The Effect of US Gas Prices on Alternative Methods of Travel

Group #104

Caroline Schmitt, Emy Ng, Matthew Kim, Mike Genovese, and Osman Yardimci

Background Information

Today, we'll look at an intriguing research project that investigates the link between changes in fuel costs and public transit usage in key cities around the United States.

This study was inspired by a simple yet deep question: 'How do changes in gasoline prices affect the usage of alternate forms of transportation?' Answering this topic has far-reaching ramifications for city planning, policymaking, environmental concerns, and affordable commute alternatives for common residents.

Navigating Complex Transit Data: From Analysis to Insights

We used a variety of data analysis techniques throughout this study, from vector-autoregressive models to Spearman Rank Correlation, with the goal of decoding complex patterns, extracting meaningful insights, and providing key stakeholders, primarily public transportation agencies, with actionable, data-driven knowledge for their decision-making processes.

Our voyage led us through the extensive transportation statistics of five of the most important US markets: New York City, Chicago, Los Angeles, San Francisco, and Washington, D.C. We dug deep into these marketplaces, eliciting dynamic interactions and watching how differences in gasoline grades added levels of complexity to these relationships.

Problem Statement

Ridership trends for alternative methods of transportation like public transit and bike share programs aren't publicly modeled.

Objective:
The purpose of our analysis will be to dig into the relationship between alternative transport ridership numbers and how the price of gasoline in the US affects these numbers.

Business Justification

Our findings will confirm if alternative transportation organizations and companies can utilize these forecasts when planning their business operations.

Impact:

This will profit maximum profits by not having a surplus of resources while demand is lower than expected or alternatively not having not enough resources available to meet demand.

Research questions

1. How does the average US gas price over time affect ridership numbers of alternative transportation methods?

2. Which areas show the most increase in transit ridership in response to increased gas prices?

3. Do all grades of gasoline fluctuate in price at roughly the same rate? And if not, does one affect alternative transportation ridership more than others?

4. Are there any particular significant events that explain any sudden spikes in alternative transportation ridership numbers?

Hypothesis

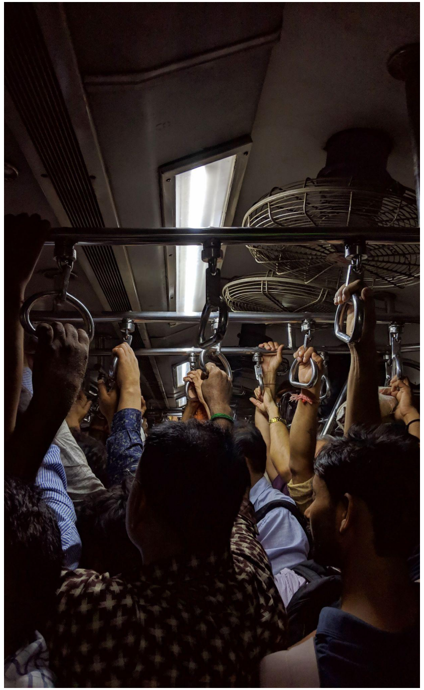
We're expecting:

01 | A positive relationship between gas prices and transit ridership.

02 | Most significant independent variable will be regular-grade gas price

03 | Some relationship between gas prices and transit ridership

04 | We don't expect to be able to explain any of our variables of interest completely.



Methodology/Approach

Spearman Rank Correlation

To help solve our research questions related to correlation and the effects of gas prices in the United States.

Granger Causality Tests

We will also be working with Granger Causality Tests to add another layer to our correlation tests.

Time Series Models

Hyperparameter Optimization (ACF and PACF plots, grid searching, and auto-ARIMA).

Utilizing multiple different model types to test data

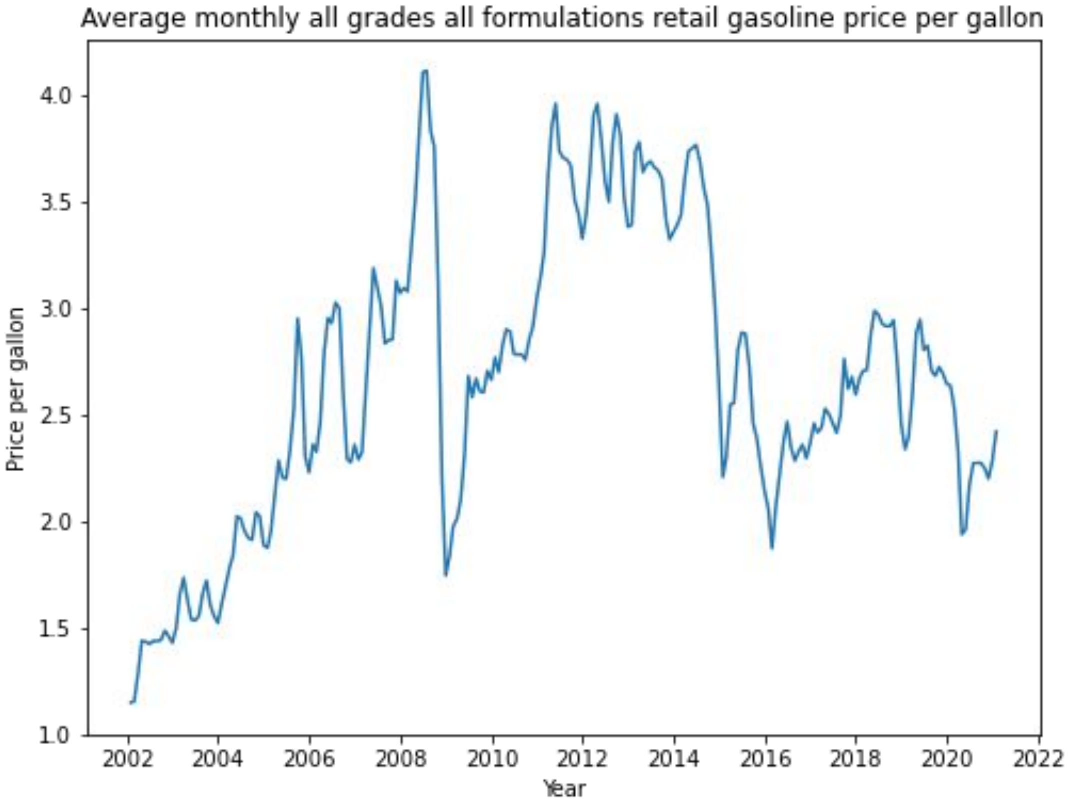
Data

U.S. Gasoline and Diesel Retail Prices dataset:

An exploratory line plot of average monthly gasoline prices, across all grades and formulations.

Note the obvious impact of the 2008 recession, the 2014-2016 oil price collapse, and COVID-2019.

Figure 1. Monthly gasoline prices

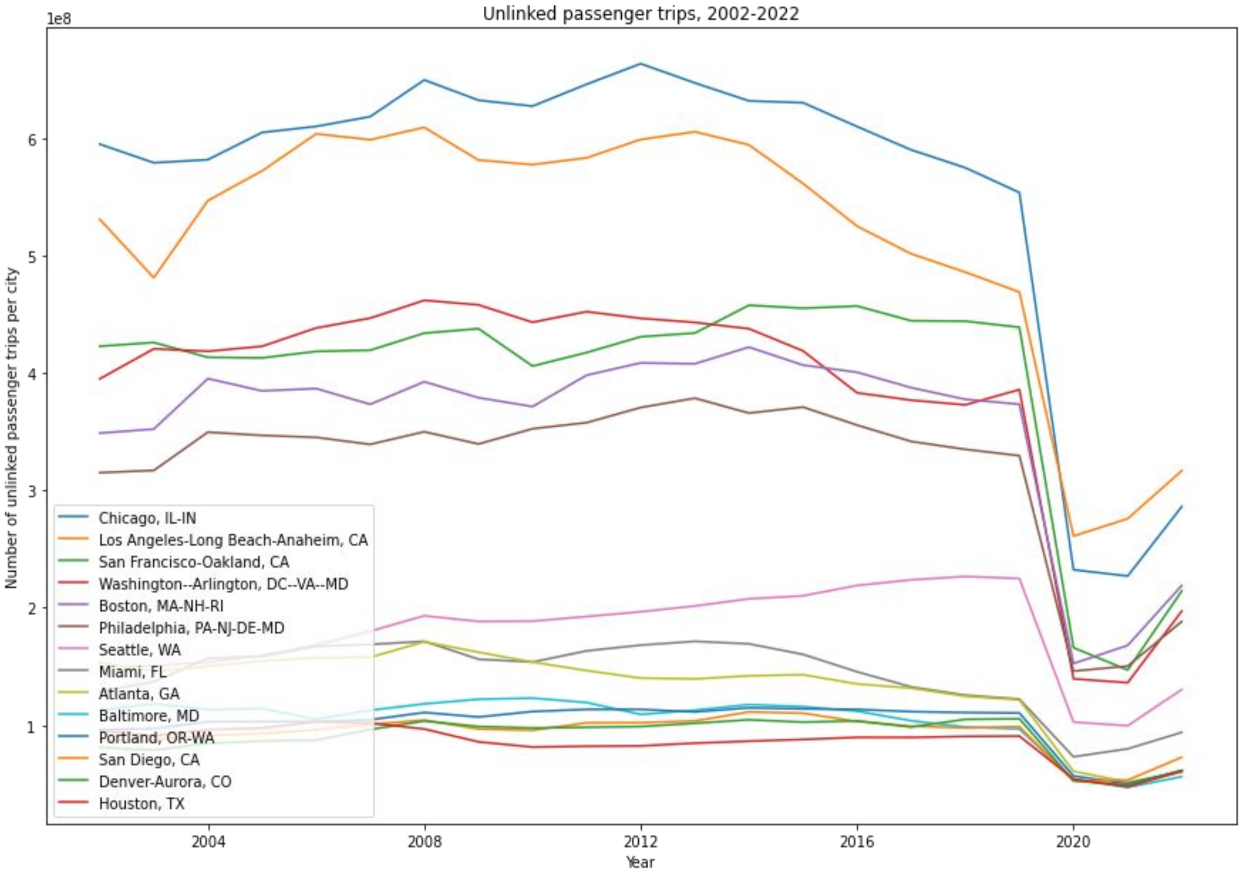


National Transit database complete monthly ridership dataset:
An exploratory line plot of unlinked passenger trip counts from 15 of the highest-trip cities.

Unlinked passenger trips is a measure of passengers who board public transit vehicles -- unadjusted for linked trips, e.g. connecting to a bus from a train as part of the same trip.

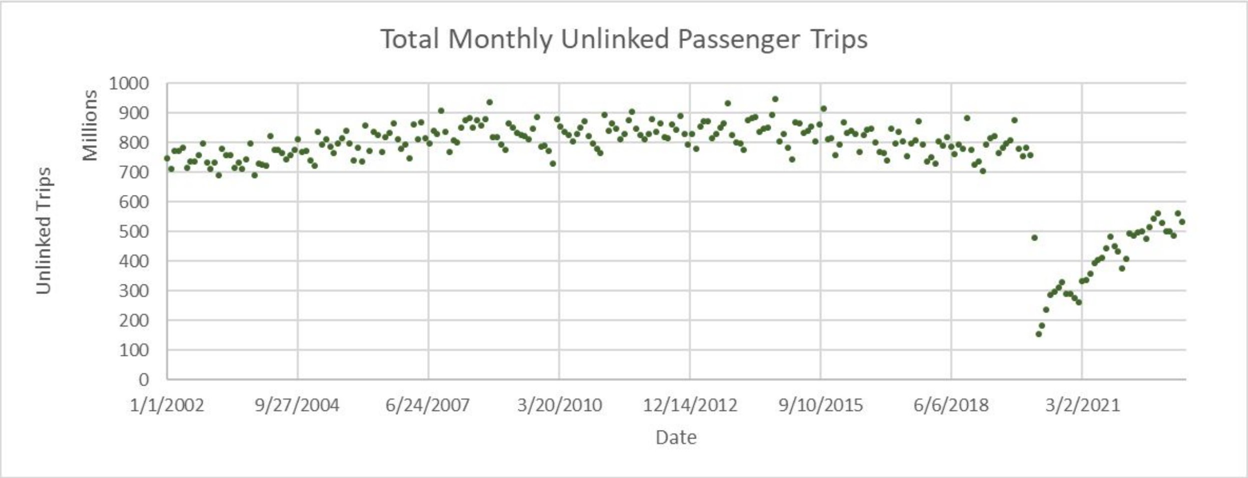
Note: NYC is an extreme outlier and is not included in this plot.

Figure 2. UPTs in top 15 cities



The exploratory analysis shows that Covid-19 had a heavy impact on the number of travelers using public transportation. This could be due to factors like transit services shutting down or decreasing operations, the increase in work-from-home jobs, and decrease in workforce participation mixed with an increase in unemployment. This could affect the data modeling stage so we need to be aware and alter our models if needed.

Figure 3. Total Transit Ridership (Jan. 2002 - Apr. 2023)



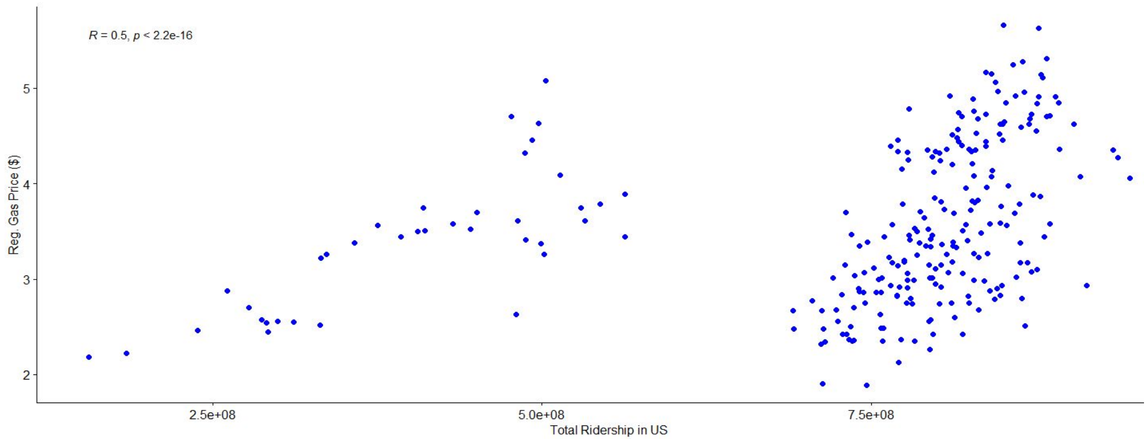
Spearman Rank Correlation Between Reg. Gasoline and Total Ridership

- Rank all monthly gasoline prices and monthly total transit ridership from lowest to highest.
- Delta of the ranks for each month taken, squared, and fit into Spearman formula.
- Resulting R value ranges from -1 to 1, with 1 being a perfected correlated positive relationship.

Test 1: All Data Points

- R = 0.5
- p = 2.2e-16

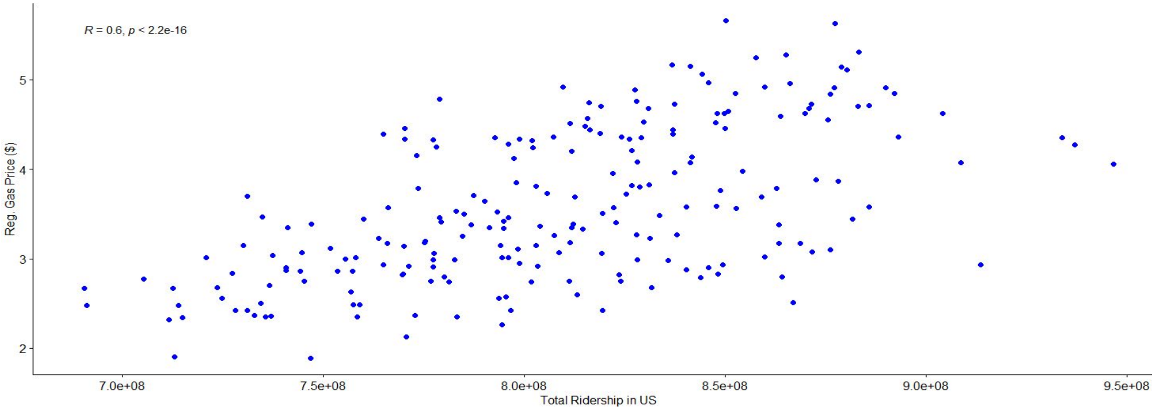
Figure 4. Reg. Gasoline vs Total Ridership in US Correlation (Jan. 2002 - Apr. 2023)



Test 2: Pre-Covid Data Points

- R = 0.6
- p = 2.2e-16

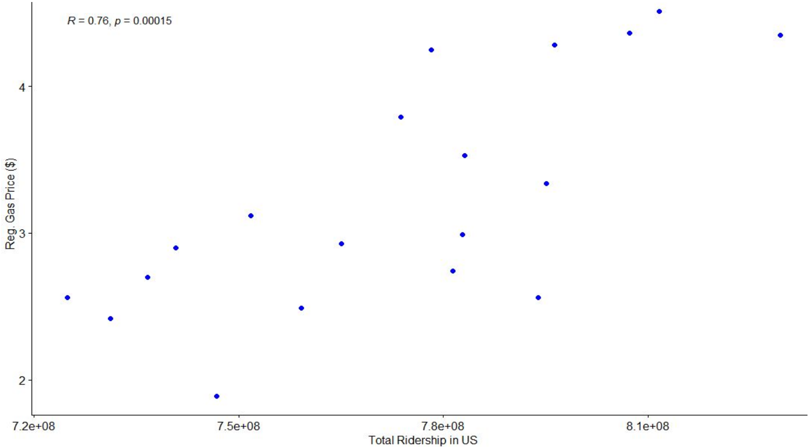
Figure 5. Reg. Gasoline vs Total Ridership in US Correlation (Jan. 2002 - Feb. 2020)



Test 3: January YoY Data Points

- R = 0.76
- p = 0.00015

Figure 6. Reg. Gasoline vs Total Ridership in US Correlation (Jan YoY 2002 - 2020)



Granger Causality Test Between Reg. Gasoline and Total Ridership

- Statistically hypothesis test - if one time-series is effective in forecasting for another
  - “Predictive causality” not necessarily direct causality

Table 1. Granger Causality Results (Reg. Gas Prices vs. Total Ridership)

Dataset	F Test Statistic	p-value
All Dates	1.8285	0.1775
Pre-Covid	42.753	4.487e-10 ***
January Year-Over-Year	5.1957	0.03769 *

Modeling approach

- Studied a subset of markets: NYC, Chicago, Los Angeles, San Francisco, and Washington, D.C.
- ARCH/GARCH models: exploratory analysis showed ridership data not volatile enough to be a good fit
- Seasonal decomposition: useful for analyzing region-level transit data, but gas prices too volatile
- Vector autoregressive (VAR) models: allow us to analyze and predict multiple measures as they change over time, we used weekly real gasoline prices and weekly total UPT/market
- Used autocorrelation function (ACF) and partial autocorrelation function (PACF) plots to identify lag order (number of previous observations to use)
- Data required scaling
- Evaluated models on both training and testing data, on unlinked passenger trips only

Vector autoregressive model performance

Table 2. Vector Autoregressive Model Results

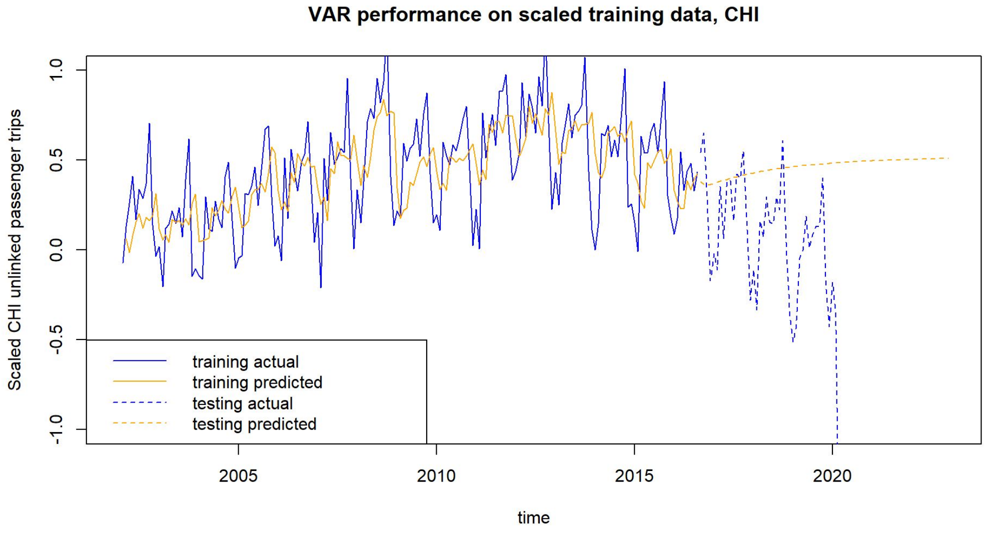
Market and order	**Training MAE**	**Testing MAE**
NYC (order=2)	0.212	1.61
CHI (order=1)	0.180	1.49
LA (order=2)	0.180	0.585
SF (order=2)	0.183	1.54
DC (order=1)	0.229	1.25

Mean absolute error: a measure of average model wrongness (smaller scores are better)

Training MAE: model error on data the model was trained on

Testing MAE: model error on data new to the model

Figure 7. VAR Performance



Looking at a plot of VAR performance, it's clear COVID-19 related shutdowns are also contributing to poor test data scores. VAR models alone don't handle sudden shocks well; adding more complex terms to the model can help address this.

Conclusions

- Through Spearman Rank and Granger Causality tests we found reg. gasoline prices are correlated and significant in forecasting alternative transportation ridership numbers.
- Although our models didn't perform very well, ability to capture movements in the training data is promising
- One problem may be that the largest transit markets in our dataset are already at "public transit saturation" -- i.e. everyone in NYC who is price sensitive already uses public transit at all times

Next steps & recommendations

- Study more markets -- likely more interesting patterns in other cities
- Study specific modes of alternative transportation from our data set
- Compare vector autoregressive models to univariate autoregressive models
- More specialized models like change point detection models

📌Citations

- Public transit usage and air quality index during the COVID-19 lockdown (PubMed)
- Transit Price Elasticities and Cross-Elasticities (Victoria Transport Policy Institute)