Assessing the relationship between vehicle miles traveled, public transit ridership, and gas prices.

Tone Agbeyegbe[1]

Rutgers University

August 2023

Abstract

This study investigates the relationship between gas prices, public transit ridership, and vehicle miles traveled (VMT) in the United States from September 2000 to March 2023. This study explored the correlation and potential effects of changing gas prices on transportation habits. The analysis included correlation analysis, regression analysis, decision trees, random forest, and gradient boosting using cross-validation. The research question driving this study is whether changes in gas prices affect the rate of change in public transit ridership and vehicle miles traveled. The findings revealed that the relationship between gas prices and transportation habits is complex, along with other factors such as the convenience of employment, income levels, and transportation decisions.

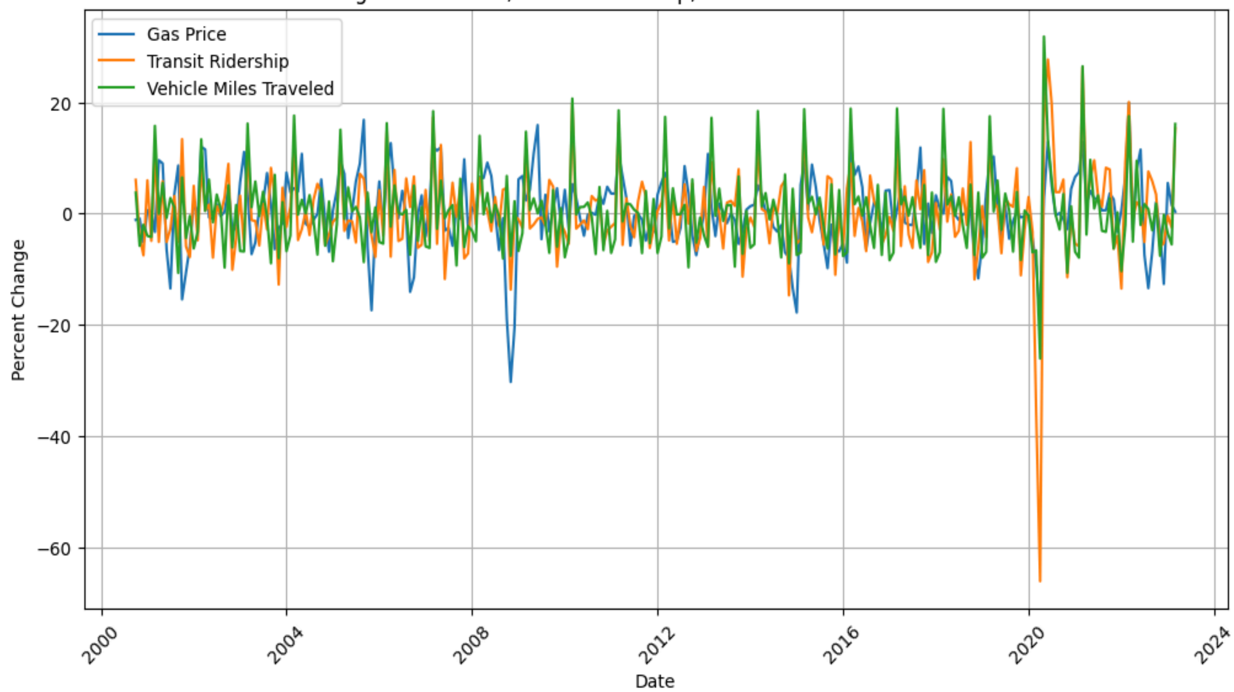
Introduction

Millions of Americans make daily decisions on gasoline purchases and modes of transportation. Common economic concepts dictate that when the price of a commodity rises, there is demand for its alternatives. Consequently, one may expect public transportation usage to be positively connected with petrol costs as people seek less expensive alternatives to driving. However, a study of the available data revealed a more complicated relationship.

Figure 1 illustrates the percentage changes in gas prices, public transit ridership, and VMT over the past 22 years. Contrary to initial expectations, gas price increases coincided with public transit ridership and the VMT. This unexpected correlation raises questions regarding the intricate dynamics of Americans' daily transportation decisions.

The data in Figure 1 were not seasonally adjusted, which may explain some of the observed variations. The influence of seasonal factors becomes apparent, as all three indicators display spikes in March, and transit ridership and vehicle miles consistently increase throughout the summer months when Americans tend to travel more.

Fig. 1 Gas Prices, Transit Ridership, and Vehicle Miles Traveled



Given these noteworthy tendencies, this study aimed to explore the pace of change in petrol prices, public transport ridership, car miles traveled, and their interplay over the period under consideration. This study aims to shed light on the complex relationship between gas prices and transportation patterns in the United States using correlation analysis, regression analysis, decision trees, random forest, and gradient-boosting approaches. The findings will help politicians, urban planners, and transportation specialists develop effective ways to promote sustainable and efficient transportation systems by providing more profound knowledge of the factors driving American mobility choices.

The remainder of this paper is organized as follows. Section 2 provides a brief review of the relevant literature. Section 3 discusses the methodology used in the study. It begins by discussing basic regression modeling and ends with a description of sophisticated machine-learning techniques. Section 4 discusses the data and empirical results. Section 5 discusses seasonal adjustments and Section 6 concludes the paper.

Section 2 Literature Review

This section presents a brief literature survey on the interrelationships between vehicle miles traveled (VMT), gas prices, transit ridership, and related topics.

Brownstone and Golob (2008) examined the correlations among residential density, household car usage, and vehicle fuel consumption. These findings indicate that residential density and gas prices notably influence VMT. Lane (2011) investigated the temporal dimensions of the correlation between public transportation and fuel costs in cities across the United States. This study demonstrates that persistent fluctuations in fuel costs can result in enduring shifts in public transportation usage, thus confirming the substantial influence of gasoline prices on transit ridership and VMT.

Maley and Weinberger (2012) confirmed a statistically significant relationship between gas prices and transit ridership increases, using Philadelphia as a case study.

In a subsequent study, Nowak and Savage (2013) analyzed the determinants of cross-elasticity between the price of gasoline and the ridership of public transit in Chicago. The findings of the study suggest that fluctuations in gasoline prices substantially affect public transportation usage, demonstrating an inverse association between gas prices and VMT as well as a positive correlation between gas prices and the number of individuals utilizing transit services.

A Congressional Budget Office Report (2014) discusses how rising gas prices influence the volume of traffic, consumers' choices of vehicles, and the shift towards public transportation from 2003 to 2006 in California. This suggests that increasing gas prices significantly impacts driving behavior, indicating a negative correlation between gas prices and VMT.

Hymel (2014) examined the factors that influence vehicular travel, focusing on why the growth in per capita VMT has stalled in recent years despite economic recovery. This indicates that changes in household demographics and economic characteristics rather than driving habits explain most of the current variation in VMT.

Dillon et al. (2015) investigated the combined impact of urban design and gasoline prices on household VMT. This finding illustrates the substantial effect of urban design on vehicle usage, suggesting that built environment characteristics may also affect VMT.

Hyrc (2017) discussed how increased gas prices could potentially lead Americans to travel more often using public transportation. Her article documented the behavior of U.S. regular conventional gas prices, public transport ridership, and vehicle miles traveled from January 1, 2012, to January 1, 2017. She found that once gas prices decreased, the relationship between these three variables became unclear.

Leard et al. (2021) examined the factors that explain changes in passenger VMT in the United States, focusing on changes in the demographics and economic characteristics of households. This study suggests that rather than changes in driving habits, these factors explain most of the recent variations in VMT.

In their study, Jung, Yu, and Kwon (2022) investigate the impact of fluctuations in gasoline prices on the decision-making process of individuals when choosing between private automobile usage and public transit attendance. The study revealed that fuel costs substantially affected individuals' choice between using a personal automobile or opting for public transportation, reinforcing the inverse correlation between gasoline prices and VMT.

The literature indicates a complex interrelationship between VMT, gas prices, transit ridership, and other factors such as residential density and urban form. Rising gas prices tend to reduce the VMT and increase transit ridership. However, other factors such as changes in household demographics and economic characteristics, urban structure, and residential density also significantly impact these trends.

Section 3 Methodology

This study utilized a combination of correlation analysis, regression analysis, decision trees, random forest, and gradient-boosting techniques to examine the relationship between gas prices, public transit ridership, and vehicle miles traveled in the United States. These analytical methods were employed to gain insight into the complex dynamics and predictive power of the investigated variables. See James, Witten, Hastie, and Tibshirani (2013) for further details. Cross-validation was performed to enhance the reliability and robustness of the analysis.

The study utilized Correlation analysis was used to quantify the strength and direction of the linear relationship between variables. By calculating correlation coefficients such as Pearson's correlation coefficient, this study assesses the degree of association between gas prices, public transit ridership, vehicle miles traveled, and other related variables. This analysis helps to determine whether any significant relationships exist and provides a preliminary understanding of the interdependencies among the variables.

In addition, this study initially performed a regression analysis of the variables of interest. Regression analysis was employed to establish statistical models that capture the relationships between variables and predict the outcomes. By fitting regression models, this study can estimate the effect of gas prices on public transit ridership and VMT, while accounting for other influential factors. Regression analysis allows for the identification of key variables and quantification of their impact, aiding in the formulation of insights and policy recommendations.

I considered several linear regression models based on regressing the dependent variable on independent variables. They follow the specification in equation (1):

The analysis starts with a basic linear regression model in which the dependent variable is vehicle miles traveled (VMT), and the independent variables are gas prices, public transit ridership, and other potential control variables in growth terms.

The general form of the model used for estimation is as follows:

$$\text{VMT} = \beta 0 + \beta 1 \times \text{Gas Price} + \beta 2 \times \text{Public Transit Ridership} + \beta 3 \times \text{Control Variables} + \epsilon \qquad \text{equation (1)}$$

where:

- β_0 is the intercept,
- β_1 is the coefficient for gas prices,
- β_2 is the coefficient for public transit ridership,

- β_3 is the vector of coefficients for the control variables
- ϵ is the random error term.

The initial simple model included growth in both gas prices and public transit ridership as the independent variables. I then gradually added more control variables to examine how they affected the relationship between changes in VMT, gas prices, and public transit ridership.

The following estimate focuses on decision trees. Decision trees are advantageous because they can handle numerical and categorical data, and their graphical representation allows for interpretability and visual inspection of the decision-making process. Decision trees explore and uncover nonlinear relationships and interactions among variables. Decision trees are hierarchical structures that recursively partition data based on features, ultimately leading to predictions or classifications. This study employs decision trees to identify essential decision points and breakpoints related to gas prices, public transit ridership, and vehicle miles traveled. This approach allows for a more nuanced understanding of the complex relationships between the variables. The goal is to identify the most significant decision points or breakpoints that best explain the variations in the target variable (e.g., public transit ridership or vehicle miles traveled) based on the input variables (e.g., gas prices).

More precisely, decision trees are an analytical technique used to uncover essential decision points and breakpoints in the relationships between variables. In the context of this study on gas prices, public transit ridership, vehicle miles traveled, and other related variables, decision trees can provide insights into the factors that influence transportation choice.

To illustrate this process, let us consider the example of using gas prices and public transit ridership. A decision tree can split data based on the threshold gas price. The tree evaluates whether the gas prices are above or below this threshold, and divides the dataset accordingly. Thus, the decision tree can identify distinct groups or segments based on the gas prices.

After the initial split, the decision tree recursively splits the data based on other features such as vehicle miles traveled or other influential factors. This process identifies additional decision points or breakpoints that contribute to the variations in public transit ridership. The decision tree algorithm selects the most informative features and breakpoints by evaluating the extent to which the data are separated into homogenous groups or subsets. It employs criteria such as information gain, Gini index, or entropy to measure the purity or homogeneity of the resulting groups. The algorithm aims to maximize the separation between groups and minimize the impurities within each group.

The resulting decision tree visually represents the decision points and breakpoints identified during the analysis. Each branch represents a decision based on a feature or attribute, leading to subsequent branches until an outcome or prediction is achieved. The decision tree highlights the most influential features and their corresponding breakpoints in explaining variations in the target variable.

In the context of this research, decision trees can help identify critical gas price thresholds, along with other influential factors that significantly impact public transit ridership or vehicle miles traveled. By examining the decision tree, researchers can gain insight into the specific conditions under which individuals are more likely to choose public transit or driving, shedding light on the complex relationships between gas prices, transportation habits, and decision-making processes.

It is important to note that decision trees are interpretable and provide a clear understanding of the decision rules learned from data. However, they are prone to overfitting or noise capture in data. Random forest and gradient boosting are often employed to improve the stability and predictive performance by combining multiple decision trees to address these limitations. See James, Witten, Hastie, and Tibshirani (2013) for details.

A random forest approach follows the decision tree estimation method. Random forest, a powerful ensemble learning technique, was employed to improve the accuracy and stability of the predictive models. Random forest constructs multiple decision trees using different data subsets and features. The RF model provides robust and reliable forecasts by aggregating the predictions of individual trees. This technique helps to mitigate overfitting and increases the generalizability of the results.

The final method is gradient boosting. The final prediction was obtained by summing the predictions from all the trees. It effectively captures nonlinear relationships, handles missing data, and provides accurate prediction. This iterative ensemble technique combines weakly predictive models to create a more potent overall model. By sequentially building new models that focus on the errors of previous models, gradient boosting improves predictive accuracy and captures complex relationships that may not be apparent in individual models.

Cross-validation was implemented to evaluate the performance and generalizability of the models. Cross-validation involves partitioning data into multiple subsets, training the models on one subset, and validating the results on the remaining subsets. This process helps to assess how well the models predict new and unseen data outcomes. Using cross-validation, the study ensures that the models are not overly optimized or biased toward the training data, thereby increasing the reliability and validity of the findings.

Including these various techniques—correlation analysis, regression analysis, decision trees, random forest, gradient boosting, and cross-validation—enables a comprehensive and robust analysis of the relationship between gas prices, public transit ridership, and vehicle miles traveled. Thus, by employing multiple analytical approaches, this study aims to uncover the underlying patterns, understand the complexities of the interactions of the variables, and provide valuable insights for policymakers, urban planners, and transportation experts to inform decision making and promote sustainable transportation systems.

Section 4: Data and Empirical Analysis

This section describes the data used in the analysis and provides the empirical results. Monthly data were obtained from The Federal Reserve Bank of St. Louis database (2023). The data cover the period from 2000.9 to 2023.3. The dataset contains several variables.

DATE

GASREGCOVM (U.S. Regular Conventional Gas Price, Dollars per Gallon, Monthly, Not Seasonally Adjusted)

TRANSIT (Public Transit Ridership, Thousands of Unlinked Trips, Monthly, Not Seasonally Adjusted)

VMTUSM227NFWA (Vehicle Miles Traveled, Millions of Miles, Monthly, Not Seasonally Adjusted)

CPIAUCSL (Consumer price index for all urban consumers: All items in the U.S. City Average, Index 1982-1984=100, Monthly, Seasonally Adjusted)

POPTHM (Population, Thousands, Monthly, Not Seasonally Adjusted)

UNRATE (Unemployment Rate, Percent, Monthly, Seasonally Adjusted)

P.I. (Personal Income, Billions of Dollars, Seasonally Adjusted Annual Rate)

RATE (Finance Rate on Consumer Installment Loans at Commercial Banks, New Autos 48 Month Loan, Percent, Monthly, Not Seasonally Adjusted)

CUSR0000SETA (Consumer Price Index for All Urban Consumers: New and Used Motor Vehicles in U.S.) City Average, Index Dec 1997=100, Monthly, Seasonally Adjusted)

TOTALSA (Total Vehicle Sales, Millions of Units, Monthly, Seasonally Adjusted Annual Rate)

The summary statistics, distributions, and plots of the relationships between variables are reported below in Table 1

Table 1. Summary Statistics - Group 1

	GASREGCOVM	TRANSIT	VMTUSM227NFWA	CPIAUCSL
count	271	271	271	271
mean	2.54945	782441	251754	225.42
std	0.755808	157242	19728.7	31.4082
min	1.072	171450	167617	173.6
25%	2.0365	779331	239586	201
50%	2.497	829115	253936	227.223
75%	3.144	869948	265102	244.713
max	4.764	993437	293308	301.808

Summary Statistics - Group 2

	POPTHM	UNRATE	PI	RATE	CUSR0000SETA	TOTALSA
count	271	271	271	271	271	271
mean	312074	5.88672	14248.4	6.01428	100.507	15.8973
std	15782.1	1.96696	3894.29	1.38477	8.07504	2.28802
min	283033	3.4	8792.1	4	91.562	8.923
25%	298286	4.5	11318.5	4.85	95.8	14.7465

50%	313454	5.4	13582.4	5.84	99.275	16.695
75%	327040	6.75	16878	7.0575	100.824	17.478
max	334753	14.7	24371.9	9.64	129.83	22.055

The summary statistics provided initial insights into the dataset.

The gas price (GASREGCOVM) ranged from $1.07 to $4.76 per gallon, with a mean price of approximately $2.55. Public transit ridership (TRANSIT) varied from 171,450 to 993,437 thousand unlinked monthly trips, with a mean of approximately 782,441. The vehicle miles traveled (VMTUSM227NFWA) ranged from 167,617 to 293,308 million miles per month, with a mean of approximately 251,754 million miles. The consumer price index (CPIAUCSL) varied from 173.6 to 301.8, averaging approximately 225.4. The Population (POPTHM) increased from 283,033 to 334,753 thousand, with an average of roughly 312,074 thousand. The unemployment rate (UNRATE) ranges from 3.4% to 14.7%, averaging approximately 5.9%. Personal income (P.I.) varied from $8,792.1 billion to $24,371.9 billion, with an average of approximately $14,248.4 billion. The finance rate of consumer installment loans for new autos (RATE) ranged from 4% to 9.64%, with an average of approximately 6%. The consumer price index for new and used motor vehicles (CUSR0000SETA) varied from 91.562 to 129.83, averaging at approximately 100.5. Total vehicle sales (TOTALSA) ranged from 8.923 to 22.055 million monthly units, averaging roughly 15.9 million units.

The graph shows the time paths of nominal gas prices, transit ridership, and VMT in the various panels of Figure 2.

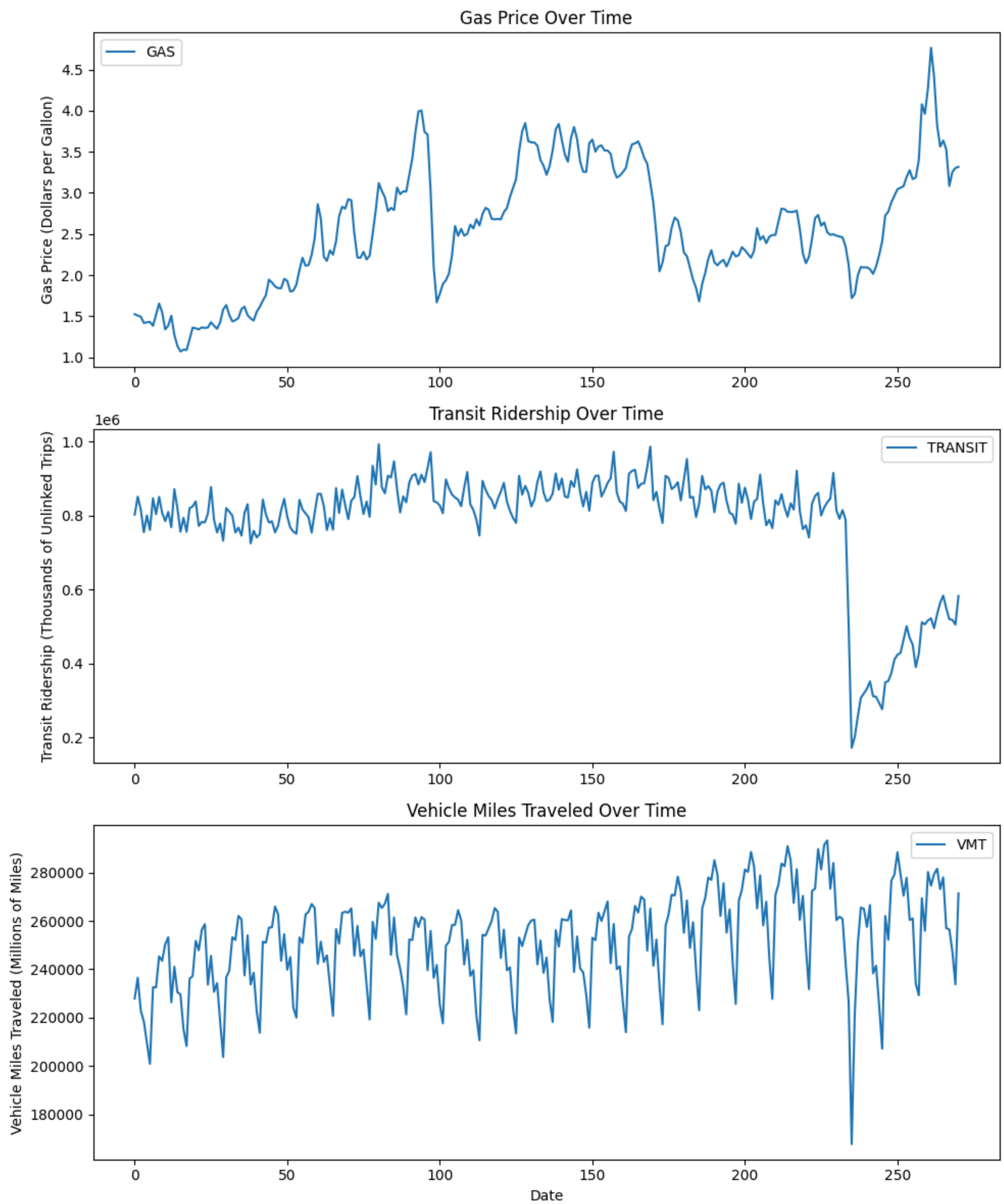


Figure 2. Top panel

This figure shows the cost of gas in dollars per gallon from 2000 to 2003. It rose until it fell dramatically to $1.60 during the 2009 stock market crash. Notably, the cost of gas has risen steadily to its peak after the COVID pandemic in 2022.

Figure 2. Middle Panel

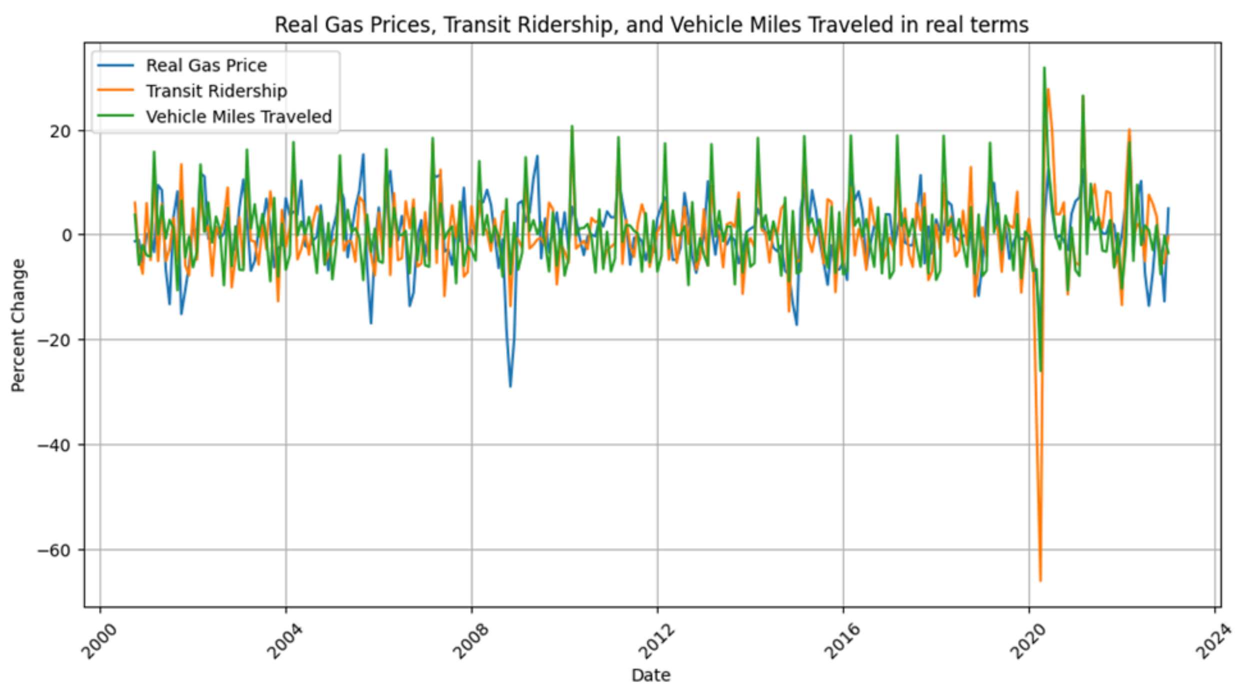
This figure shows that public transit ridership fluctuated between 0.75 and 1 thousand unlinked trips from 2000 to the start of 2020. After the COVID-19 pandemic, public transit ridership dropped to 0.18 thousand unlinked trips. However, it begins to increase again.

Figure 2. Bottom Panel

This figure shows that millions of vehicle miles traveled, fluctuating until the 2020 COVID pandemic when we saw a dramatic dip.

Given that real effects are more important than nominal effects, I examine the relationships between the variables by adjusting for the impact of inflation. Figure 3 shows the relationship between the three variables when gas prices are deflated. The graphs exhibit strong seasonal effects; however, there is no clear evidence of a connection between the three variables.

Figure 3



The relationship between gas prices, vehicle miles traveled, and transit use became more ambiguous as gas prices decreased. For instance, in April 2022, gas prices dropped by 2.9%; however, public transit ridership and vehicle miles traveled increased by 5.7%. Similarly, in September 2022, gas prices fell by 6.8%, and public transit ridership increased by 2%. In December 2022, gas prices fell by 12.7%, whereas public ridership increased by 2.3%. These findings underscore the impact of various factors beyond gas prices in shaping Americans' transportation habits. Daily obligations, such as commuting, visiting families, and shipping goods, often remain relatively stable, regardless of fluctuations in gas prices. Factors such as travel convenience, infrastructure availability, employment, income level, and seasonal opportunities appear to be more influential determinants of transportation choice than the price of gas alone.

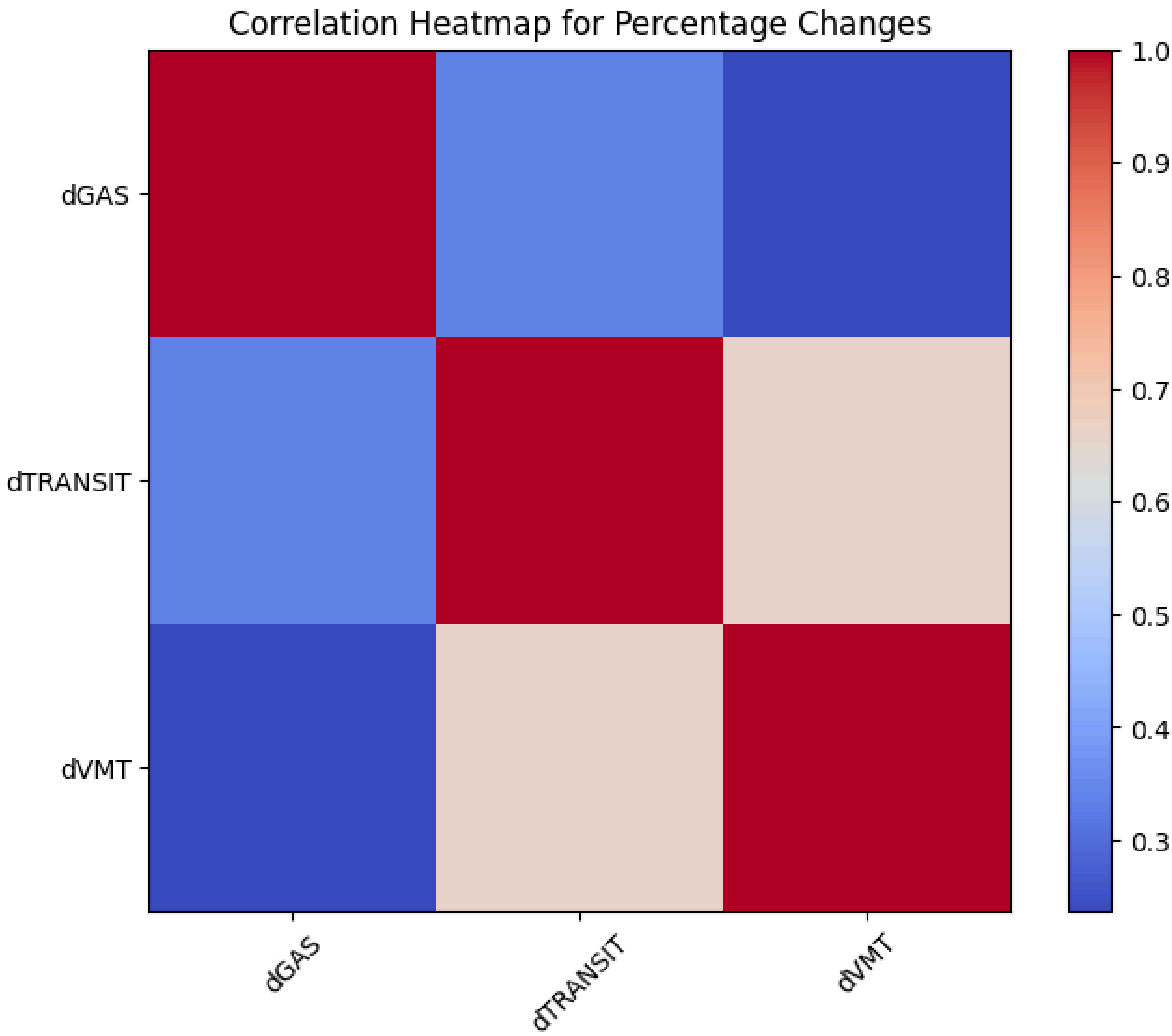
Next, using correlation analysis and visual plots, I looked at the relationships between the variables of interest (vehicle miles traveled, public transit ridership, and gas prices) and the control variables were examined.

Table 2. Correlation Matrix for Selected Variables

GAS	TRANSIT	VMT	CPI	PI
1.00	0.34	0.24	0.77	-0.09
0.34	1.00	0.66	0.25	-0.11
0.24	0.66	1.00	0.03	0.05
0.77	0.25	0.03	1.00	-0.20
-0.09	-0.11	0.05	-0.20	1.00

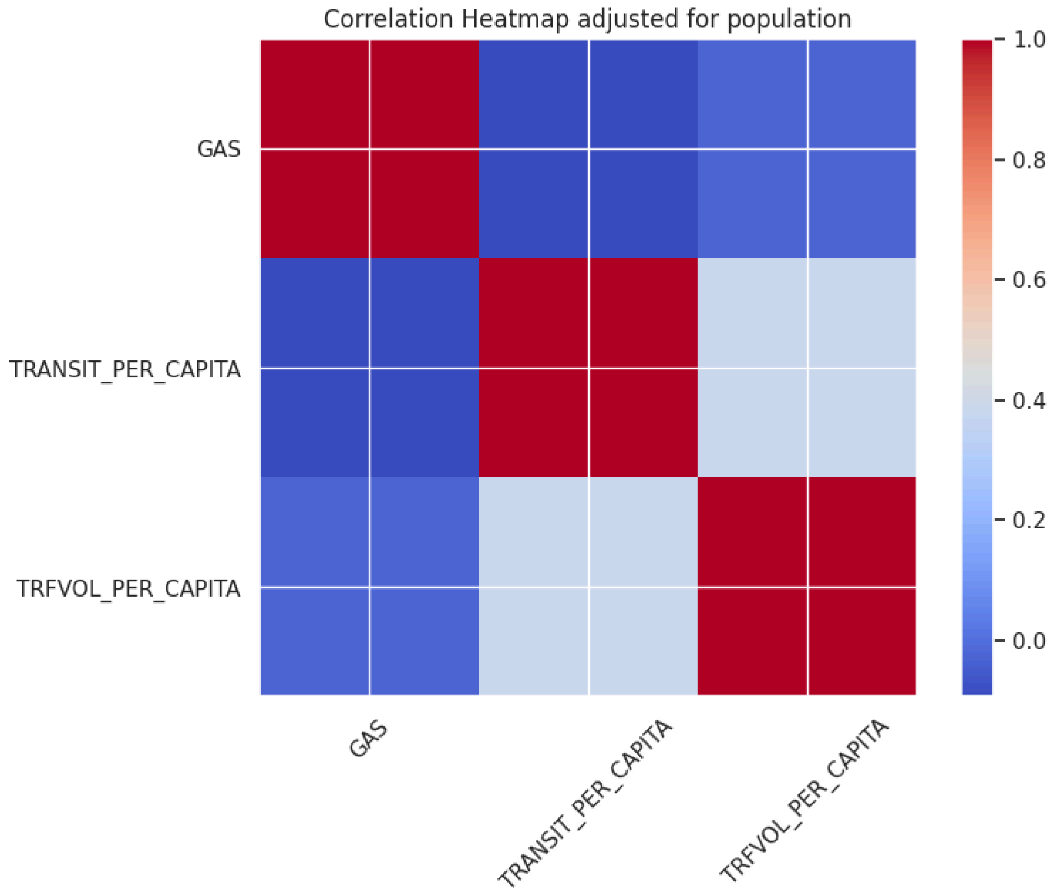
The correlation matrix in Figure 4, which depicts growth rate changes in the variables, reveals specific facts: changes in gas prices positively correlate with growth in vehicle miles traveled (0.32). As gas prices increase, vehicle mile travel also tends to grow. This might seem counterintuitive but could be due to factors such as increased economic activity leading to increased travel and demand for gas, which increases its price. Public transit ridership has a positive but weak correlation with gas prices (0.03), and vehicle miles traveled (0.11), indicating that changes in these variables might have a small impact on public transit ridership. Gas prices have a strong positive correlation with the consumer price index (0.58) and population (0.51), suggesting that the general price level and population increase as gas prices increase. Vehicle miles traveled negatively correlated with the unemployment rate (-0.30), indicating that when more people are employed (and therefore likely commuting to work), more vehicle miles are traveled. Vehicle miles traveled were positively correlated with personal income (0.39), meaning that people tend to travel more as personal income increases.

Figure 4.A



Note: d represents the change in a variable.

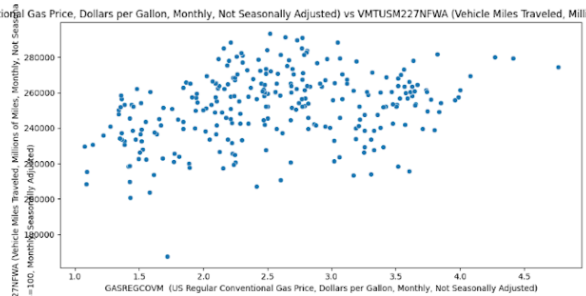
Figure 4.B



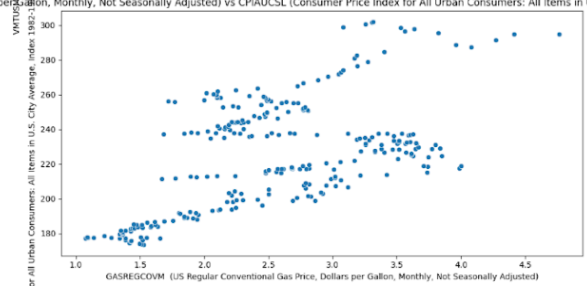
To better visualize these relationships and obtain a better understanding, scatter plots for pairs of variables with significant correlations are shown.

Figure 5 shows scatter plots of the two variables. Figures 5.A, 5.B, 5.C, 5.D and 5.E shows the relationships between gas prices and transit ridership, between gas prices and transit ridership, between real gas prices and transit ridership, and between real gas prices and vehicles traveled.

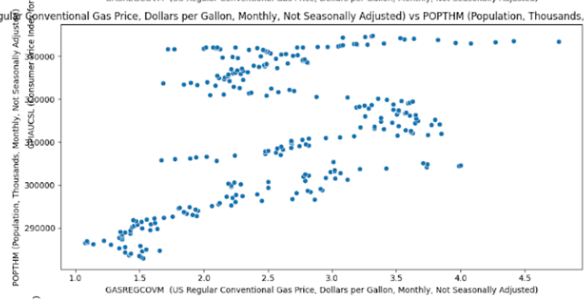
GASREGCOVM (US Regular Conventional Gas Price, Dollars per Gallon, Monthly, Not Seasonally Adjusted) vs VMTUSM227NFWA (Vehicle Miles Traveled, Millions of Miles, Monthly, Not Seasonally Adjusted)



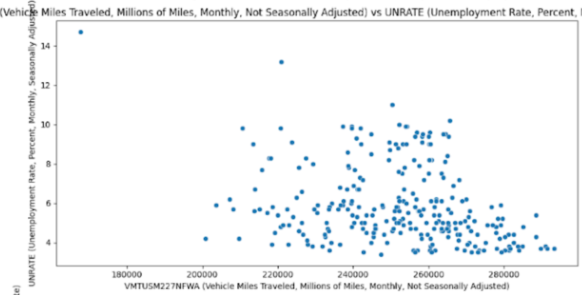
GASREGCOVM (US Regular Conventional Gas Price, Dollars per Gallon, Monthly, Not Seasonally Adjusted) vs CPIAUCSL (Consumer Price Index for All Urban Consumers: All Items in U.S. City Average, Index 1982-1984=100, Monthly, Seasonally Adjusted)



GASREGCOVM (US Regular Conventional Gas Price, Dollars per Gallon, Monthly, Not Seasonally Adjusted) vs POPTHM (Population, Thousands, Monthly, Not Seasonally Adjusted)



VMTUSM227NFWA (Vehicle Miles Traveled, Millions of Miles, Monthly, Not Seasonally Adjusted) vs UNRATE (Unemployment Rate, Percent, Monthly, Seasonally Adjusted)



VMTUSM227NFWA (Vehicle Miles Traveled, Millions of Miles, Monthly, Not Seasonally Adjusted) vs PI (Personal Income, Billions of Dollars, Seasonally Adjusted Annual Rate)

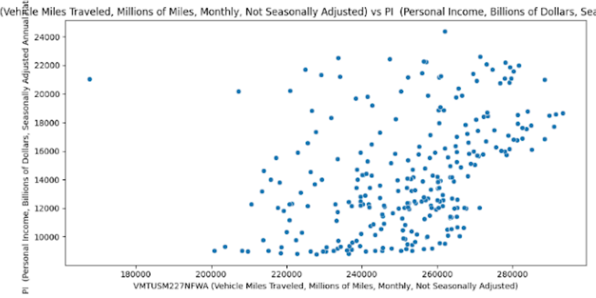
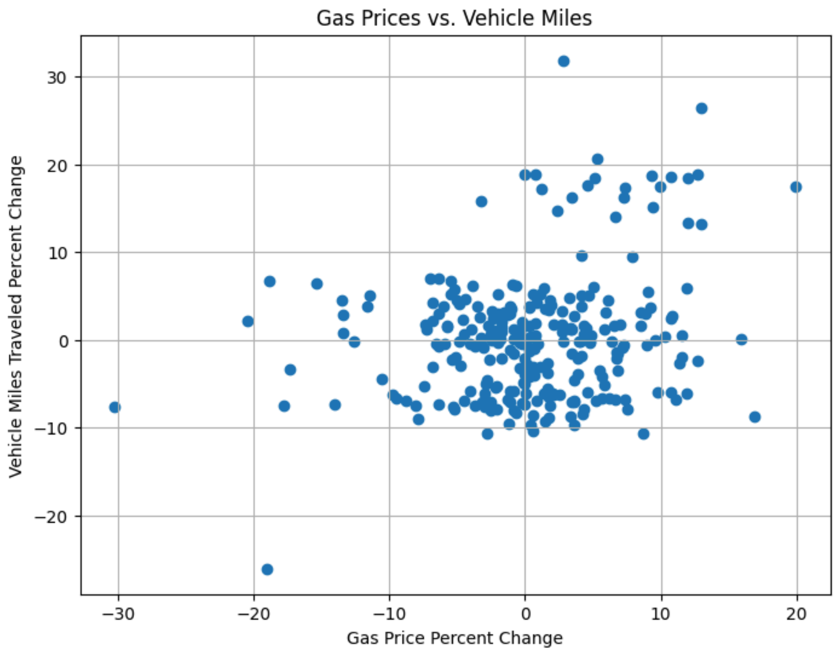
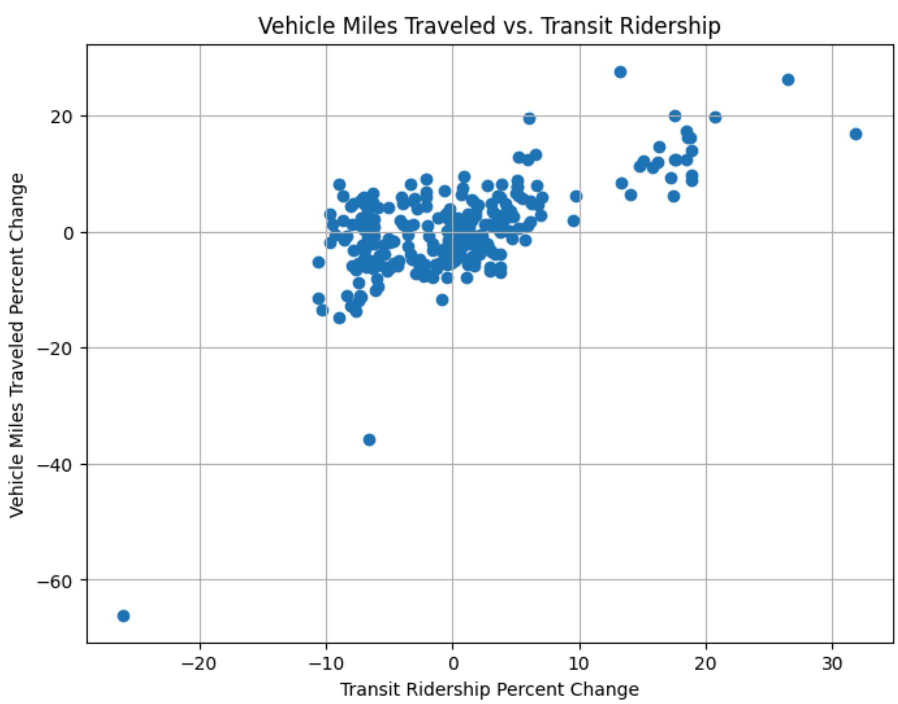


Figure 5.A



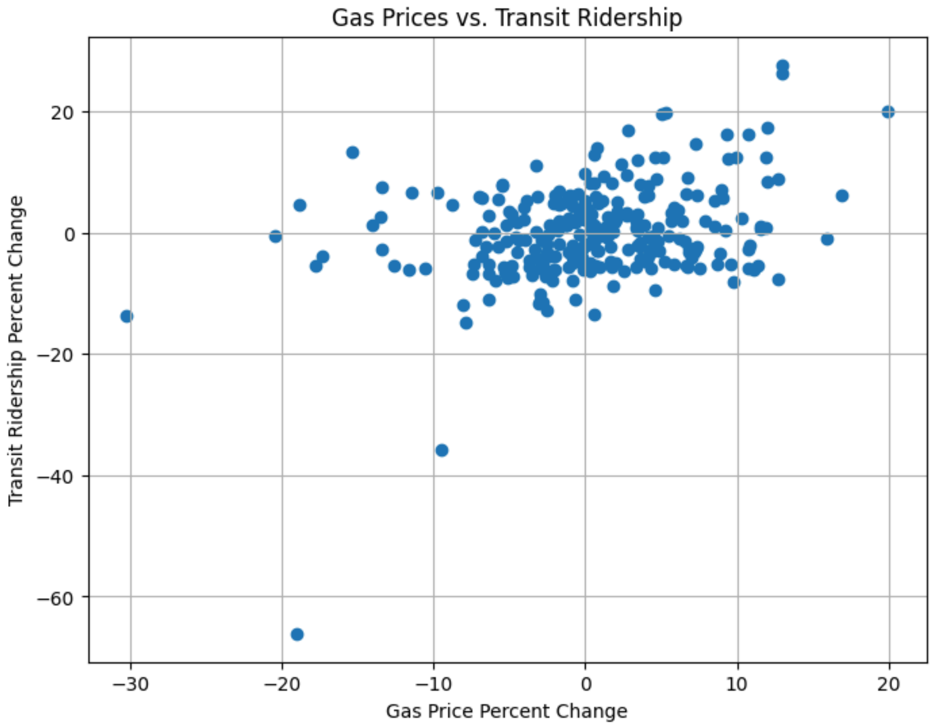
The scatter plot shows a weak positive correlation between gas prices and vehicle miles.

Figure 5.B



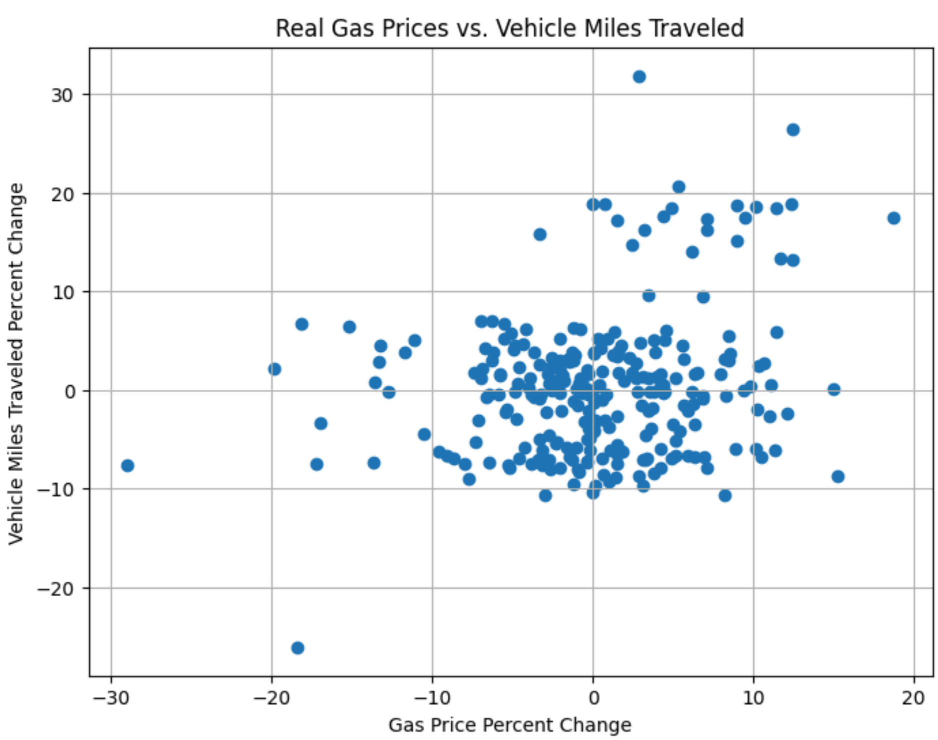
In this figure we can see a moderate positive relationship between vehicle miles traveled percentage change and transit ridership percentage change

Figure 5.C



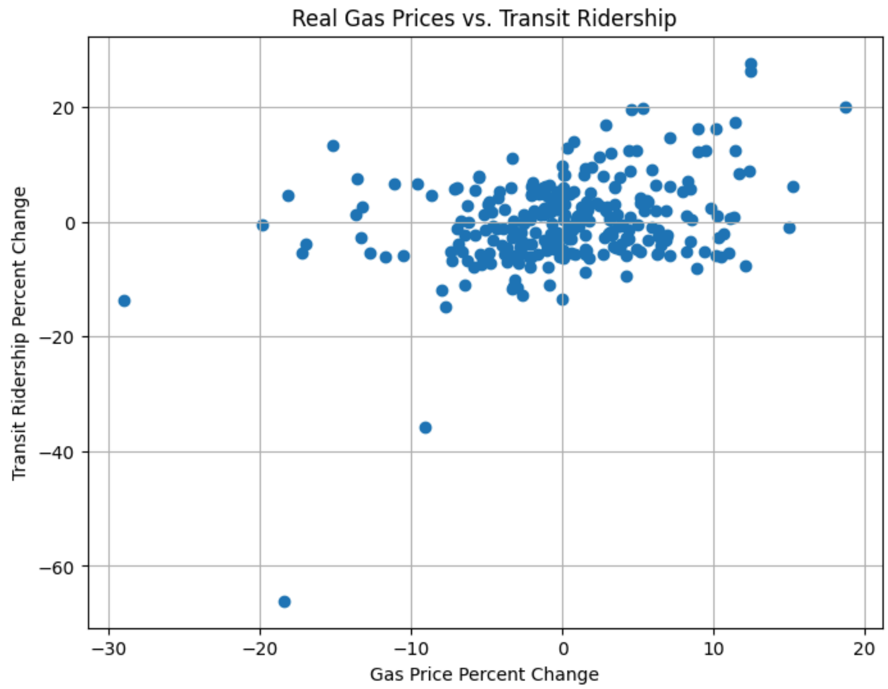
This figure shows that gas prices and transit ridership are not strongly related.

Figure 5.D



This figure shows a low positive correlation between the real gas prices and vehicle miles traveled.

Figure 5.E



Scatter plots confirmed the relationships identified in the correlation matrix.

1. Gas price vs. vehicle miles traveled: This plot shows a positive relationship between gas prices and vehicle miles traveled. As previously mentioned, this might seem counterintuitive. However, this could be due to various factors, such as increased economic activity leading to increased travel and demand for gas, which increases its price.

2. Gas price vs. consumer price index: This plot shows a strong positive relationship between gas prices and the consumer price index. This suggests that when the gas price increases, the general price level in the economy also tends to increase.

3. Gas price versus Population: This plot shows a positive relationship between gas prices and population. As the population increases, the demand for gas may increase, leading to higher gas prices.

4. Vehicle miles traveled versus unemployment rate: This plot shows a negative relationship. When the unemployment rate is high (and thus fewer people commute to work), the number of vehicle miles traveled tends to be lower.

5. Vehicle miles traveled versus personal income: This plot shows a positive relationship. As personal income increases, people might be more likely to travel (for work, leisure, etc.), leading to more vehicle miles traveling.

These insights can help us to understand the relationship between different economic factors. For example, knowledge of the relationship between gas prices, VMT, and public transit ridership can inform transportation, energy, and environmental policies. However, it is important to note that this correlation does not imply causation. The relationships identified were associations, and other factors might be at play that have not been considered. Furthermore, a more rigorous statistical analysis is needed to make definitive conclusions. Hence, we used regression methods.

Using the Ordinary Least Squares (OLS) method to estimate the model parameters, we obtain the following results:

Table 3 Regression Results

Model		1	2	3	4	5	6	7	8
β_0	Constant	0.142 0.420	1.715 0.000	3.326 0.003	3.261 0.037	3.002 0.056	2.093 0.282	-3.842 0.557	-3.528 0.589
B_1	GAS	0.018 0.741	0.323 0.000	0.322 0.000	0.321 0.000	0.306 0.000	0.305 0.000	0.338 0.000	0.334 0.000
B_2	TRANSIT	0.585 0.000	0.583 0.000	0.585 0.000	0.585 0.000	0.591 0.000	0.591 0.000	0.587 0.000	0.571 0.000
B_3	CPI		-8.419 0.000	-8.645 0.000	-8.618 0.000	-8.321 0.000	-8.381 0.000	-9.146 0.000	-9.132 0.000
B_4	POP			-25.16 0.121	-25.17 0.121	-23.252 0.152	-31.225 0.103	-21.113 0.335	-20.141 0.356
B_5	UNRATE				0.010 0.952	0.008 0.964	0.025 0.882	0.063 0.720	0.026 0.884
B_6	PI					0.274 0.101	0.270 0.107	0.266 0.112	0.260 0.119
B_7	RATE						0.219 0.430	0.221 0.426	0.220 0.428
B_8	CUSR0000SETA							0.052 0.342	0.050 0.355
B_9	TOTALSA								0.077 0.113
	Adj. R-squared	0.434	0.485	0.488	0.486	0.49	0.49	0.49	0.49
	Log-Likelihood	-843.	-830.	-828.	-828	-827	-827	-826	-825
	AIC	1692.	1667.	1667	1669	1668	1669	1670	1670
	BIC	1703.	1682.	1685	1690	1693	1698	1703	1706
	Durbin-Watson	2.17	2.28	2.30	2.30	2.29	2.30	2.31	2.31
	F-statistic	104.0	85.60	65.15	51.93	44	37.75	33.13	29.90

Note: p-values are below the coefficient estimates.

Interpretation of Coefficients

Constant (β_0): The model's constant or intercept term is the value of the dependent variable when all independent variables are zero. As model complexity increases (from Model 1 to Model 8), the value of the constant term increases and then decreases. The constant term has no significant practical interpretation in this context, especially considering that setting all the independent variables to zero does not make sense.

Gas Price ($\beta 1$): The coefficient for gas prices starts at 0.018 in Model 1 and increases to 0.323 in Model 2, remaining more or less stable and significant in the remaining models. This positive correlation suggests that, as gas prices rise, VMT also increases, which is somewhat counterintuitive, as we expect people to drive less as gas prices increase.

Public Transit Ridership (β_2): The coefficient of public transit ridership remains significant and positive across all models. This indicates that when more people use public transit, the VMT also increases. This could be due to various reasons, such as increased economic activity, or it could reflect urban growth and development, where both public transit and vehicle usage increase.

CPI (β_3): The coefficient of the Consumer Price Index is negative and significant across all models where it is included. This indicates that, as CPI increases (implying inflation), VMT decreases. This could be due to increased vehicle operation and maintenance costs during inflationary periods.

Population (β_4): The coefficient for population is negative, but not significant in all models. This indicates that the population size does not significantly affect VMT in the models where it is included.

Unemployment Rate (β_5), Personal Income (β_6), Finance Rate (β_7), and Consumer Price Index for Motor Vehicles (β_8): The coefficients of these variables are not significant in any of the models, indicating that they do not have a significant effect on VMT.

Total Vehicle Sales (β_9): The coefficient for total vehicle sales is positive but not significant, suggesting that total vehicle sales do not significantly affect the VMT.

The goodness-of-fit (Adj.. R-squared): The adjusted R-squared values range from 0.434 in Model 1 to 0.49 in Models 5 to 8. The adjusted R-squared value indicates the proportion of the variance in the dependent variable (VMT), which is predictable from the independent variables. An adjusted R-squared of 0.49 in Model 8 means that the model's predictors can explain about 49% of the change in VMT.

Specification Statistics.

Durbin-Watson: The Durbin-Watson statistic tests for autocorrelation in the residuals from a statistical regression analysis. The values in all the models were close to 2, suggesting no autocorrelation in the residuals.

AIC and BIC: The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used for model selection. Lower values for both statistics indicated a better model. Here, AIC and BIC both decrease from Model 1 to Model 2, suggesting that Model 2 is a better fit than Model 1. However, after Model 2, the AIC and BIC values are more or less stable, indicating that adding more variables does not significantly improve the model.

Log-Likelihood: The log-likelihood also measures the goodness-of-fit of a model. Higher values indicated a better fit. Here, the log-likelihood increases from Model 1 to Model 8, suggesting that model fit improves as more variables are included.

F-statistic: The F-statistic is used to test the overall significance of the model. A higher F-statistic indicated a more significant model. However, it seems to decrease as more variables are added, indicating that the additional variables do not contribute significantly to the model's explanatory power.

In conclusion, the models suggest that gas prices, public transit ridership, and CPI are significant predictors of VMT, whereas the other variables are not significant. The goodness-of-fit and specification statistics suggest that Model 2 provides a reasonable fit to the data and that adding more variables does not significantly improve the model. However, the counterintuitive positive relationship between gas prices and VMT suggests further investigation. This counterintuitive result could be due to various reasons, such as omitted variable bias (where other important variables that affect VMT are not included in the model) or complex underlying relationships between gas prices, VMT, and other factors. For example, it could be that when gas prices increase, it is during economic growth, and people travel more despite the higher costs. It is also possible that the data capture longer-term effects such as people buying more fuel-efficient cars when gas prices are high, which could lead to more driving. Thus, we considered methods other than linear regression.

The results below are the feature importance scores from the three machine learning models: Decision Tree, Random Forest, and Gradient Boosting. Feature importance scores represent the usefulness or value of each feature in constructing the machine-learning model. The higher the score, the more influential the feature in making predictions.

Table 4 Feature Importance

Feature Importance	Decision Tree	Random Forest	Gradient Boosting
GAS	0.0431	0.0457	0.0446
TRANSIT	0.6156	0.6158	0.6392
CPI	0.03784	0.0391	0.0410
POP	0.1143	0.0908	0.1362
UNRATE	0.0156	0.0404	0.0251
PI	0.0413	0.0464	0.0118
RATE	0.0424	0.0317	0.0268
CUSR0000SETA	0.0166	0.0303	0.0140
TOTALSA	0.0735	0.0598	0.0613
R^2	-0.0103	0.4077	0.4023
Standard Deviation	0.3138	0.0790	0.1125

Decision Tree Feature Importance

The essential feature of the decision tree model is 'TRANSIT' (change in Public Transit Ridership), with a score of 0.6155. Followed by 'POP' (change in Population) with a score of 0.1143. The least essential feature in this model is 'UNRATE' (Unemployment Rate), with a score of 0.0155.

Random forest feature importance.

Similar to the decision tree model, the most crucial feature in the random forest model is 'TRANSIT' with a score of 0.6158. This was followed by 'POP' with a score of 0.0908. The least essential feature in this model is 'CUSR0000SETA' the Consumer Price Index for All Urban Consumers (‘Used Motor Vehicles), with a score of 0.0303.

Importance of gradient boosting features

Again, the most important feature is 'TRANSIT,’ with a score of 0.6392. This was followed by 'POP' with a score of 0.1361. The least essential feature in this model is 'P.I.' (change in Personal Income), with a score of 0.0118.

In all three models, 'TRANSIT' is the most essential feature, indicating that changes in public transit ridership are vital in predicting the dependent variable (presumably VMT, based on your previous messages). 'POP' also consistently ranks high in importance.

Interestingly, 'GAS' (change in Gas Price), a significant factor in the regression models, does not have very high feature importance in these machine learning models. This suggests that while 'GAS' may have a statistically significant relationship with VMT, it might not be as important for predicting the interaction and nonlinear relationships between all variables.

These results provide a different perspective on the data and highlight the complexity of the relationships between these variables. It is also essential to note that correlations between features can influence feature importance scores, and they do not provide information about the direction of the relationship (positive or negative) between each feature and the target variable.

Section 4 Seasonal adjustment

The above results were obtained using unadjusted seasonal data for the most part. One might inquire what the results would look like using seasonally adjusted data. This exercise is beyond the scope of this study, but an attempt is made to examine the seasonal nature of the data for VMT, to which I now turn.

The data provided are seasonally unadjusted. Unadjusted data from the FRED were analyzed. This is possible using a technique involving time-series decomposition. A time-series decomposition technique separates a time-series variable into its underlying components: trend, seasonal, and residual. Several statistical methods can be used to produce decompositions. These techniques are widely used in forecasting and time series analyses. Some primary methods include classical decomposition, moving averages, X-12-ARIMA, SEATS, and Hodrick-Prescott filter. See, for example, Granger (1989), Bowerman and O'Connell (1993), and Enders (1995).

The classical decomposition method separates a time series into trends, seasonality, and residual components. It assumes that seasonality and trend components are additive. This method can be applied when data have no irregular or random components.

Moving Averages: Moving averages involve computing the average of a fixed number of consecutive observations in a time series. The moving average can help identify the underlying trend by smoothing short-term fluctuations. Various moving averages can be used, such as simple moving averages, weighted moving averages, and exponential moving averages.

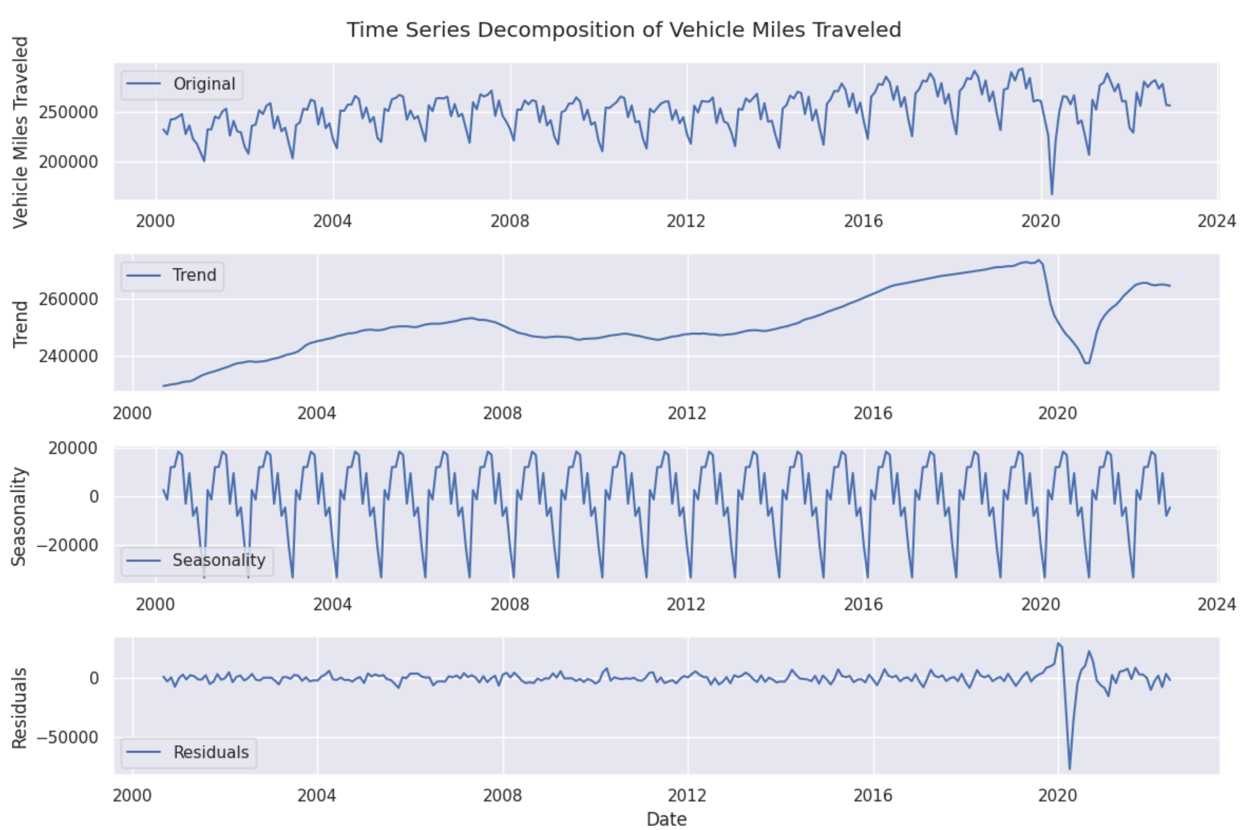
X-12-ARIMA and SEATS: X-12-ARIMA and SEATS are seasonal adjustment programs developed by the U.S. Census Bureau and Bank of Spain These programs use advanced statistical methods, including autoregressive integrated moving average (ARIMA) models, to decompose time series into trend, seasonality, and residual components, while accounting for irregular components and calendar effects.

Hodrick-Prescott Filter: The Hodrick-Prescott (H.P.) filter is widely used for separating a time series into trend and cyclical components. It applies a smoothing parameter to penalize the deviations between the observed data and the estimated trend. The resulting decomposition provided an estimate of the underlying trend component.

These techniques offer various advantages and are suitable for different types of time-series data. The choice of an appropriate decomposition method depends on the characteristics of the time series and the specific objectives of the analysis. This study uses the X-12-ARIMA seasonal adjustment program in the U.S. Census Bureau. Findley (1998).

The figure below shows the time-series decomposition of the vehicle miles traveled.

Figure 6



Time series decomposition is a statistical technique that breaks down time-series data into their constituent components, typically comprising trend, seasonality, and noise. The goal of decomposition is to understand the underlying patterns and structures within time-series data. After looking at the time series decomposition of vehicle miles traveled, we can see that the COVID-19 Pandemic had a strong negative effect on vehicle miles traveled. This is why the trends and residuals fell around March 2020.

This figure shows the vehicle miles traveled in terms of trends, seasonality, and residuals. The trend shows that vehicle miles traveled increased until approximately 2019, when there was a dip until 2021. The residuals of vehicle miles traveled also fell significantly owing to the 2020 COVID pandemic. The seasonal pattern shows peaks in June and dips during the winter months. It dipped later and rose again in August.

Section 6 Conclusion

This study investigated the relationship between gas prices, public transit ridership, and vehicle miles traveled in the United States over 23 years. This study explored the complex dynamics of transportation habits and the impact of changing gas prices using correlation analysis, regression analysis, decision trees, random forest, and gradient boosting techniques.

Contrary to the initial expectations, the findings reveal a nuanced relationship between gas prices and transportation choices. Increases in gas prices coincided with increases in public transit ridership and vehicle miles traveled. This unexpected correlation indicates factors beyond gas prices, such as convenience of employment, income levels, and seasonal opportunities. Seasonal factors were particularly notable, with spikes in transit ridership and vehicle miles traveled during the summer.

The analysis also employed machine learning techniques, including decision trees, random forests, and gradient boosting, to capture nonlinear relationships and complex interactions between variables. These techniques yield accurate predictions and provide insight into the factors that influence transportation habits.

Overall, the findings of this study contribute to a deeper understanding of the multifaceted relationship between gas price and transportation choice. They underscored the importance of considering a range of factors beyond gas prices when developing policies and strategies to promote sustainable and efficient transportation systems. Policymakers, urban planners, and transportation experts can leverage these insights to design effective interventions that account for the various influences on transportation decisions.

While this study sheds light on the intricate dynamics of play, further investigation is warranted to explore additional factors that may impact transportation choices. Future studies should consider variables such as environmental consciousness, technological advancements, and policy interventions to gain a more comprehensive understanding of the evolving transportation landscape.

In conclusion, this study provides valuable insights into the relationship between gas prices, public transit ridership, and vehicle miles traveled in the United States. The complexity of this relationship highlights the need for a holistic approach when analyzing transportation habits and developing strategies to shape sustainable transportation systems in the future.

Notes:

1. Undergraduate Student (BS), Statistics/Economics, Rutgers University, New Brunswick

Acknowledgments

I'd like to thank Professor T. Agbeyegbe for his guidance and suggestions. I have benefitted from conversations with Microsoft (2023).

References

Ahangari, et al. (2014). Investigation into Impact of Fluctuations in Gasoline Prices and Macroeconomic Conditions on Road Safety in Developed Countries.

Bowerman, B. L., & O'Connell, R. T. (1993). Forecasting and time series: an applied approach

Brownstone, D., & Golob, T. F. (2009). The impact of residential density on vehicle usage and energy consumption. Journal of Urban Economics, 65(1), 91-98.

Congressional Budget Office. (2014). The 2014 long-term budget outlook. Washington, DC: Congressional Budget Office.

Dillon, H. S., Saphores, J. D., & Boarnet, M. G. (2015). The impact of urban form and gasoline prices on vehicle usage: Evidence from the 2009 National Household Travel Survey. Research in Transportation Economics, 52, 23-33.

Enders, Walter. (1995). Applied Econometric Time Series. New York: John Wiley and Sons, Inc.

Findley, D. F., B. C. Monsell, W. R. Bell, M. C. Otto, and B.-C. Chen. "New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program." Journal of Business &

Economic Statistics. Vol. 16, Number 2, 1998, pp. 127–152.

Federal Research Bank of St. Louis Databank. (2023)

Granger, C. W. J. (2014). Forecasting in business and economics. Academic Press.

Hymel, K. M. (2014). Factors Influencing Vehicle Miles Traveled in California: Measurement and Analysis.

Hyrc, M. (2017). Gas Prices and Transportation Habit https://fredblog.stlouisfed.org/2017/06/gas-prices-and-transportation-habits/ Accessed May 10, 2023.

James, et al. (2021) An Introduction to Statistical Learning, with Applications in R (Springer, second edition, 2013/2021)

Jung, H., Yu, G. J., & Kwon, K. M. (2016). Investigating the effect of gasoline prices on transit Ridership and unobserved heterogeneity. Journal of Public Transportation, 19(4), 56-74.

Lane, B. W. (2012). A time-series analysis of gasoline prices and public transportation in US Metropolitan areas. Journal of Transport Geography, 22, 221-235.

Leard, B., Linn, J., & Munnings, C. (2016). Explaining the Evolution of Passenger Vehicle Miles Traveled.

Maley, D. W., & Weinberger, R. (2009). Rising gas price and transit ridership: case study of Philadelphia, Pennsylvania. Transportation research record, 2139(1), 183-188.

Munyon, V. V., & Bowen, W. M. (2018). Vehicle fuel economy and vehicle miles traveled: An Empirical investigation of Jevon's Paradox.

Nowak, W. P., & Savage, I. (2013). The cross elasticity between gasoline prices and transit use: Evidence from Chicago.