and days, and is also provided in the APTA dataset where the UPT values were extracted.

[4] The author thanks Charlotte Barham from the BART management with kind assistance in giving this piece of vital information

Table 1: **List of Regression Variables**

Variable Name	Variable Description
$\log(Ridership_{it})$ (Dependent Variable)	This is the natural logarithm of monthly ridership figures. This dependent variable measures the cumulative effect of the gasoline prices along with control variables on the total public transit ridership.
$\log(GasPrice_t)$ (Independent Variable)	This is the natural logarithm of average monthly gasoline price. The main independent variable is the average monthly gas price which is derived from the average weekly gas prices obtained from the EIA.
$\log(GasPricelag_t)$ (Independent Variable)	This alternate independent variable is used in the Behavioral Lag model where I investigate if people respond to public transit ridership with 6 months of delay in gasoline prices.
M_t (Control Variable)	These month dummy variables (January to November) will help to to account for inherent seasonal differences experienced in travel patterns between different time periods of the year.
$Year_t$ (Control Variable)	These year dummy variables (2003-2015) will account for the year-on-year effects on public transit ridership.
$Unemp_t$ (Control Variable)	The Unemployment rate in the Bay Area gives the demographic profile of the population of study, and could possibly explain for sudden shocks in transit ridership when there are macroeconomic restructures or factors that affect consumption patterns.
$FareInc_{it}$ (Control Variable)	This variable represents the Fare Increase since the first period of investigation (Jan 2002) will take into account of the increased cost of travelling in either the MUNI or BART as a factor that could affect the transit ridership.
Upg_i (Control Variable)	Increase in service coverage (new line extension) might bring about a bigger demographic population who access the service and hence might have explanatory power of a shock of transit ridership.
$log(VehicleHours_{it})$ (Control Variable)	The vehicle revenue hours varies from month to month, depending on the number of days in the month and also holiday schedules and strikes etc. This variable helps to make the comparison between months fair.

Applicability of Data chosen:

Ridership Data from the APTA segments the SF Muni Ridership into its different modal forms (Light Rail, Cable Car, Bus etc), hence choosing Bus and Light Rail subsections gives us the relevant figures needed. Similarly, as the BART only operates on the Heavy Rail segment, the figures obtained from the dataset can be directly applied onto our analysis.

For the gas prices, even though the regression utilizes the state gas average, the graphical analysis provided here from data supplied by 'Gasbuddy.com' shows us that the usage of State Average does not generally cause a loss of generality, as both follow very similar price movement trajectories.

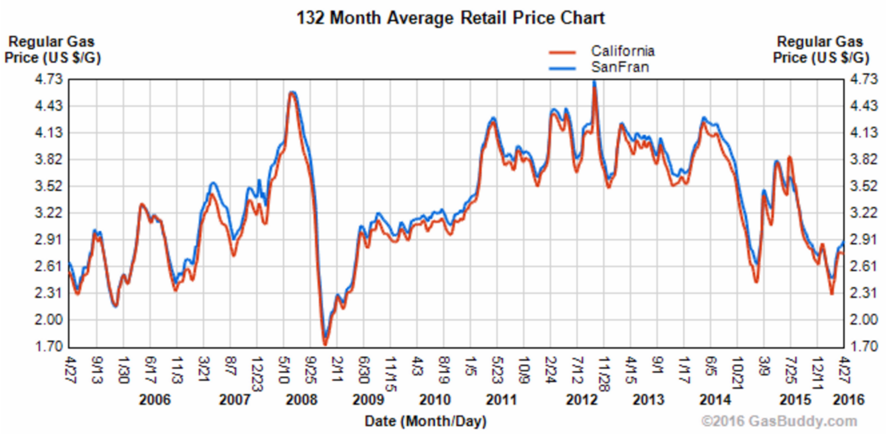

Figure 8: Trend of CA State Avg and San Francisco Gas Prices over the past 11 years

3.3 Descriptive Statistics

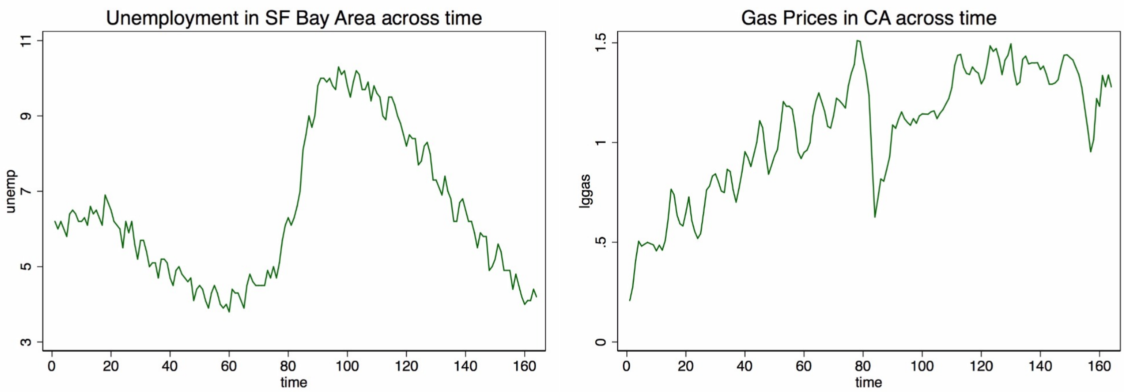

Figure 9: Trend of Unemployment and Gas Prices in San Francisco over period of investigation

Table 2: **Summary Descriptive Statistics of Key Variables**

Key Variables	Obs	Mean	SD	Min	Max
Log (BART Ridership)	164	16.04	0.121	15.79	16.35
Log (MUNI Ridership)	164	15.16	0.113	14.86	15.47
Log (MUNI Bus Ridership)	164	15.86	0.076	15.62	16.04
Log (Gas)	164	1.065	0.308	0.208	1.511
Log (BART Vehicle Hours)	164	12.00	0.119	11.58	12.20
Log (MUNI Vehicle Hours)	164	10.78	0.101	10.47	10.96
Log (MUNI Bus Vehicle Hours)	164	11.73	0.048	11.58	11.83

3.3.1 Interpretation

Gas Prices: We can see from the graph in Figure 9 that the gas prices have been on a steady increase up until a point in around 2008 when it had a steep drop. The prices following that time period has been largely volatile, showing much fluctuations in price changes. However, it has since shown a significant decrease in price which extends beyond the dataset (beyond April 2015). There was also a huge increase in the unemployment rate that coincides with around the same time of the gas price drop.

As we are estimating the cross demand elasticity, the responsiveness of quantity demanded for a good to a change in the price of another good, the terms being measured have to be in natural logarithmic terms (Ln). The same applies to the vehicle revenue hours for all three services, which are presented here.

However, for ease of understanding on how the variation of the riderships vary across the months, graphs in Figure 10 depicting how much each month's average ridership deviates from the average are plotted, split into the various transportation modes. As evident, there is higher ridership throughout the spring and summer months, and lesser ridership over winter months. This shows the presence of seasonal fluctuations of ridership across the modes.

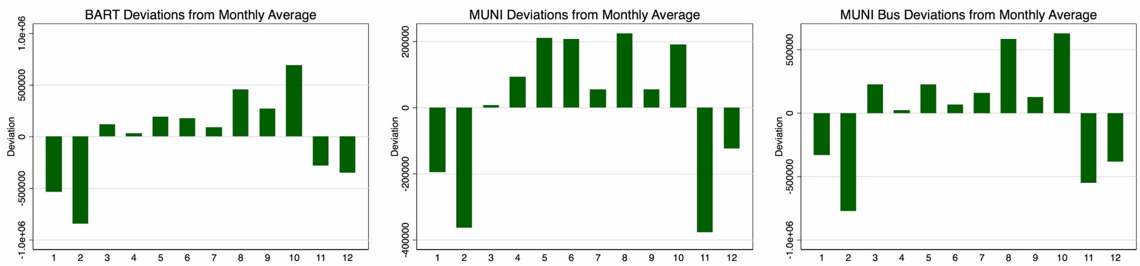

Figure 10: Deviations of Monthly ridership from Monthly Average

3.4 Empirical Strategy

This paper considers four different models to estimate the cross elasticity. The Constant Elasticity, Event Elasticity and Behavioral Lag models estimates the elasticities using the *Ordinary Least Squares (OLS)* whereas the Instrumental Model utilizes a two-stage least squares (2SLS) *instrumental regression* to estimate the coefficients. The 2SLS method is where the dependent variable's error terms are correlated with the independent variable, causing an endogeneity problem, and an instrument variable is used to estimate the correct coefficient.

The Constant and Event models' rudimentary structure were adapted from Currie and Phung (2007), whereas the Instrumental and Behavioral Lag were derivations created by the author.

For each mode of transport, the four models are applied:

1. **Constant Elasticity Model**:

$$\begin{aligned}\log(Ridership_t) = \alpha + e_0 \log(GasPrice_t) + \sum_{i=1}^{11} \gamma_i M_i + \sum_{i=2003}^{2015} \delta_i Year_i \\ + \beta_1 \log(VehHrs_t) + \beta_2 Unemp_t + \beta_3 Upgrade_i + \beta_4 FI_t + \mu_t\end{aligned} \tag{2}$$

This model measures elasticity by assuming constant elasticity, for the entire duration of investigation from 2002-2015.

2. **Event Elasticity Model**:

$$\begin{aligned}\log\left(Ridership_t\right) = \alpha + (e_0 + \beta_1 Upgrade_i + \beta_2 FI_i)\log\left(GasPrice_t\right) + \sum_{i=1}^{11}\gamma_i M_i \\ + \sum_{i=2003}^{2015}\delta_i Year_i + \beta_3\log\left(VehHrs_t\right) + \beta_4 Unemp_t \\ + \beta_5 Upgrade_i + \beta_6 FI_t + \mu_t\end{aligned} \tag{3}$$

This model measures elasticities, allowing some flexibility in that the elasticity can change as events occur, such as system upgrades and fare price increases, which are reasonable changes that can possibly change the demand elasticity of public transit.

3. **Instrumental Model**:

$$\begin{aligned}\log\left(Ridership_t\right) = \delta + e_0\log\left(CrudePrice_t\right) + ControlVars + \sum_{i=1}^{11}\gamma_i M_i + \sum_{i=2003}^{2015}\delta_i Year_i \\ + \beta_1\log\left(VehHrs_t\right) + \beta_2 Unemp_t + \beta_3 Upgrade_i + \beta_4 FI_t + \xi_t\end{aligned} \tag{4}$$

To account for possible endogeneity between gas prices and public transit ridership, an instrument of crude oil prices (Western Texas Intermediate) price is used. This instrument is exogenous and relevant, as there is a high correlation between the crude gas prices and gasoline prices, but yet there is no direct relationship between crude oil prices and public transit ridership directly given that consumers are not affected directly through the crude oil price except via gasoline price, conditional on controls.

4. **Behavioral Lag Model**:

$$\begin{aligned}\log\left(Ridership_t\right) = \alpha + e_0\log\left(GasPrice_{t-6}\right) + \sum_{i=1}^{11}\gamma_i M_i + \sum_{i=2003}^{2015}\delta_i Year_i \\ + \beta_1\log\left(VehHrs_t\right) + \beta_2 Unemp_t + \beta_3 Upgrade_i + \beta_4 FI_t + \mu_t\end{aligned} \tag{5}$$

This model is exactly the same as the Constant Elasticity model except on the independent variable which is a time-lagged Log Gas Price. It might be possible that the response to gasoline prices is not immediate. As noted in both Yanmaz-Tuzel & Ozbay (2010) and Currie and Phung (2007), it takes several months before travelers respond to gasoline price changes, hence a 6-month gasoline lag variable is used to see if this relationship is viable. (For instance, a January 2002 ridership figure will correspond to a July 2001 gasoline price)

3.4.1 Instrument Validity

The Federal Reserve Economic Data (FRED) from St. Louis Federal Reserve Bank provides data on the monthly West Texas Intermediate (WTI) prices at Cushing, Oklahoma. The WTI is a grade of crude oil typically used as a benchmark in oil pricing, and it is also the underlying commodity of the New York Mechantile Exchange's oil futures contracts, hence it is a reasonable easy to access public domain data.

Prior to analyzing the results from regressions, the instrument is tested to see if it is relevant and exogenous. The WTI price is highly correlated to the gasoline prices because the crude oil is a key component of the price of gasoline prices, along with refinery, distribution and taxes. Intuitively, there is also no direct relationship between transport ridership and crude oil price. Based on economic reasoning, the exogenous and relevance requirements are satisfied.

To confirm this, using the estat first stage function on STATA, the F-statistic for the instrument choice of WTI is generated. With the F-stat from the partial F test >10, which satisfies the rule of not being a weak instrument, the instrument can be said to be valid for the IV regression.

```
estat first

First-stage regression summary statistics
```

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	Robust F(1,130)	Prob > F
lggas	0.9745	0.9688	0.6388	136.794	0.0000

Figure 11: Testing for Instrument Relevance using the BART Instrumental Regression

Similarly, using the estat endog function on STATA, the exogeneity of the instrument is also confirmed.

```
estat endog

Tests of endogeneity
Ho: variables are exogenous

Robust score chi2(1)            =  1.97498  (p = 0.1599)
Robust regression F(1,128)      =  1.54628  (p = 0.2160)
```

Figure 12: Testing for Instrument Exogeneity using the BART Instrumental Regression

4 Gasoline Prices and Public Transit Ridership

This paper was able to find statistically significant and positive coefficients for the gasoline price coefficients in each of the modes being investigated, which satisfies the original hypothesis. This paper first replicates the Constant and Events models of what past papers have done, and then proceeds to add in the other control variables using the proposed four models described above. The results will be interpreted by the mode of transport for clarity purposes, and are displayed in summary of regression results tables. 6 regressions were conducted for each transportation type, as represented by the short forms in the column heading of the regression output table: the Replicate Constant Model (Rep Const), the Replicate Events Model (Rep Events), Constant Model (Constant), Events Model (Event), Instrumental Model (Instrument) and Behavior Lagged Model (Lag).

4.1 BART

The replicated models of Constant (0.312) and Events (0.0799) elasticities both returned statistically significant coefficients and positive figures for Log Gas. However, the main problem with this model is that because the Yearly fixed effects were not added on, there are unexplained deviations in ridership numbers. As a result, the Replicate Models do not have as high an R-squared value than the expanded models.

Looking at the enhanced models, the Log Gas coefficient is statistically significant only in the Lagged variable of 6 months. As the four expanded models all have sufficiently high R-squared values (0.943), the choice to choose the most suitable model lies in the economic reasoning behind it.

The positive and significant coefficient of 0.0581 for Log Gas Lag suggests that an increase in 1% of lagged gas prices comes with an increase of 0.06% of BART Ridership.

A possible explanation for this occurrence lies in the psychology of perception that people have in favor for rail systems over the bus. Studies have been conducted that under very similar conditions, such as one in Hensher and Mulley (2015), when presented with the same images with the only difference being a bus and a train as the mode of transport, a vast majority (73%) chose the rail images. This biased perception of the transportation essentially points to the fact that there is lesser demand elasticity for the heavy rail service, particularly in the context when heavy rail systems like the BART operate on its own dedicated lanes of operation, and provide a much more reliable service schedule and expectations (unlike buses and lightrail which may be stuck in traffic jams etc). Hence, the elasticity values that are lower for the Heavy Rail can be explained and accounted for this way for the relative demand inelasticity.

This therefore also supports the Behavioral Lag model that is favored over the Constant elasticity model. Given the relative demand inelasticity of BART Ridership, it takes a much longer time for consumers to change their behavior in terms of choices to stop riding the BART when gas prices decrease. For the reverse case, when prices increase, there is also a higher "barriers to entry" because of the relative network effects of the BART system compared to the Bus and Light Rail systems.

There are generally less accessible options for one whose access time[5] to a BART station as compared to bus and light rail services that may be more accessible within the neighborhoods.

The time fixed effects show that there is significant variance in the transport numbers at different phases of time and year. The month fixed effects return statistically significant results in most months, from January to October. The year fixed effects were also statistically significant for most years, from 2007-2009 and 2012-2014. These provide evidence there is seasonal and time fluctuation that the month and yearly fixed effects help to add explanatory power to the regression.

Unemployment was an important dummy variable with a negative coefficient of -0.0103, as there are many key stations along the San Francisco and Oakland segment of the service that commutes downtown, where the commercial centres of the cities are. Hence, in a time period of unemployment, there will be significant lesser passengers who use the service to commute to work.

However, the BART Upgrades Dummies, along with the Fare Increases were statistically insignificant. It could be that the capacity of the train service was already reaching to the maximum capacity, which is currently still operating on, and hence the decreases or increases in passenger services were not accurately picked up. We also acknowledge that there perhaps could be unaccounted for factors that may have confounded this results, but in summary, the results give us a good explanation of the occurrence.

Model chosen: Behavioral Lag Model, Model 6 in Table 3

[5] Refer to the model in 3.1 with regards to the disutility of access time

Table 3: **Summary of Regression Results for BART**

Dependent Variable **Log(Ridership)**	(1) Rep Const	(2) Rep Events	(3) Const	(4) Events	(5) Instrument	**(6)** **Lag**
Log Gas	0.312***	0.0749**	0.0299	-0.0484	-0.00301	
	(0.0175)	(0.0376)	(0.0305)	(0.0680)	(0.0343)	
Log GasLag						0.0581**
						(0.0250)
BART Fare Inc		0.626***	-0.282	-0.723*	-0.174	0.0717
		(0.0854)	(0.232)	(0.421)	(0.234)	(0.259)
BART Upg1		-0.0292	-0.0182	-0.0456*	-0.0166	-0.0198
		(0.0195)	(0.0154)	(0.0234)	(0.0140)	(0.0147)
BART Upg2		0.0386***	0.0239**	-0.138	0.0232**	0.0156
		(0.0124)	(0.0112)	(0.102)	(0.00992)	(0.0110)
BART Upg3		0.0239	-0.0292	0.114	-0.0354*	-0.0316
		(0.0187)	(0.0220)	(0.159)	(0.0198)	(0.0209)
BART Fare Inc*Log Gas				0.423		
				(0.285)		
BART Upg1*Log Gas				0.0483*		
				(0.0260)		
BART Upg2*Log Gas				0.120		
				(0.0835)		
BART Upg3*Log Gas				-0.0858		
				(0.116)		
Unemp		-0.0145***	-0.00991	-0.00357	-0.0111*	-0.0103*
		(0.00209)	(0.00675)	(0.00701)	(0.00600)	(0.00614)
Yearly FE	N	N	Y	Y	Y	Y
Observations	164	164	164	164	164	164
R-squared	0.632	0.887	0.944	0.947	0.943	0.945

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Fixed effects here refer to the yearly and monthly fixed effects (not shown in table), Fare Increase, System Upgrades, Interaction variables and Unemployment.

4.2 MUNI Light Rail

In both replicate models, the coefficient for gasoline prices were significant. However, the sign on the coefficient was not consistent. In the Replicate Events model, the -0.107 coefficient was not intuitive, as a rise in gasoline prices was correlated with a decrease in the MUNI train ridership. It is possible that there is a wrong regression specification or omitted variable bias present. Also, in both regressions, the R-squared values were very low (0.021 and 0.388). Therefore, the expanded models which had much higher R-squared values are looked upon for the explanation of the Ridership values.

Looking at the enhanced models, the Constant model seems to be the best suitable model for two reasons. It is the only model in which we have a statistically significant coefficient for Log Gas variable of 0.147, which makes economical sense. Furthermore, comparing the Constant model with that of the other models, the control variables seem to be more predictive in power, with their statistical significance.

The positive and significant coefficient of 0.147 for Log Gas suggests that an increase in 1% of gas prices comes with an increase of 0.15% of MUNI ridership, which is about twice the effect of that of the BART elasticity. Interesting, this relationship is the reverse of what previous papers have derived, given that past papers like Currie and Phung (2007) argued for higher sensitivity in elasticity values for the Heavy Rail over Light Rail and Bus.

The time fixed effects show that there is significant variance in the transport numbers at different phases of time and year. The month fixed effects statistically significant results for a few months in January, February and November. Similarly, for the Year fixed effects, it is also significant in the years 2004-2006 and 2004-2015.

The MUNI Fare Increase was a strongly significant negative coefficient of -0.213. Due to the nature of the MUNI Light Rails, and the general traffic conditions in San Francisco, the fare increase prompted much of a substitution to the mode of transport being used. Alternatives such as biking, walking or even app-based taxis such as Uber and Lyft, which have made a rise in recent years are all possible contributors to the shift in the ridership figures.

Going off that tangent, it is possible that with MUNI possibly being the cheapest form of public transportation (other than walking), that when unemployment rates increase, there is an increase in ridership, as depicted by the statistical significant coefficient of 0.0744.

Interestingly though, the MUNI Upgrade of the T line did not seem to have caused a significant difference in ridership. Comparing this to the BART Models, it does seem that there might be unexplained deviations due to the relatively lower R-squared values.

Model chosen: Constant Model, Model 3 from Table 4

Table 4: **Summary of Regression Results for MUNI Light Rail**

Dependent Variable **Log Ridership**	(1) Rep Const	(2) Rep Events	**(3) Constant**	(4) Events	(5) Instrument	(6) Lag
Log Gas	0.0530**	-0.107**	0.147**	-0.00231	0.117	
	(0.0235)	(0.0521)	(0.0688)	(0.100)	(0.0755)	
Log GasLag						-0.0514
						(0.0590)
MUNI Fare Inc		0.0657	-0.213**	-0.222	-0.204**	-0.189*
		(0.0468)	(0.0988)	(0.188)	(0.0924)	(0.0971)
MUNI Upg		0.0341	0.00144	-0.221	0.00354	0.00857
		(0.0224)	(0.0350)	(0.166)	(0.0317)	(0.0330)
MUNI FareInc*Log Gas				-0.00670		
				(0.177)		
MUNI Upg*Log Gas				0.203		
				(0.144)		
Unemp		0.0151***	0.0744***	0.0751***	0.0738***	0.0695***
		(0.00481)	(0.0141)	(0.0147)	(0.0128)	(0.0152)
Yearly FE	N	N	Y	Y	Y	Y
Observations	164	164	164	164	164	164
R-squared	0.021	0.388	0.662	0.670	0.662	0.652

Robust standard errors in parentheses

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

Fixed effects here refer to the yearly and monthly fixed effects (not shown in table), Fare Increase, System Upgrades, Interaction variables and Unemployment.

4.3 MUNI Bus

For the replicated models, only the Events model showed a statistically significant coefficient that is negative, -0.0460 which is not intuitive in nature, given that a rise in gas prices goes along with a decrease in the MUNI Bus ridership. It is possible that there is a wrong regression specification or omitted variable bias present. Furthermore, the R-squared values for the replicated models were significantly lower than the expanded models. Hence, we look at the enhanced models for more explanations.

Looking at the enhanced models, there are three models (Constant, Instrument and Behavioral Lag) that return statistically significant coefficients. However, the Constant model seems to be the best suitable model for three reasons. This is chosen through the processes of elimination.

Firstly, in the Lagged model, the coefficient for Log Gas Lag is -0.110, which is significantly negative, which does not make economic intuitive sense. Furthermore, upon looking at the Yearly fixed effects dummy variables, they are not statistically signficant unlike the other models. Hence we look at the other two possible models. We conduct the tests of endogeneity on STATA to decide if the instrumental model is viable. There are mixed results obtained, with the Durbin test showing a need for instrumental regression, and the Wu-Hausman test displaying the reverse results. Therefore, the conservative approach would be to adopt will be the Constant model. Furthermore, the adjusted

R-squared value for the Constant model is also the highest amongst all the expanded models, at 0.686, which suggests that this model has the highest explanatory power.

The positive and significant coefficient of 0.125 for Log Gas suggests that an increase in 1% of gas prices comes with an increase of 0.13% of MUNI ridership, which is about twice the effect of that of the BART elasticity. Similarly to the earlier section on MUNI, this relationship is the reverse of what previous papers have derived, given that past papers like Currie and Phung (2007) argued for higher sensitivity in elasticity values for the Heavy Rail over Light Rail and Bus.

The time fixed effects show that there is significant variance in the transport numbers at different phases of time and year. The month fixed effects also return statistically significant results across the different models in most months, from February to October. The year fixed effects were also statistically significant in most years from 2004 to 2015. These provide evidence there is seasonal and time fluctuation that the month and yearly fixed effects help to add explanatory power to the regression.

The Fare Increase variable was also statistically significant with a positive coefficient of 0.106. This could be explained possibly by the fact that the bus is already one of the most cost effective means of travel, and that the cost increase would not have a significant effect on pushing people towards the alternatives.

Going along with this line of logic, similarly, the Unemployment variable which is not significant, might also be because of the inelasticity of the transit to changes in the macroeconomic environment. Given that the buses operate to more residential areas, this might be the best form of transportation that some residents have, and hence even if they do not have a job to get to, they will still depend on this mode of transportation to get around their daily tasks and routines.

Model Chosen: Constant Model, Model 3 from Table 5

Table 5: **Summary of Regression Results for MUNI Bus**

Dependent Variable	(1)	(2)	**(3)**	(4)	(5)	(6)
Log Ridership	Rep Const	Rep Events	**Constant**	Events	Instrument	Lag
Log Gas	0.0144	-0.0460*	0.125***	0.0999	0.147***	
	(0.0154)	(0.0244)	(0.0378)	(0.0687)	(0.0450)	
Log GasLag						-0.110***
						(0.0416)
MUNI Fare Inc		0.0321*	0.106*	0.0617	0.0993*	0.0998*
		(0.0169)	(0.0551)	(0.113)	(0.0527)	(0.0572)
MUNI Fare Inc*Log Gas				0.0441		
				(0.0918)		
Unemp		-0.00355	0.0144	0.0140	0.0147*	0.00770
		(0.00219)	(0.00929)	(0.00929)	(0.00849)	(0.0114)
Yearly FE	N	N	Y	Y	Y	Y
Observations	164	164	164	164	164	164
R-squared	0.003	0.522	0.686	0.686	0.685	0.684

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Fixed effects here refer to the yearly and monthly fixed effects (not shown in table), Fare Increase, Interaction variables and Unemployment.

4.4 Robustness Checks

As with other empirical papers, robustness checks are done to ensure the structural validity and feasibility of the coefficients for the regressors.

Firstly, this paper replicates the constant model and events model suggested and adapted from Currie & Phung (2007). These models do not involve the complete set of fixed effects controls that we eventually use, but provide us with the structure and some intuition on how the regression results turn out eventually. The results were reported in the first two columns of each type of transportation mode, under the replication models.

In investigating the model of elasticity, Currie and Phung (2007) also discovered that demand elasticities may not be constant over time, when they compared a world events model which took into account significant events such as Hurricane Katrina and 9/11 and a standard constant elasticity model. Including these interaction terms, the world events model returned a higher adjusted R^2 than in a constant elasticity model. Their proposed model from past research was the Events elasticity model. Based on this paper's analysis, it had seemed that a Constant model was most suitable due to the following interpretations.

Based on standard economic reasoning, elasticity which is a behavioral response to change in prices has several factors that affect it. Firstly, the type of product that is investigated, Transportation, is a daily necessity. Given the available options to people's transit, there is a limited range of substitutes available. This favors the constant elasticity model given that a person's preferences should not change significantly within such a short span of time. Also, the changes in prices for public transportation are usually small proportions of a person's income percentage for most segments of the population. It also coincides with the rather low values of elasticity derived in the regressions for each transport mode. Based on BLS Statistics of the Consumer Price Index (CPI), the changes in proportion of transportation from 2012-2013 and 2013-2014 on average was only 0.1% and 0.8% respectively. Therefore, this paper has decided to follow the analysis through with the Constant Elasticity and Behavioral Lag model, which both assumes a constant elasticity during the time period of investigation.

Hence, subsequently, for the chosen model for each transportation type (Behavioral Lag & Constant), there were 4 separate regressions being run, with the final one being the actual regression used in the result analysis, involving all of the fixed effects variables. Using the equations displayed below, we start by running regressions with lesser control variables to see how different the relationship is in magnitude, and logic test checks whether the estimated coefficients were consistent with expectations.

The equations are listed as below:

1. $$\log\left(Ridership_t\right) = \alpha + e_0 \log\left(GasPrice_t\right) + \mu_t \tag{6}$$

   This is the most basic model that only includes Log Gas Price as regressors.

2. $$\log\left(Ridership_t\right) = \alpha + e_0 \log\left(GasPrice_t\right) + \sum_{i=1}^{11} \gamma_i M_i + \mu_t \tag{7}$$

   This is a basic model that only includes the monthly fixed effects as control variables.

3. $$\begin{gathered} \log\left(Ridership_t\right) = \alpha + e_0 \log\left(GasPrice_t\right) + \sum_{i=1}^{11} \gamma_i M_i \\ +\beta_1 \log\left(VehHrs_t\right) + \beta_2 Unemp_t + \beta_3 Upgrade_i + \beta_4 FI_t + \mu_t \end{gathered} \tag{8}$$

   This is the model without the Yearly fixed effects added.

4. $$\begin{gathered} \log\left(Ridership_t\right) = \alpha + e_0 \log\left(GasPrice_t\right) + \sum_{i=1}^{11} \gamma_i M_i + \sum_{i=2003}^{2015} \delta_i Year_i \\ +\beta_1 \log\left(VehHrs_t\right) + \beta_2 Unemp_t + \beta_3 Upgrade_i + \beta_4 FI_t + \mu_t \end{gathered} \tag{9}$$

   This is the complete model with all the fixed effects added.

4.4.1 BART

In all of the regressions, the coefficient on Log Gas Lag is statistically significant.
For the most basic regression, the coefficient on Log Gas Lag is 0.278, with an adjusted R-squared value of 0.571. As we add more control variables, the explanatory power of the regression increases, as the adjusted R-squared value increases to 0.768, 0.905 and 0.945 in the final regression. The decrease in coefficient value also showed that we accounted for potential Omitted Variable Bias inherent in the first regression equation.

4.4.2 MUNI Light Rail

The coefficients on Log Gas variable was significant in the first, third and fourth regressions. The sign however was inconsistent, although the most basic regression also returned a positive coefficient of 0.0429.
The adjusted R-squared value increased from the original 0.021 to 0.242, 0.387 and 0.663 for the final model equation. This shows that each time we add control variables to the regression there is increased explanatory power in the regression. The increase in coefficient value while maintaining the positive sign also showed that we accounted for potential Omitted Variable Bias inherent in the first regression equation. also showed that we reduced the Omitted Variable Bias inherent in the first regression equation.

4.4.3 MUNI Bus

The coefficients on Log Gas was only statistically significant in the final model, although the most basic regression also returned a positive coefficient of 0.0144.
The adjusted R-squared value also increased from the original 0.034 to 0.511, 0.522 and to 0.686 in the final model equation. This shows that each time we add control variables to the regression there is increased explanatory power in the regression. The increase in coefficient value while maintaining the positive sign also showed that we accounted for potential Omitted Variable Bias inherent in the first regression equation.

4.4.4 Newey West Standard Errors

Given that this paper looks at a time-series data where presumably there is potential for autocorrelation, or cross-autocorelation between the adjacent time periods, the regressions with the models that are best selected are re-run on STATA using the Newey West estimator to take into account that time lag.

The results are consistent with the model chosen for each transportation mode. For the BART, the Log Gas Lag variable is still consistent, after taking into account a possible 6 month autocorrelation, which is consistent with the model of the Lagged Gas Price of 6 months. For the MUNI, the Log

Gas variable is only significant up onto a point of 1 month lag, which is consistent with that of the Constant Model. For the MUNI Bus, the Log Gas variable is significant in both the 1 month and 6 months lagged variant, which is consistent in the fact that the MUNI Bus has both the Constant and Lagged Model displaying significant results in the section above.

5 Are Transit Services supplied in response to Gasoline Prices?

The revenue hours variable of the various transportation modes was included in the previous segment because of the possible collinear relationship with the other independent and control variables that will confound the relationship between the variables of interest. This next series of regression will attempt to discover the relationships, if any, between the revenue hours and gasoline prices, controlling for the variables used in the previous segment of regressions.

On the whole, it seems that the agencies do not practice adjusting the service schedules to respond to the changes in the gasoline price, as shown by the statistically insignificant figures of the gas variable obtained. On a policy level, these findings suggest that the BART District (BARTD) and San Francisco Municipal Transportation Agency (SFMTA) for them to consider adjusting their schedules and rosters in response to the gasoline price changes for more efficient fleet management and service rating schedules. This seems especially important given the relatively poor performance of commuter feedback and satisfaction ratings in audit reports done by both the SF City Government and the BARTD themselves.

5.1 BART

In the previous analysis, we determined that the most suitable model for the BART was the Lagged Constant model, hence we regress the BART Revenue Hours to the Log Lagged Gas prices. Using the BART Revenue Hours as the Dependent Variable, we find that the Log Lagged Gas variable is insignificant when controlling for the other variables.

However, it does seem that there has been slight service adjustments since the Upgrades 1 and 2, namely the SFO Extension and West Dublin/Pleasanton stations have been opened. Although the coefficient is slightly negative in the Upgrade 1 variable, it is worthy to note that the service to SFO Airport is less routine and has its service more spaced out than the other segments. Hence, the longer average waiting time for that segment of the service may have affected the total service hours, despite the increase in ridership that might be experienced.

Furthermore, time seasonality is an important factor of consideration as the month variables of February, April, August, October and November are statistically significant. For the year variables, all years investigated from 2002-2015 are statistically significant.

Finally, the negative unemployment coefficient seem to suggest the overall trend in providing less commute services in times of recession, when the agency expects a lower volume of passengers to be demanding of its services. That is consistent with standard economic intuition.

Table 6: **Summary of Regression Results for BART**

Dependent Variable	(1)	(2)
Log BART VehHrs	Base	Lag
Log Gas Lag	0.0216	-0.00343
	(0.0355)	(0.0275)
BART Fare Inc		-0.495*
		(0.290)
BART Upg 1		-0.0394***
		(0.0139)
BART Upg 2		0.0241**
		(0.0119)
BART Upg 3		-0.00488
		(0.0172)
Unemp		-0.0141*
		(0.00716)
Constant	11.97***	11.85***
	(0.0411)	(0.0486)
Observations	164	164
R-squared	0.004	0.934

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

5.2 MUNI Light Rail

In the previous analysis, we determined that the most suitable model for the MUNI was the Constant model, hence we regress the MUNI Revenue Hours to the Log Gas prices. Using the MUNI Revenue Hours as the Dependent Variable, we find that the Log Gas variable is insignificant when controlling for the other variables. This is somewhat not surprising given the similarity in the BART regression.

There does seem to be a significant increase in the MUNI service since the opening of the T line, which is intuitively correct because of the increase in running service of the extended service line. The Fare Increase variable is also strongly significant, presumably as a means to increase the revenue made by the MUNI service. With only 20% of the revenues MUNI make coming from train fares, presumably this relationship is demonstrated by the want for the agency to recover more revenues from the train service that it operates.

Time seasonality continues to be an important consideration factor as the month variables of January, March and June are statistically significant. For the year variables, all years investigated from 2002-2015 are statistically significant in the regression.

Finally, the negative unemployment coefficient seem to suggest the overall trend in providing less commute services in times of recession, when the agency expects a lower volume of passengers to be demanding of its services. This is also consistent with standard economic intuition and what the BART Agency does.

Table 7: **Summary of Regression Results for MUNI Light Rail**

Dependent Variable **Log MUNI VehHrs**	(1) Base	(2) Constant
Log Gas	-0.0526**	0.0296
	(0.0203)	(0.0514)
MUNI Fare Inc		0.248***
		(0.0538)
MUNI Upg		0.0738***
		(0.0247)
Unemp		-0.0536***
		(0.00980)
Constant	10.83***	11.16***
	(0.0236)	(0.0723)
Observations	164	164
R-squared	0.026	0.709

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

5.3 MUNI Bus

In the previous analysis, we determined that the most suitable model for the MUNI Bus was the Constant model, hence we regress the MUNI Bus Revenue Hours to the Log Gas prices. Using the MUNI Bus Revenue Hours as the Dependent Variable, we find that the Log Gas variable is insignificant when controlling for the other variables. This is somewhat not surprising given the similarity in the BART and MUNI regressions.

Time seasonality is an important consideration factor as the month variables of February, March, June, August, September, October and November are statistically significant. For the year variables, all years investigated except 2015 are statistically significant in the regression.

Compared to the BART and MUNI, it is somewhat surprising that the other control variables such as Unemployment and Fare Increases do not affect the service hours significantly. This could also perhaps be the routes of service, that are not as specific on the routes to employment (offices) compared to the BART and MUNI, and hence the demand for travel is relatively more inelastic.

However, given that there are also no service upgrade variables within MUNI Bus that we can account for, due to the complexity of the lines present, it might be possible that there are Omitted Variable Biases that we have failed to account for that have skewed the results to become insignificant.

Table 8: **Summary of Regression Results for MUNI Bus**

Dependent Variable	(1)	(2)
Log MUNI Bus VehHrs	Base	Constant
Log Gas	-0.0863***	-0.00532
	(0.00973)	(0.0152)
MUNI Fare Inc		0.00416
		(0.0167)
Unemp		-0.00153
		(0.00327)
Constant	11.82***	11.82***
	(0.0108)	(0.0210)
Observations	164	164
R-squared	0.308	0.891

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6 The Correlations of Highway Speeds and Transit Ridership

In an attempt to check the validity and provide extra evidence for the relationship between gasoline prices and public transit ridership, the extension is built to try and investigate the direct effect on driving has on public transportation. Would the relationship be a direct substitution effect, or something more complicated? By analyzing the average driving speed on selected freeways around the Bay Area at their peak usage time (5-7pm), the correlation between the modes of transportation is studied.

In order to control for the quality of the analysis, data points with a data quality below 50% were removed from the analysis. Although this reduced the sample that was being investigated, it maintains the accuracy of the analysis because of the huge amount of data that is being aggregated.

The multivariate regressions from Sections 4 and 5, with the appropriate model chosen for each mode of transport, was used for this analysis. The dependent variable was substituted for the respective highway speeds.

6.1 CA 24E

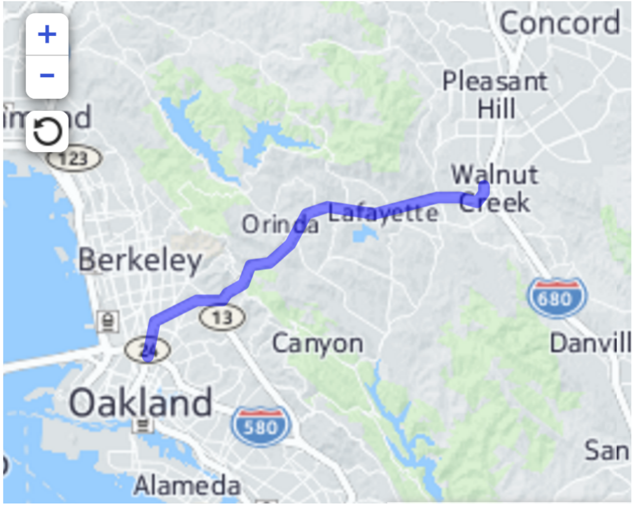

Figure 13: Route Map of the CA 24 Highway Segment studied

This segment of the CA-24E connects San Francisco to the Inner East Bay cities towards Walnut Creek, Lafayette and Orinda. There are more cars owned on average in these cities that lie further away from San Francisco. In these cities, there is a higher proportion of drivers to work. Amongst the three transport means, only the variable for the Log Gas Lag of the BART was significant. This is reasonable as there is a BART line that operates along this stretch, unlike the MUNI and MUNI Bus that operates within the San Francisco City itself. However, the negative sign of the Log Gas Lag variable does not initially match the expectations, as the anticipated effect would be the opposite, as presumably people would switch to the BART when gas prices become more expensive.

However, it seems that people's choices are quite inelastic here, and perhaps are hoping that other people would make the switch due to other reasons such as convenience and comfort etc. Another possibility could also be that there is indirect substitution, such as the switch from driving individually to carpooling or ride-share services like Uber and Lyft, or other public transportation means such as buses.

In general, the adjusted Rsquared values for each of the three transportation modes are high, and have sufficient explanatory power. Also, the significance on month variables like January, April and May also suggest that there is some time-seasonality involved on highway speeds on the CA 24E.

Table 9: **Summary of Regression Results for CA24E**

Dependent Variable	(1)	(2)	(3)
24E	BART	MUNI	MUNI Bus
Log Gas		0.907	0.981
		(2.034)	(2.022)
Log GasLag	-3.381*		
	(2.014)		
BART Fare Inc	-24.89		
	(16.47)		
BART Upg2	1.155		
	(0.810)		
BART Upg3	0.179		
	(0.653)		
MUNI Fare Inc		-2.373	-2.594
		(2.112)	(2.087)
MUNI Upg		0.926	
		(1.170)	
Unemp	-0.335	-0.219	-0.148
	(0.667)	(0.732)	(0.681)
Yearly FE	Y	Y	Y
Observations	86	86	86
R-squared	0.901	0.896	0.895

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

6.2 Interstate 80W

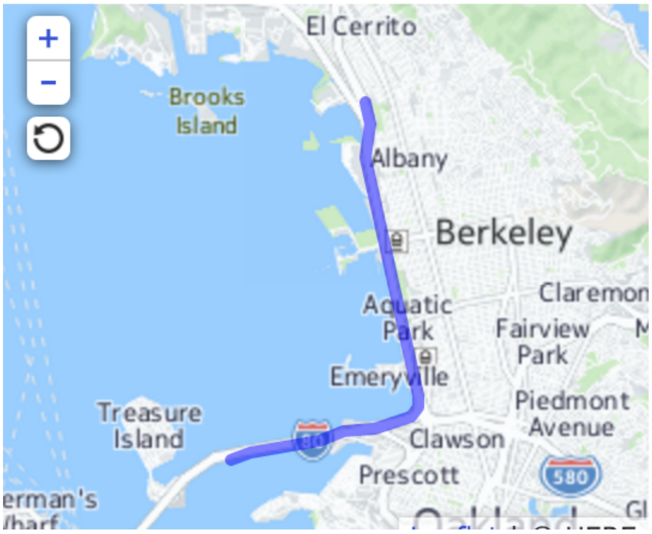

Figure 14: Route Map of the I-80W Highway Segment studied

This segment of the I-80W runs in the Alameda County segment of the freeway, where the BART Service also runs in the East Bay towards Berkeley and Richmond. This can be considered the alternate route taken by drivers who will otherwise be utilizing the BART service.

An important point to note would be that the road lanes in this segment of the freeway are constantly congested during the peak hours of usage, and hence that might affect the coefficients obtained because it may be hard to differentiate the speed change as caused by the difference by consumer behavior.

Based on the regressions, surprisingly, only the MUNI and MUNI Bus models seem to be affected despite this route being a corridor of the BART. A possible explanation could be that it seems that people's choices are quite inelastic here, and that there is indirect substitution, such as the switch from driving individually to carpooling or ride-share services like Uber and Lyft, or other public transportation means such as buses such as the AC Transit which offers a Transbay Bus service into the city.

The MUNI Upgrade variable also returned a significant coefficient. The positive value means that the MUNI Light Rail upgrade is correlated with an increase in the highway speed, which means that presumably more people are substituting away for public transportation within the City.

In general, the adjusted Rsquared values for each of the three transportation modes are high, and have sufficient explanatory power. Also, the significance on all month variables except July and the years 2011-2013 also suggest that there is some time-seasonality involved on highway speeds on the I80W.

Table 10: **Summary of Regression Results for I80W**

Dependent Variable	(1)	(2)	(3)
80W	BART	MUNI	MUNI Bus
Log GasLag	-3.435		
	(2.134)		
Log Gas		6.476***	7.152***
		(2.002)	(1.979)
BART Fare Inc	3.945		
	(16.33)		
BART Upg1	-4.338***		
	(1.125)		
BART Upg2	0.683		
	(1.142)		
BART Upg3	-4.490		
	(3.248)		
MUNI Fare Inc		-0.600	-1.761
		(2.370)	(2.354)
MUNI Upg		2.754***	
		(0.852)	
Unemp	-1.383**	-0.899*	-0.807
	(0.686)	(0.521)	(0.513)
Yearly FE	Y	Y	Y
Observations	99	99	99
R-squared	0.744	0.761	0.751

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6.3 US 101

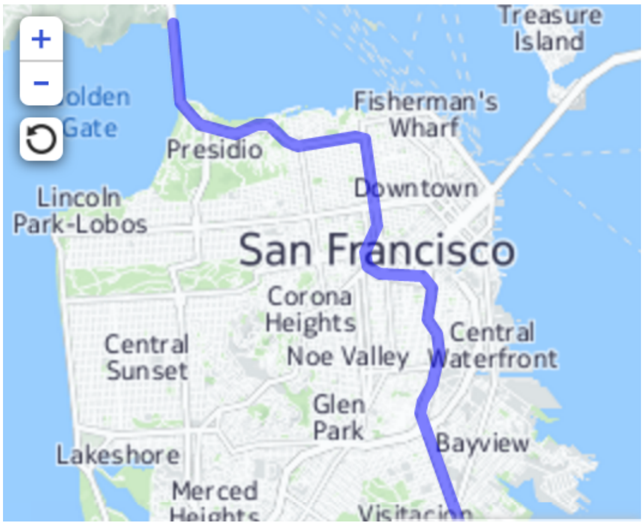

Figure 15: Route Map of the US-101N Highway Segment studied

The US-101N portion was considered a control segment of the highway because none of the major public transit modes studied reaches to Marin County, towards the North of San Francisco. Hence, presumably, the consumer behavior towards driving, which would be a main form of transportation into the city, would likely be more inelastic, and this is demonstrated by the low correlation coefficients to Gas.

However, because there is a significant overlap in route near to Downtown where the BART operates, there is an expectation that perhaps the BART effect might also be captured. Indeed, as expected, all the Log Gas variables are positive in nature, which represents that a higher gas price causes a reduction in the driving, hence resulting in the higher speeds on the freeway observed. In particular, the coefficients on MUNI and MUNI Bus are strongly positive, and depict this effect.

The MUNI Upgrade also is strongly significant to have helped divert some of the driving away. Presumably, one living in the North would have taken the ferry or bus services before making a transfer to reach their intended destination.

In general, the adjusted Rsquared values for each of the three transportation modes are high, and have sufficient explanatory power. Also, the significance on most month variables from January to May and the years 2005/2006 also suggest that there is some time-seasonality involved on highway speeds on the US101.

Table 11: **Summary of Regression Results for US101**

Dependent Variable	(1)	(2)	(3)
US-101	BART	MUNI	MUNI Bus
Log GasLag	1.733		
	(3.118)		
Log Gas		5.646*	6.697**
		(2.954)	(2.906)
BART Fare Inc	-8.477		
	(31.27)		
BART Upg2	-0.256		
	(7.085)		
BART Upg3	-1.185		
	(2.022)		
MUNI Fare Inc		-2.341	-3.491
		(4.821)	(4.740)
MUNI Upg		4.374***	
		(1.510)	
Unemp	-0.282	-0.0315	0.0672
	(0.962)	(0.766)	(0.746)
Yearly FE	Y	Y	Y
Observations	72	72	72
R-squared	0.725	0.756	0.742

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6.4 Takeaways from this analysis

Apart from the inherent problems that data quality presented with, the infrastructure present could also be a reason why we were not able to test out the hypothesis properly. The highway infrastructure systems in Northern California are nearing its maximum capacity, and if they do reach to a state like that in Southern California, minor differences as measured by our regressions will be hard to pick up through the driving speeds.

It could also be that perhaps there is not a simple substitution relationship that we can look into between public transit and private driving. With the advent of many new forms of transportation such as rideshare services such as Uber and Lyft, and the increased platforms for carpooling and incentives from companies to encourage their employees to carpool, the substitution equation is not as discrete as initially perceived.

Although we were not able to get much insights from the I-80 and CA-24E segments, the US101 provided a good control that does show the substitutability effect, confirming the value that the previous segments have proven, leaving us on a good concluding note for further analysis in future research.

7 Further Research Opportunities

There are many potential areas of extensions to the paper to allow the analysis to be more comprehensive. Firstly, in this paper, a short term elasticity was estimated, based on how ridership figures respond to gasoline prices in the relatively short run. However, papers such as Tsai et.al (2014) have argued that the Long Run and Short Run elasticities are different, and that a trend in gasoline prices may change travel pattern and gasoline consumption pattern such as converting to electric or hybrid cars etc. This would be an interesting supplement to this research.

Secondly, this paper does not consider the network effects generated by the various modes of transportation in totality. For instance, a person might choose to take public transport if the modes of transfer between the BART and MUNI were seamless and hassle free. Hence, papers such as Sorensen & Longva (2011) will be useful references to consider on expansion between the coordination effects between transport modes that will refine the estimates in this paper.

Finally, to scope down on the specifics of the data, it would be useful to better understand actual consumer behavior. For instance, the effects of ridership can be separated to monthly pass users and single trip users. Given that the monthly pass offers unlimited access to the mode of public transportation over the specified duration, the relative transit demand would be rather inelastic, given the sunk cost of purchasing the monthly pass, whereas the single trip user will be more likely to consider factors like gasoline prices to decide finally on what mode of transport to utilize. With a finer dataset to work with, this will be helpful in these aspects of expansions. As mentioned in the previous section, the investigation can be expanded to include more recent developments in policy and business through the rideshare applications and carpooling policies and measures.

8 Conclusion

This paper has examined the relationship between gasoline prices and public transit ridership in the San Francisco Bay Area using time series data analysis. Numerous past studies have suggested the relationships between the two, although research for the San Francisco locality and the time-relevant time frame were not present. The significance of this study was to provide an updated cross demand elasticity figure of gasoline on public transportation specific to the San Francisco Bay Area. Following the intuition given from the average cross elasticity values derived from past studies like Currie and Phung (2007) of 0.12, the range of values arrived at for this paper range within 0.0581 to 0.147, which is similar to other study values.

A trend noticed from this study that was different from past studies was the difference in the order of magnitude in the classification of elasticity values. Past studies have determined a greater sensitivity of elasticity for Heavy Rail compared to Light Rail and Bus, but in the present study, a reversal of this trend was observed. As mentioned in the analysis, it is posited that the relative inelasticity arising from the difference in service choice and preference as the main explanation for this phenomenon.

Another key insight justified in this study was the usage of a Constant elasticity model that gives the most explanatory power at least when applied to the San Francisco locality. This is in contrast to the conclusions derived from past studies which used a Events based elasticity model in determining the relationships between gasoline price and transit ridership.

Using the Revenue Hours to regress on the set of independent and control variables, we also discover that the transit agencies currently do not adjust its service schedules to fit the potential gasoline demand relationship this paper has discovered. This opens a realm of policy possibilities to improve the current less-than-ideal service ratings and feedback these agencies are facing. Furthermore, cross checking with that of the highway speed data around East Bay and San Francisco, we can also preliminary see the effects of the substitution between public transportation and private driving through the changing gasoline prices that also has interesting implications to policy decision-making.

It is acknowledged that there are limitations to the data, given the scope of 164 observations and the usage of monthly Unlinked Passenger Trips. Using additional data points from PEMS to analyze the driving speeds, and the revenue hours analysis, we now have a clearer interpretation of the transport ridership figures. It is hoped that these results apply to the San Francisco locality and can possibly serve in ways to the Transportation departments and service operators for more efficient resource allocation to further improve the service standards and quality of public transit.

9 References

1. APTA (n.d.). Public Transportation Use is Growing – Here Are the Facts. Retrieved April 1, 2016, from `http://www.apta.com/mediacenter/ptbenefits/Pages/Public-Transportation-Use-is-Growing-.aspx`

2. APTA (n.d.). Ridership Report. Retrieved December 9, 2015, from `http://www.apta.com/resources/statistics/Pages/ridershipreport.aspx`.

3. BART Marketing & Research Department, Corey, Canapary & Galanis Research. (2015). 2014 BART Customer Satisfaction Study. Retrieved April 1, 2016, from `https://www.bart.gov/sites/default/files/docs/CustSat2014Report_Final.pdf`.

4. Basso, L., & Silva, H. (2014). Efficiency and Substitutability of Transit Subsidies and Other Urban Transport Policies. American Economic Journal: Economic Policy, 1-33.

5. Bureau of Labor Statistics (BLS). San Francisco-Oakland-Fremont, CA Economy at a Glance. (2015, December 8). Retrieved December 9, 2015, from
`http://www.bls.gov/eag/eag.ca_sanfrancisco_msa.htm`.

6. Borenstein, S., Cameron, A., & Gilbert, R. (1997). Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes? The Quarterly Journal of Economics, 305-339.

7. Brady, J. (2015, December 23). Low Gas Prices Expected To Continue In 2016. National Public Radio. Retrieved April 25, 2016, from
`http://www.npr.org/2015/12/23/460826535/low-gas-prices-expected-to-continue-in-2016`

8. Brian, T., & Camille, F. (2003). The Factors Influencing Transit Ridership: A Review and Analysis of the Ridership Literature.

9. California Gasoline and Diesel Retail Prices. (2015, December 7). Retrieved December 9, 2015, from
`https://www.eia.gov/dnav/pet/pet_pri_gnd_dcus_sca_w.htm`.

10. City and County of San Francisco (2016). Customer Rating of Overall Satisfaction with Transit Services. Retrieved April 1, 2016, from
`http://sfgov.org/scorecards/customer-rating-overall-satisfaction-transit-services`

11. Currie, G., & Phung, J. (2007). Transit Ridership, Auto Gas Prices, and World Events: New Drivers of Change? Transportation Research Record: Journal of the Transportation Research Board, 3-10.

12. Fueling California. (n.d.). Quick Facts. Retrieved March 1, 2016, from
`http://www.fuelingcalifornia.org/resources/quick-facts/`

13. Goodwin, P. (1992). A review of new demand elasticities with special reference to short and long run effects of price changes. Journal of Transport Economics and Policy, 26(2), 155-169.

14. Haire, A., & Machemehl, R. (2007). Impact of Rising Fuel Prices on U.S. Transit Ridership. Transportation Research Record: Journal of the Transportation Research Board, 11-19.

15. Hensher, D. A., Mulley, C. (2014). Modal image: Candidate drivers of preference differences for BRT and LRT. Transportation, 42(1), 7-23.

16. Lane, B. (2010). The relationship between recent gasoline price fluctuations and transit ridership in major US cities. Journal of Transport Geography, 214-225.

17. Maley, D., & Weinberger, R. (2009). Rising Gas Price and Transit Ridership. Transportation Research Record: Journal of the Transportation Research Board, 183-188.

18. Small, K., & Dender, K. (2008). Long-run Trends in Transport Demand, Fuel Price Elasticities and Implications of the Oil Outlook for Transport Policy. Oil Dependence ITF Round Tables, 165-200.

19. Sørensen, C., & Longva, F. (2011). Increased coordination in public transport—which mechanisms are available? Transport Policy, 117-125.

20. Sullivan, A. (2003). Urban Economics (5th ed.). Boston: McGraw-Hill/Irwin.

21. Taylor, B., Miller, D., Iseki, H., & Fink, C. (2008). Nature and/or nurture? Analyzing the determinants of transit ridership across US urbanized areas. Transportation Research Part A: Policy and Practice, 60-77.

22. Tsai, C., Mulley, C., & Clifton, G. (2014). A Review of Pseudo Panel Data Approach in Estimating Short-run and Long-run Public Transport Demand Elasticities. Transport Reviews, 102-121.

23. Yanmaz-Tuzel, O., & Ozbay, K. (2010). Impacts of Gasoline Prices on New Jersey Transit Ridership. Transportation Research Record: Journal of the Transportation Research Board, 52-61.