Working Paper: Examining The Effects Of COVID-19 Stimulus Checks on Air Travel

Lucas Mantoani

2024-03-20

Preview

This project serves as the Capstone for Harvard's Professional Certificate in Data Science. Project requirements entail sourcing data, conducting two regressions (with one regression being non-linear or non-logistic), and compiling a report summarizing the findings. The objective is to attain a clear understanding of the subjects covered across the nine sections of the Data Science Series.

Introduction

In March of 2020, the COVID-19 infectious disease was classified as a national pandemic in the United States. Americans were advised to quarantine, and only essential workers were permitted to continue going to work. As part of the American Government's response to avoid an economic crisis, three sets of stimulus checks were provided. Labeled the CARES Act, Consolidated Appropriations Act, and the American Rescue Plan Act respectively, these checks were meant to supplement household income and prevent an economic recession. The CARES Act provided $1,200 per income tax filer and $500 per child. The second stimulus, the Consolidated Appropriations Act, provided $600 per income tax filer and $600 per child. Finally, the American Rescue Plan Act gave $1,400 per income tax filer and $1,400 per child.

Although the CARES Act received bipartisan support, subsequent rounds of stimulus faced challenges in a Republican-led Senate. Proponents argued that sizable stimulus checks could shorten the recession and alleviate economic hardship. Opponents contended that it would be costly and wasteful. A common critique was that the fiscal support disproportionately benefited economically stable households (DeParle, 2021).

This report seeks to investigate whether the stimulus checks influenced air travel patterns. If families did not allocate the stimulus towards essential expenses such as rent, food, or transportation, they might have saved the funds. After prolonged periods of staying at home due to quarantining restrictions, households may have developed a desire to venture beyond their residences. The objective of this report is to assess whether there was any discernible impact on air travel resulting from changes in disposable income.

Method's and Analysis

Data Integration

The following data was sourced from the Federal Reserve Bank of St. Louis(FRED), the Bureau of Transportation Statistics(BTS), and the Bureau of Labor Statistics(BLS).

The table below describes each variable in the data set.

Table 1: Variable Desciptions

Variable	Description	Units
International_RPM	International Revenue Passenger Miles(RPM)	Millions of RPM
Domestic_RPM	Domestic Revenue Passenger Miles(RPM)	Millions of RPM
System_RPM	Total Revenue Passenger Miles(RPM)	Millions of RPM
M1	M1 Money Supply	Billions of Dollars, Seasonally Adjusted
M2	M2 Money Supply	Billions of Dollars, Not Seasonally Adjusted
Real_Disposable_Income	Real Disposable Personal Income	Trillions of Chained 2017 Dollars, Seasonally Adjusted Annual Rate
CPI_Stable_Baskets	Consumer Price Index for All Urban Consumers: All Items Less Food and Energy in U.S. City Average	Index 1982-1984=100, Seasonally Adjusted
Unemployment_Rate	Unemployment Rate	Percent, Seasonally Adjusted
CPI_Airline_Fares	Airline fares in U.S. city average, all urban consumers, not seasonally adjusted	Index 1982-84=100, Not Seasonally Adjusted
GDP	GDP	Billions of Chained 2017 Dollars, Seasonally Adjusted Annual Rat
Rent_Prices	Rent Prices	Index 1982-1984=100, Seasonally Adjusted
Food_Price_Index	Food Prices	Index Dec 1984=100, Not Seasonally Adjusted
Ave_Gas_Prices	Average Gasoline Price	Dollars per Gallon, Not Seasonally Adjusted

The first few rows are displayed below to demonstrate variable storage.

Table 2: Data Storage

Period	International_RPM	Domestic_RPM	System_RPM	M2	M1	Real_Disposable_Income	CPI_Stable_Baskets	Unemployment_Rate	CPI_Airline_Fares	GDP	Rent_Prices	Food_Price_Index	Ave_Gas_Prices
2000-01-01	14111488	34933923	49045412	4666.2	1122.1	9799.9	2.335519	4.0	221.8	13812.35	180.9	126.7	1.28880
2000-02-01	13282332	36023970	49306303	4679.4	1108.6	9837.9	2.379221	4.1	230.2	13871.53	181.3	127.2	1.37700
2000-03-01	16530620	43912920	60443541	4710.2	1107.5	9864.0	2.483286	4.0	240.7	13951.08	181.9	127.4	1.51625
2000-04-01	16322844	41963837	58286680	4766.1	1115.6	9913.7	2.409239	3.8	239.5	14133.23	182.3	128.1	1.46475
2000-05-01	17355458	43178325	60533783	4753.9	1104.9	9954.5	2.558647	4.0	241.3	14113.45	182.8	129.3	1.48680
2000-06-01	19018771	45884524	64903295	4771.8	1102.6	9982.3	2.701492	4.0	245.0	14146.44	183.4	129.4	1.63325

Below are the summary statistics:

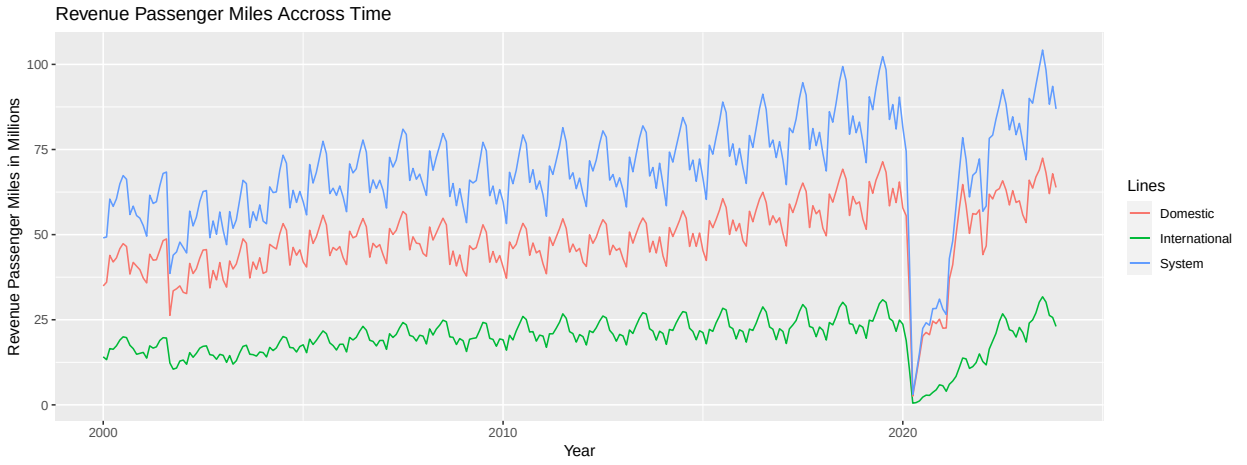
Table 3: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
International_RPM	287	19,544,207.000	5,579,789.000	503,559	31,749,731
Domestic_RPM	287	48,714,600.000	10,414,362.000	2,583,242	72,471,792
System_RPM	287	68,258,807.000	15,404,396.000	3,086,801	104,221,524
M2	287	10,916.930	5,019.509	4,666.200	21,703.500
M1	287	4,666.804	6,123.774	1,088.600	20,664.500
Real_Disposable_Income	287	13,219.590	2,109.746	9,799.900	20,422.600
CPI_Stable_Baskets	287	2.587	1.062	0.664	6.538
Unemployment_Rate	287	5.771	1.971	3.400	14.800
CPI_Airline_Fares	287	263.334	32.362	197.134	344.853
GDP	287	17,777.830	2,408.994	13,812.350	22,677.430
Rent_Prices	287	269.118	57.759	180.900	408.366
Food_Price_Index	287	181.885	35.079	126.700	256.595
Ave_Gas_Prices	287	2.608	0.789	1.086	4.929

Data Interpretation

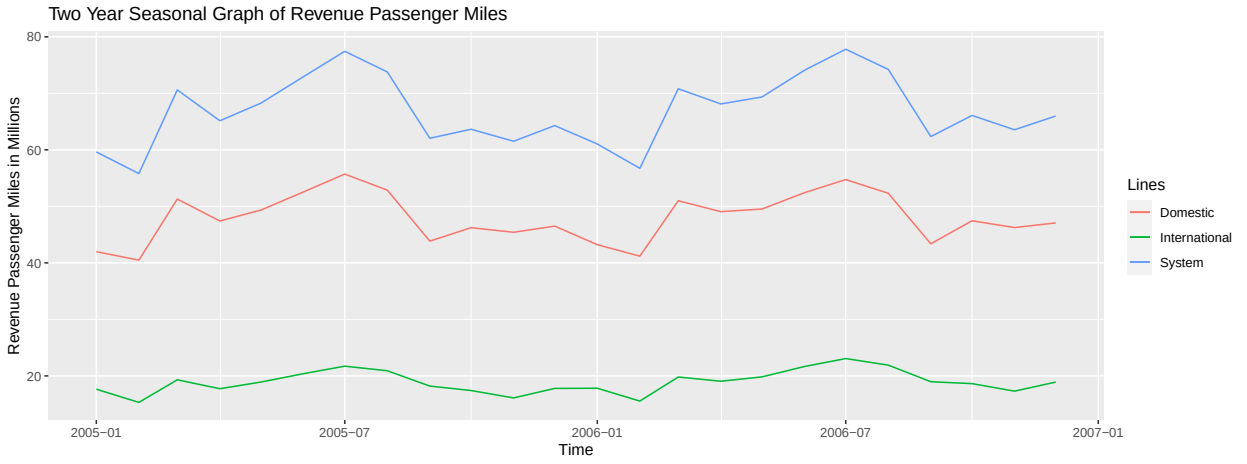
In this section visualization of the data will give insight about the variables changed overtime through graphical visualization.

The dependent variable Revenue Passenger Miles(RPM) is displayed below showing its fluctuations across time.



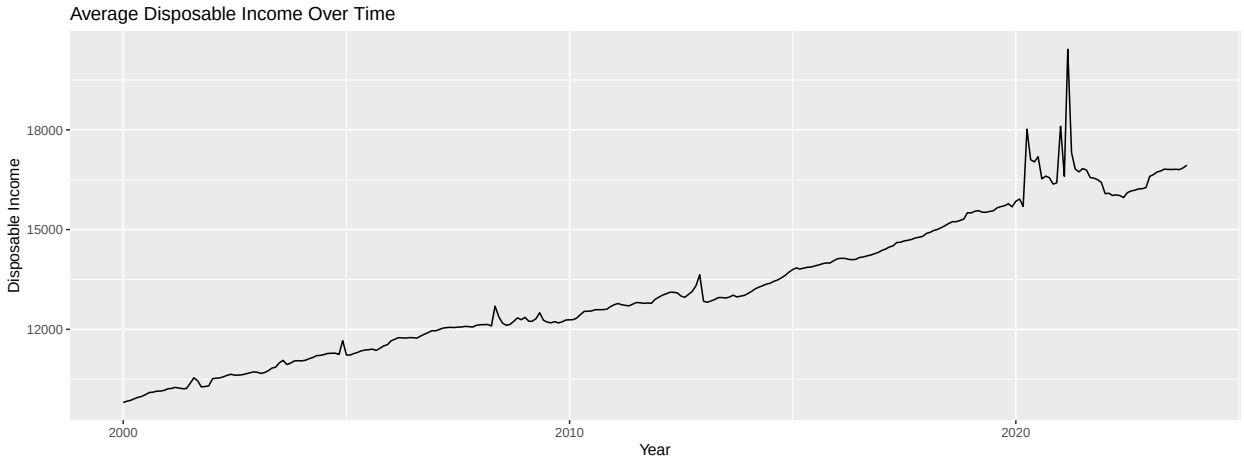
RPM over time highlights a couple of important characteristics. The first insight displayed is the shock effects of demand across the time period. Notably there is a small dip in 2001 from the terrorist attack on 9-11-2001 and there is a major dip from the COVID-19 pandemic. The project analysis will have consider these factors into our regressions to not skew our results.

The second insight that RPMs have seasonal fluctuations that seem consistent across the time period. A clearer depiction of the variations between high and low RPMs becomes apparent when condensing down to two seasons.

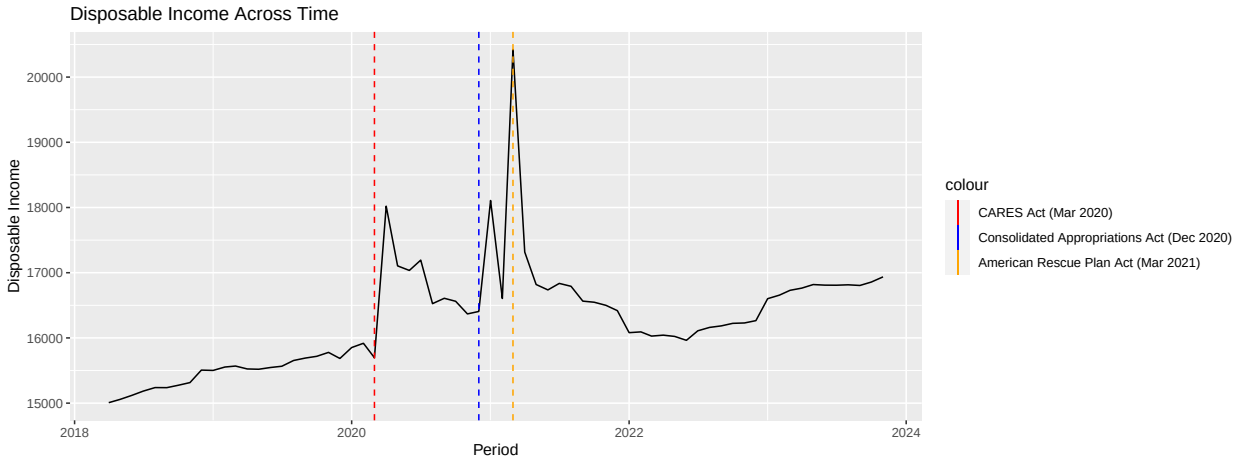


Summer is the peak of travel in a given year with dips when transitioning from fall to winter. International and domestic travel follow similar peaks and valleys.

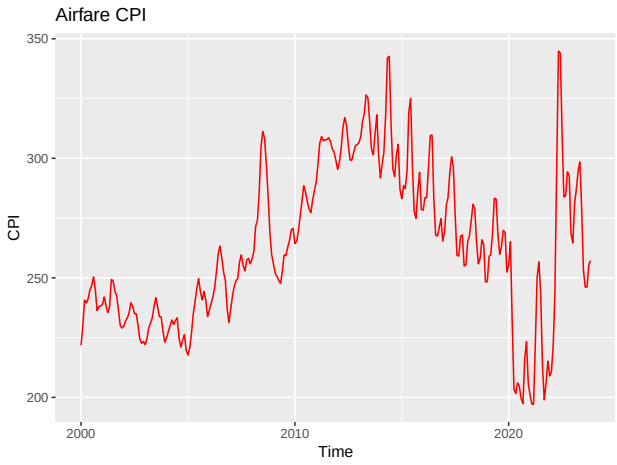
Real Disposable Income is graphed below:

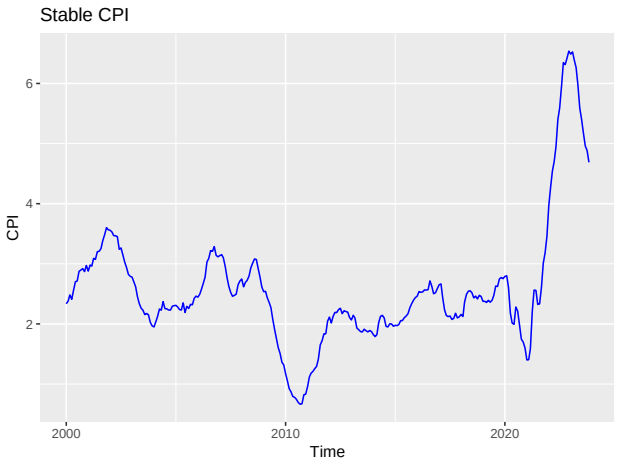


The graph is linear with small fluctuations before the year 2020. The huge spikes observable are the stimulus checks that were provided to prevent the economy from collapsing from the global pandemic.These can be observed more closely in the graph below:

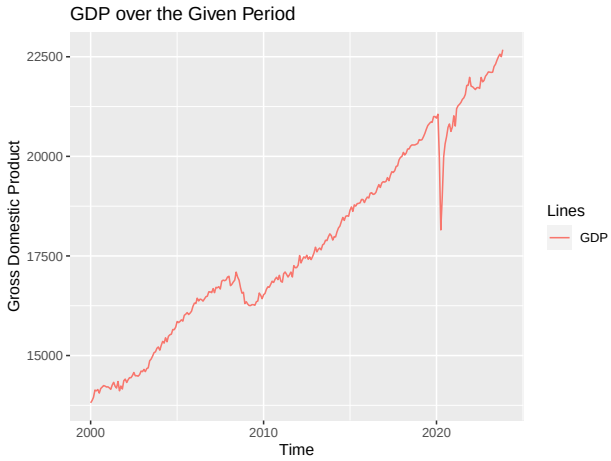


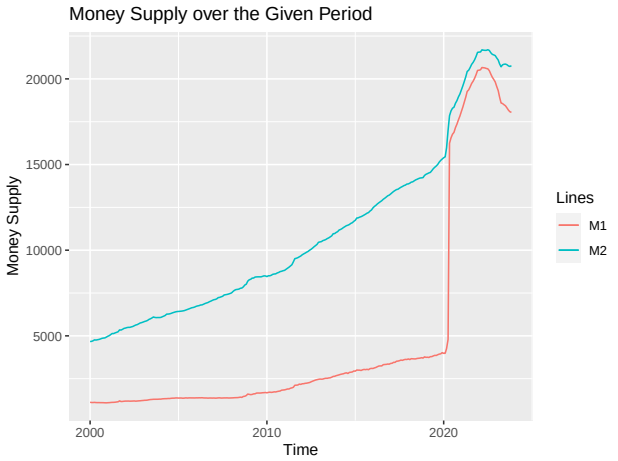
The next graphs compares the CPI of flight prices compared to stable goods:



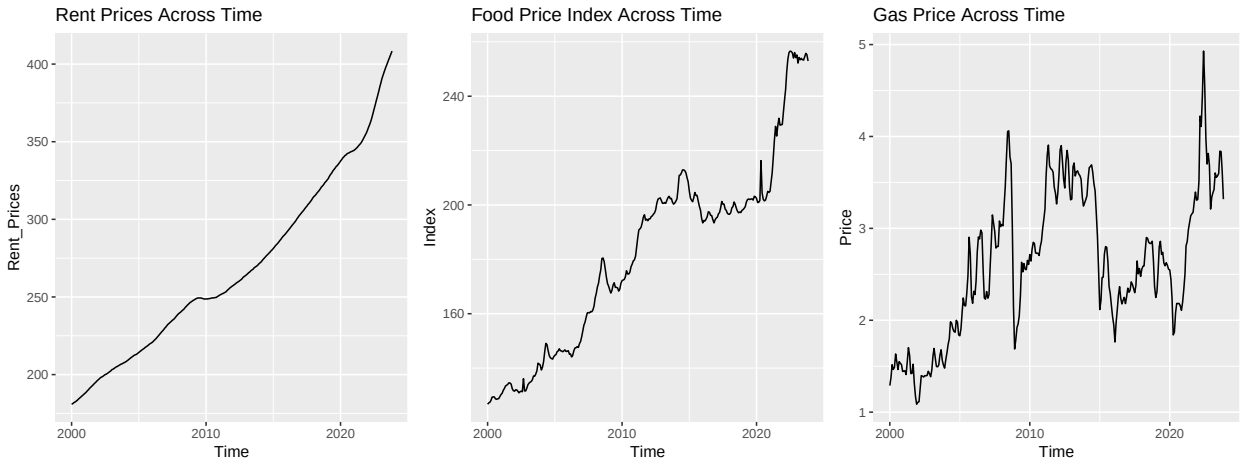


The graphs do not display a consistent relationship over the entire time period. The only period showing correlation is during the pandemic. This behavior can be clarified by analyzing macroeconomic variables. Initially, a decrease in demand led to a dip in prices as US citizens refrained from traveling and remained indoors. The significant spike in prices can be attributed to the surge in the money supply, as depicted below. Gross Domestic Product (GDP) declined as production slowed due to quarantine measures. To sustain economic activity, the government resorted to printing money, distributing it to businesses and citizens to stimulate demand. This led to inflation, explaining the spike in prices, while also facilitating an upward trend in GDP. Both GDP and the Money Supply are depicted in the graphs below.





The next series of graphs are provided to measure substitute goods of travel. If the prices of your base necessities increase, there is less wealth to spend on vacations and visiting families during holidays.



Gas prices are the most comparable substitute good, as they also constitute a travel expense. When planning a trip, one can choose between driving or flying to their destination. The other two graphs depict a consistent upward trend over time, indicating increasing prices.

Modeling

To determine if the personal stimulus check had an impact on RPMs this project will find if there are statistically significant effects. The first way to tell this is by running a regression model. The equation for this model is given below:

$$Y_i = \beta_0 + \beta_1 + ... + \beta_n + \epsilon$$

where: - Y_i is the dependent variable, - β_0 is the intercept , - β_1 is the first independent variable, - β_n is the nth independent variable, - ϵ is the error term.

There are several considerations and assumptions for this model. The first aspect is the seasonal nature of RPMs. The final model will hold the seasons constant in order to measure if real disposable income effected RPMs. If the variables show no statistical significance they will be included in the final model to account for the cyclical nature seen visualized in the data. The second aspect of consideration is removing the demand shock data from our model. This project aims to see if there was an increase in RPM after the stimulus checks. Including economic shocks it will skew our equation negative because the huge dip seen in the data. Finally the models will not include the macroeconomic variables to prevent confounding results.

Regression Model 1

The first model will only include the outcome variable and the primary input variable. The equation for the base model is given below:

$$Y = \beta_0 + \beta_1 * x_1 + \epsilon$$

where: - Y_i is the predicted outcome Revenue Passenger Miles, - β_0 is the intercept , - β_1 is the coefficient of the independent variable Real Disposable Income, - x_1 is the value of the predictor variable Real Disposable income, - ϵ is the error term.

[1] “Model 1 Coefficients” (Intercept) Real_Disposable_Income 4.430536 5.129234

Regression Model 2

The second equation will encompass additional variables such as flight prices, rent costs, food costs, and gasoline expenses. Flight prices were incorporated due to the assumption that recreational travel is an elastic good. Elastic goods are those for which consumers are less likely to purchase if the price increases because they are not essential. The following two variables were included because they represent inelastic goods. Inelastic goods are essential items that consumers will continue to purchase regardless of price increases. Rent is included for its role in determining the cost of accommodation in an area. Lastly, gasoline prices are included as they serve as a substitute for flight costs. If driving becomes cheaper than flying, consumers are more likely to opt for driving, considering the opportunity cost. The equation is outlined below: drive then fly consumers will more often drive if it is worth the opportunity cost. This equation is given below:

$$Y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + ... + \beta_5 * x_5 + \epsilon$$

where: - β_2 through β_5 are the coefficient of the variables held constant, x_2 through x_5 are the value of the variables held constant

[1] "Model 2 Coefficients" (Intercept) Real_Disposable_Income CPI_Airline_Fares -5.9856037 3.6885598 0.1233849 Rent_Prices Food_Price_Index Ave_Gas_Prices 0.2165173 -0.3996787 4.0801191

Regression Model 3

The final regression is a time series. This will include the recorded date of the data and the seasonal variables stored as dummy variable 0 or 1. There are only three included seasons because the forth season is the omission of the other 3.

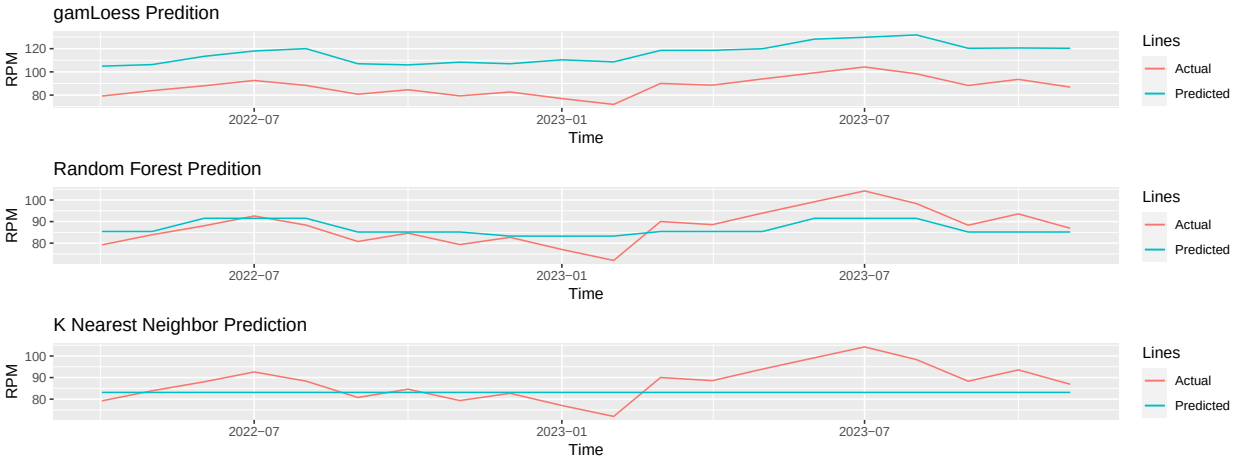
$$Y = \beta_0 + \beta_1 * x_1 + ... + \beta_6 * (t) + \beta_7 * x_7 + \beta_8 * x_8 + \beta_9 * x_9 + \epsilon$$

where: - β_6 is the coefficient for time,-t is the value of time, -β_7 through β_9 are the seasonal coefficients, -x_7 through x_9 are dummy variable values for the season

[1] "Model 3 Coefficients" (Intercept) Real_Disposable_Income CPI_Airline_Fares -4.504817e+00 6.595039e+00 -2.262121e-03 Rent_Prices Food_Price_Index Ave_Gas_Prices -1.661160e-02 -1.507072e-01 2.234299e+00 Period SeasonSpring SeasonSummer 9.748509e-09 4.144734e+00 1.273533e+01 SeasonWinter -3.167937e+00

Machine Learning Predictions

This project will employ machine learning algorithms to assess the impact on air travel. This project splits the data into a training and testing set with dates preceding 2020 and compares predicted values against actual values to evaluate the influence of COVID stimulus checks. Three distinct models - GAM (Generalized Additive Models), Random Forest, and K-Nearest Neighbors - will be utilized to discern any substantial effects stemming from the stimulus checks. If testing predictions prove accurate while validation predictions significantly underestimate actual results, it suggests that the stimulus checks contributed to an increase in RPMs. Below are three graphs of the validation data and the prediction data.


gamLoess Predition
RPM
120
100
80
2022−07
2023−01
2023−07
Time
Lines
Actual
Predicted
Random Forest Predition
RPM
100
90
80
2022−07
2023−01
2023−07
Time
Lines
Actual
Predicted
K Nearest Neighbor Prediction
RPM
100
90
80
2022−07
2023−01
2023−07
Time
Lines
Actual
Predicted

Results

The results of the three regression models discussed above are displayed in the table below.

Table 4: Regression Results

	Dependent variable:		
	System_RPM		
	Base	Full Regression	Time Series
	(1)	(2)	(3)
Real_Disposable_Income	5.129***	3.689*	6.595***
	(0.262)	(1.918)	(2.151)
CPI_Airline_Fares		0.123***	−0.002
		(0.037)	(0.028)
Rent_Prices		0.217**	−0.017
		(0.087)	(0.058)
Food_Price_Index		−0.400***	−0.151**
		(0.083)	(0.059)
Ave_Gas_Prices		4.080***	2.234***
		(1.081)	(0.750)
Period			0.000
			(0.00000)
SeasonSpring			4.145***
			(0.858)
SeasonSummer			12.735***
			(0.859)
SeasonWinter			−3.168***
			(0.856)
Constant	4.431	−5.986	−4.505
	(3.420)	(8.092)	(6.830)
Observations	261	261	261
R^2	0.596	0.643	0.854
Adjusted R^2	0.595	0.636	0.849
Note:			*p<0.1; **p<0.05; ***p<0.01

Real disposable income showed significance in the first and third regressions, with a positive relationship observed. For every unit increase in real disposable income, there was an associated increase of 6.595 million revenue passenger miles. However, the significance of variables in the second and third regressions changed with the inclusion of seasonal variables. In the third regression, seasonal variables, food prices, and gasoline prices were found to be significant. Gas prices exhibited a positive coefficient, indicating that an increase in gas prices led to higher airline travel. Conversely, food prices had a negative coefficient, suggesting that

lower food costs corresponded to increased travel. The data included in the third regression explained 85% of the resulting change in revenue passenger miles.

Regarding the machine learning algorithms, the mean average error (MAE) for our testing data was within one standard deviation of the actual result. Upon examining the graphs and table, the results did not align with the hypothesis that stimulus savings increased flight travel. The Mean Percentage Error (MPE) would be expected to have the same magnitude across all three methods. Although the MAE increased from testing to validation, the magnitude of the error was the key factor under examination.

Standard Deviation of System RPM
11.05309

Model	MAE	MPE
GamLoess Test	3.269890	-2.485899
RandomForest Test	3.528833	-2.649087
Knn Test	6.073475	-3.004581
GamLoess Validation	28.321690	-32.686360
RandomForest Validation	28.321690	0.271388
Knn Validation	28.321690	-3.004581

Conclusion

In conclusion, this report serves as the culmination of the Capstone for Harvard's Professional Certificate in Data Science, fulfilling project requirements by sourcing data, conducting regressions, and summarizing findings across the nine sections of the Data Science Series. The analysis focused on investigating the impact of COVID-19 stimulus checks on air travel patterns. Through regression models and machine learning algorithms, the study aimed to discern any significant effects stemming from the economic stimuli.

Real disposable income emerged as a significant factor affecting air travel, demonstrating a positive relationship in regression analyses. However, the machine learning analysis did not conclusively support the hypothesis that stimulus savings had a substantial effect on air travel. Furthermore, this report acknowledges the limitations of using aggregate data and suggests that employing a random sample of individuals may provide deeper insights into the trends observed.

Overall, while this study sheds light on the complex interplay between economic stimuli and air travel, further research utilizing more granular data and refined methodologies may yield a more comprehensive understanding of the phenomenon.

Citations

Context Review

- Coronavirus disease (COVID-19) pandemic
  - **Source:** World Health Organization. (2023). Coronavirus disease (COVID-19) pandemic. World Health Organization. Link. Accessed on 2024-03-12.
- Stimulus Checks Substantially Reduced Hardship, Study Shows (2021, June 2). *The New York Times* by Jason DeParle Link. Accessed on 2024-03-13.
- "Trump tells aides he supports second round of stimulus checks, but White House divisions remain" by Jeff Stein, Josh Dawsey, and Erica Werner. *Washington Post*. June 23, 2020. Link.

Data Sources

FRED St. Louis FED:

- Consumer Price Index for All Urban Consumers: All Items Less Food and Energy in U.S. City Average (CPILFESL)
  - **Source:** U.S. Federal Reserve Bank of St. Louis. (2023). Consumer Price Index for All Urban Consumers: All Items Less Food and Energy in U.S. City Average [CPILFESL]. FRED, Federal Reserve Bank of St. Louis. Link. Accessed on 2024-02-22.
- Consumer Price Index for All Urban Consumers: Food in U.S. City Average
  - **Source:** U.S. Federal Reserve Bank of St. Louis. (2023). Consumer Price Index for All Urban Consumers: Food in U.S. City Average [CPIUFDSL]. FRED, Federal Reserve Bank of St. Louis. Link. Accessed on 2024-02-22.
- Gross Domestic Product (GDP)
  - **Source:** U.S. Federal Reserve Bank of St. Louis. (2023). Gross Domestic Product [GDP]. FRED, Federal Reserve Bank of St. Louis. Link. Accessed on 2024-02-22.
- M1 (M1SL)
  - **Source:** U.S. Federal Reserve Bank of St. Louis. (2023). M1 Money Stock [M1SL]. FRED, Federal Reserve Bank of St. Louis. Link. Accessed on 2024-02-22.
- M2 (M2SL)
  - **Source:** U.S. Federal Reserve Bank of St. Louis. (2023). M2 Money Stock [M2SL]. FRED, Federal Reserve Bank of St. Louis. Link. Accessed on 2024-02-22.
- Noncyclical Rate of Unemployment (NROU)
  - **Source:** U.S. Federal Reserve Bank of St. Louis. (2023). Noncyclical Rate of Unemployment (NROU) [NROU]. FRED, Federal Reserve Bank of St. Louis. Link. Accessed on 2024-02-22.
- Real Disposable Personal Income
  - **Source:** U.S. Federal Reserve Bank of St. Louis. (2023). Real Disposable Personal Income [DSPIC96]. FRED, Federal Reserve Bank of St. Louis. Link. Accessed on 2024-02-22.
- US Regular All Formulations Gas Price
  - **Source:** U.S. Federal Reserve Bank of St. Louis. (2023). US Regular All Formulations Gas Price [GASREGW]. FRED, Federal Reserve Bank of St. Louis. Link. Accessed on 2024-02-22.

Bureau of Labor Statistics (BLS):

- Airline fares in U.S. city average, all urban consumers, not seasonally adjusted
  - **Source:** U.S. Bureau of Labor Statistics. (2023). Airline fares in U.S. city average, all urban consumers, not seasonally adjusted. U.S. Bureau of Labor Statistics. Link. Accessed on 2024-02-22.

Bureau of Transportation Statistics:

- U.S. Air Carrier Traffic Statistics through November 2023
  - **Source:** Bureau of Transportation Statistics. (2023). U.S. Air Carrier Traffic Statistics. Bureau of Transportation Statistics. Link. Accessed on 2024-02-22.