

# Final Project

Yihao Zhang

4/30/2020

## Abstraction

This study is aimed to find the in changes in people's travel behavior in responses to the fluctuations of gasoline prices in Austin city to provide a more comprehensive understanding of the net effects of gasoline prices on transit ridership. We used both static regression and dynamic regression anlaysis to fitted the most accurate predictive model for transit ridership and use that model to determine the estimate effect of change in gasoline price on transit ridership.

The main finding of this study is:

- The price elasticity of gasoline on the demand of transit ridership is 0.042, which indicates a 0.42 percent increase in ridership in response to a 10 percent increase in the current gasoline prices.

This findings provides insight and guidance for how transit agencies can plan and prepare for accommodating higher transit travel needs of the public.

## Introduction

Due to the outbreak of COVID-19 as a global pandemic, oil storage facilities were reaching their limits. On April 20, 2020, U.S. crude oil prices dropped by almost 300 percent to turn negative for the first time in history and reach negative \$37.63 a barrel.

Retail gasoline is a consumer product that is refined from crude oil and its closely related to our daily life. It is a complementary good of motor vehicles, which is one of the main tools for people's daily transportation. The recent dramatical fluctuations of crude oil prices have affected the price of retail gasoline, and this brought the interest in analyzing how people's travel behavior responses to the fluctuations of gasoline prices. Will an increase in retail gasoline price affect people's travel behavior by increasing regional demand on public transportation as an effort to reduce travel expenditure? If so, a comprehensive understanding of how people's demand for public transportation changes in response to a change in gasoline price can provide an important guideline to public transit agencies on capacity management and pricing strategy during a period when gasoline price fluctuates.

Based on this scope, this analysis will use the panel data of public transportation ridership and retail gasoline price from the Austin area from 2016 to 2020 along with a control group of specific factors to estimate the short-run effect of change in gasoline prices on transit ridership.

## Methodology

This analysis used both static linear regression model and a dynamic regression model for metro ridership estimation. The first model estimated the contemporaneous relationship of metro ridership and relevant external variables, and the second model takes both the effect of past observations on metro ridership and relevant external variables to improve the forecast accuracy. The aim here is to estimate the elasticity of

metro ridership to gasoline price and to specify; this study is predicting the percentage change in metro ridership in response to one percentage change in retail gasoline price.

Table 1 illustrates the information of dependent and independent variables included in both regression analysis as well as the source of data. The range of the data site for all variables is from October 2016 to February 2020 monthly.

Table 1: Dependent and Independent Variables Used in Studies as well as Their Specifications.			
Variable	Feature	Description	Source
<b>Metro Ridership</b>	Dependent Variable	Measured the sum of Monthly ridership with MetroAccess, MetroBus, MetroExpress, MetroRail and MetroRapid in Austin City, implemented log transformation afterwards.	U.S. Energy Information administration
<b>Retail Gasoline price</b>	Independent Variable	Used Texas regular conventional retail gasoline price to represent the gasoline price in Austin City. Raw Data acquired is in Weekly basic, transformed into average monthly data, implemented log transformation afterwards.	Capital Metropolitan Transportation Authority
<b>Temperature</b>	Independent Variable	Measured the monthly average temperature (Fahrenheit) in Austin City, implemented log transformation afterwards.	National Weather Service Forecast
<b>Precipitation</b>	Independent Variable	Measured the monthly total precipitation in inch in Austin City, implemented log transformation afterwards.	National Weather Service Forecast
<b>Unemployment</b>	Independent Variable	Measured the monthly unemployment rate in Austin City.	The Bureau of Labor Statistics

As seen in Table 1, in addition to Metro Ridership and Retail Gasoline Price, Temperature, Precipitation, and unemployment rate have been considered as explanatory variables. Temperature and Precipitation are weather indicators that can strongly affect people’s travel behavior. In mild sunny days, more people will be more willing to walk to and from their destinations while on rainy days, more people will prefer to take public transportations. Unemployment is another substantial explanatory variable; a lower unemployment rate means there will be more people who have chances to take the metro to work.

Besides these 4 explanatory variables, several other external variables have been considered to involve in the study and were dropped out later. These variables include metro fare, average household income, and metro service rating. Metro fare and average household income were dropped from the study because both variables remain stable from 2016-2020 every month. Metro service rating has not been considered due to the non-existence of an official data report.

The following sub-sections describe the regression modeling process within the framework of both static data and panel data analysis for transit ridership estimation.

## 1. Static Regression Model

Here, we are modeling a contemporaneous relationship between Metro Ridership and other independent variables with the assumption that the change among the independent variables will have an immediate effect on the dependent variable (metro ridership).

Before building the prediction model, the whole dataset was separated into a training group and a testing group. The training group counts for 80% of the observations recorded from October 2016 to May 2019, and the testing group counts for the rest 20% observations, which is from June 2019 to February 2020.

Next, two-way interaction automatically iterative model selection method is implemented for our model selection. Given the result, models generated from forwarding Selection and Stepwise Selection provided the same information criteria.

- Model1 :  $\log(y) = 14.350 + 0.142\log(x_1) + 0.058\log(x_2) - 0.006\log(x_3)$ ,  $AIC = -152.42$
- Model2 :  $\log(y) = 14.507 - 0.045\log(x_1) + 0.123\log(x_2) - 0.008\log(x_3) - \log(x_4)$ ,  $AIC = -152.42$

where,

$y$  denotes metro ridership in Austin area.

$x_1$  denotes average retail gasoline price in Austin area.

$x_2$  denotes average monthly temperature.

$x_3$  denotes average monthly total precipitation.

$x_4$  denotes average monthly average unemployment rate.

Then, out-of-sample estimation accuracy was computed, and the result is in table 2.

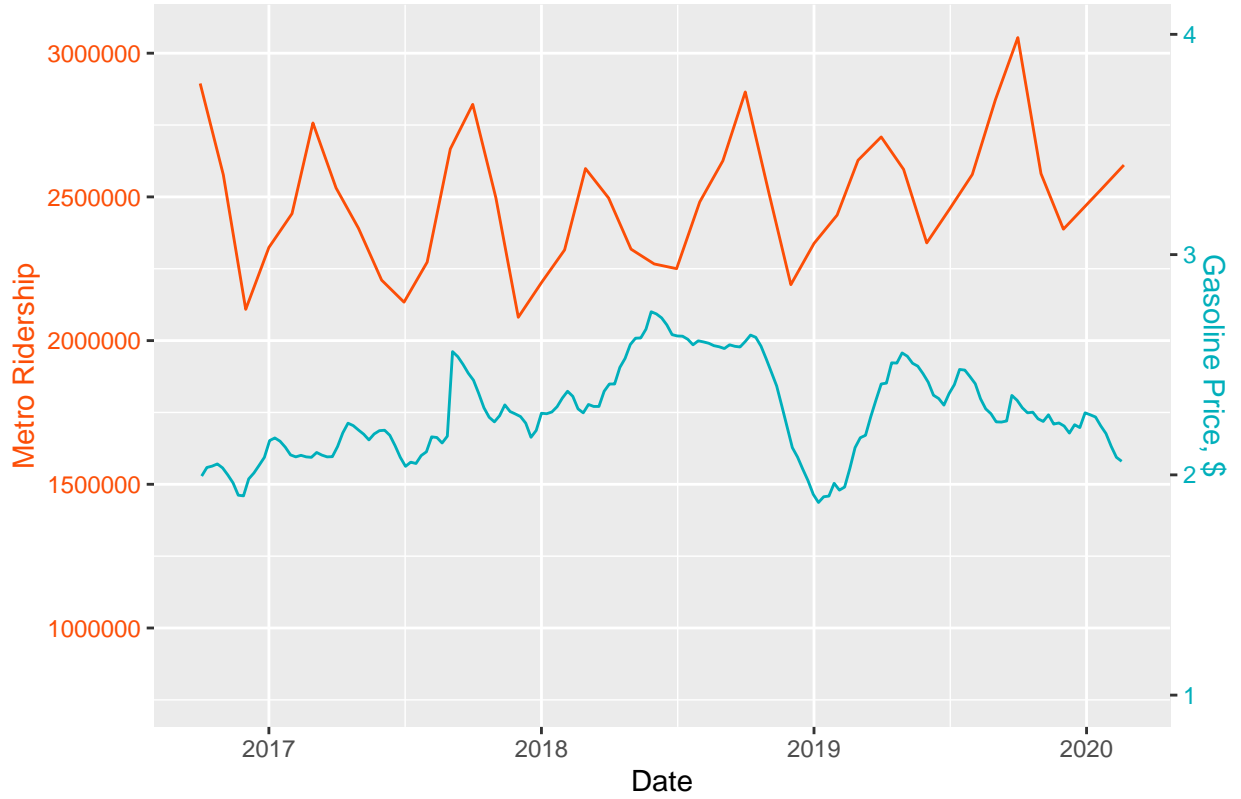
Table 2: Out of Sample Prediction Accuracy		
	MAPE	RMSE
<b>Model 1</b>	0.5063061	0.0973204
<b>Model 2</b>	0.4271024	0.0795031

As seen in Table 2, Model 1 made a better out of sample accuracy performance; it generated a lower RMSE and MAPE. We can conclude that Model 1 is our best static regression model for metro ridership estimation.

## 2. Dynamic Regression Model

Figure 1 shows the relationship between metro ridership - representing the sum of ridership with MetroAccess, MetroBus, MetroExpress, MetroRail, and MetroRapid and average retail gasoline prices in Austin City.

Figure 1: Retail Gasoline Price and Metro Ridership in Austin Area



These two variables follow a similar trend pattern during a certain period. Furthermore, in fact, due to information asymmetry, it is reasonable to suggest that the effect of the gasoline price change on ridership is inherently temporal since people will not be notified simultaneously when the gasoline price changes. So, here we are going to perform a dynamic regression analysis to examining temporal and lagged effects of changes in variables.

Here we have equation 1 which express the specification that the model I will fit using panel data with a group of control variables.

equation1:

$$y_t = a_0 + \beta x_t + n_t + \epsilon_t$$

where,

$y_t$  denotes metro ridership in Austin area at time  $t$ .

$a_0$  denotes the intercept parameter.

$\beta$  are vector of slope parameters associated with external influential factors.

$x_t$  is the vector of external influential factors that include gasoline price, average temperature, total precipitation, unemployment rate in Austin area at time  $t$ .

$n_t$  denotes the stochastic error term corresponding to the regression for the mode.

$\epsilon_t$  denotes error from the ARIMA model.

After implemented the same training and testing set separation method that used for the static model, that the training set counts for observations recorded from October 2016 to May 2019, and the testing set counts for observations from June 2019 to February 2020. Stationarities of the training series has been tested to be true afterward, and several dynamic models have been fitted to the series.

Next, we select model 3 as our best prediction model using the dynamic regression method, which provides the lowest information criteria.

Model 3:

$$\log(y_t) = 14.061 + 0.0422\log(x_{1t}) + 0.151\log(x_{2t}) - 0.001\log(x_{3t}) - 0.038\log(x_{4t}) + n_t$$

$$n_t = 0.681n_{t-1} + 0.887n_{t-12} - 0.604n_{t-13} + \epsilon_t$$

where,

$x_{1t}$  denotes average retail gasoline price in Austin area at time t.

$x_{2t}$  denotes average monthly temperature at time t.

$x_{3t}$  denotes average monthly total precipitation at time t.

$x_{4t}$  denotes average monthly average unemployment rate at time t.

Then, we compute this model's out-of-sample accuracy performance. It generates a MAPE of 0.27 and a RMSE of 0.045, comparing to our Model 2, Model 3 provides a better of sample accuracy and we conclude that model 3 is our best prediction model for metro ridership.

## Results

This section presents results from a series of panel data using regression analyses of our best prediction model. After performed log-log transformation on both metro ridership and gasoline price, the estimated coefficient on gasoline can be interpreted as the price elasticity of gasoline on metro ridership. Here the coefficient is 0.042, which tells that a 10% increase in gasoline price in Austin city is expected to cause metro ridership to increase by 0.42%. This result is consistent with our hypothesis and indicate that the increase in gasoline price has a positive effect on metro ridership demand, people change their travel behavior by increasing regional demand on public transportation as an effort to reduce travel expenditure in response to the increase in gasoline price. But the sensitivity of this effect is low. This could be explained that people do not always notify the small level of changes in gasoline price and it is not possible for them simultaneously receive the information of the change in gasoline price. And most people will notify the change in gasoline price only when the difference is huge.

Meanwhile, the coefficient on our control variables have the same as we expected. The positive effect of temperature, the negative effect of precipitation and the negative effect of unemployment on metro ridership. The estimated effect on all explanatory variables has listed in table 3:

Table 3: Coefficient of Explanatory Variables	
variable	coefficient
Constant	+14.061
Log of monthly gasoline price	+0.0422
log of average monthly temperature	+0.151
log of average monthly total precipitation	-0.001
monthly unemployment rate	-0.038

## Conclusion

This study examined the net effects of changes in gasoline prices on metro ridership in Austin city. A 10% increase in gasoline price in Austin city is expected to cause a 0.42% of increase in metro ridership demand.