Investigating the Relationship Between US Fuel Price and Stock Performance of Major Automotive Producers in the US Under Time Series Context

Yijian Jin

Perm# 8916256

PSTAT 174

Section: W 6:00pm

Submission Date: Mar 13th, 2023

PART I - Executive Summary

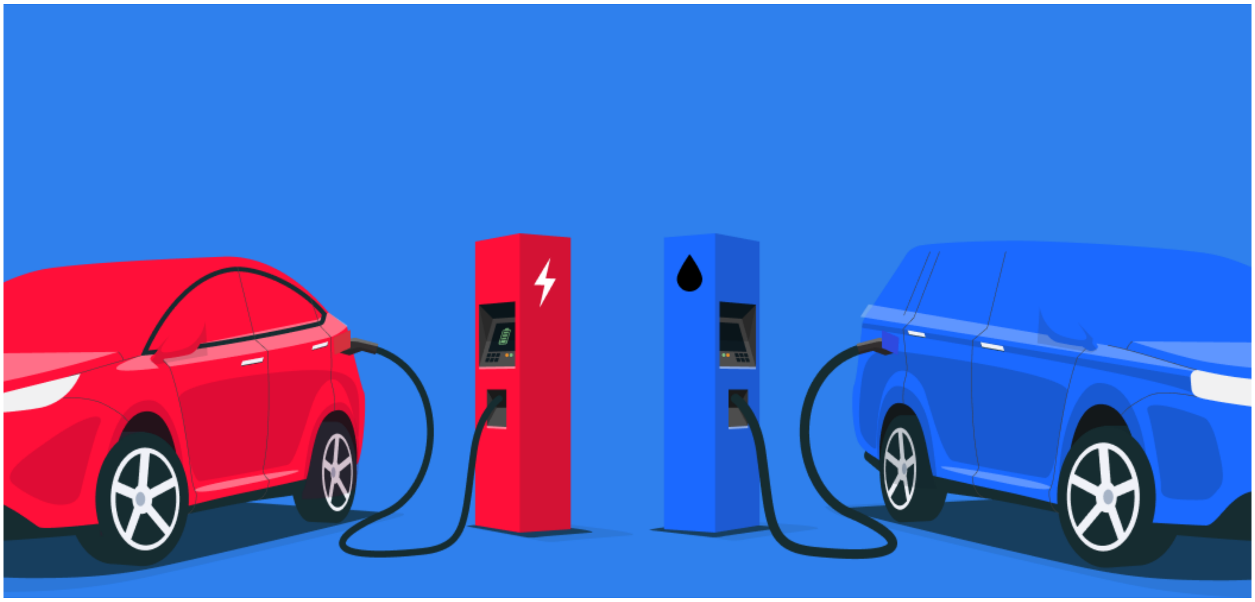


Image excepts from <https://blog.float.sg/ev-vs-gas-what-car-is-cheaper-to-own/>

This project is dedicated to investigating how will the gas price influence stock price of different car manufactures under time series context, particularly electric vehicle (TESLA) and gas vehicle manufactures (FORD, TOYOTA). This report is divided into four major sections: section I will be dedicated to the discussion of the data sets, including overview of research context, interests of study, and the empirical properties of the data; section II will be about model application, which includes model selection, estimation, and model performance diagnostic; lastly, section III will be about model analysis and conclusion, which will be focus on selecting the best models, accuracy analysis, and discussion on all findings relating to the research context. Lastly, the last part of the report will be dedicated to bibliography including source of datasets and reference materials. By fitting data to various ARIMA models, we concluded that electric vehicle manufacturer (Tesla) tends to experience greater growth compared to the traditional car manufactures (Ford and Toyota) since the gas price was experiencing instability throughout the years. Nevertheless, the standard deviation of our models is significantly high, causing exceptionally large prediction interval due to the instability nature of gas and stock price.

PART II – INTRODUCTION

Research Question

Investigating the relationship between US gasoline price and stock performance of representative EV manufacturer and traditional fuel vehicle manufactures under time-series context.

Interest of Study

The future of electric vehicles (EVs) looks promising, with increasing demand for sustainable and eco-friendly transportation solutions. In the coming years, advancements in battery technology are expected to result in longer ranges, faster charging times, and lower costs. Additionally, EV infrastructure and charging networks are rapidly expanding, making it easier for consumers to adopt EVs as their primary means of transportation. Governments around the world are also offering incentives and implementing regulations to support the growth of the EV market. Overall, the trend towards electrification in the automotive industry is expected to continue, with EVs becoming increasingly affordable and accessible to consumers.

People are switching to electric vehicles (EVs) for a variety of reasons, including:

- Sustainability: EVs emit fewer greenhouse gases and pollutants compared to gasoline-powered vehicles, making them a more environmentally friendly option.
- Cost savings: Despite their higher upfront costs, EVs often have lower operating costs due to lower fuel and maintenance expenses.
- Performance: Many EVs offer quick acceleration and smooth, quiet rides.
- Government incentives: Governments around the world are offering tax credits, rebates, and other incentives to encourage people to adopt EVs.
- Convenience: With the growth of charging networks, it is becoming increasingly convenient for people to own and operate an EV.

- Technological advancements: Continuous improvements in battery technology are enabling EVs to travel longer distances on a single charge, making them more practical for everyday use.

Overall, these factors are driving increased interest in EVs, leading more and more people to consider switching from traditional gasoline-powered vehicles.

Gas prices can play a role in people's car purchasing decisions, especially for those who drive frequently and cover long distances. When gas prices are high, it can make owning a gas-powered vehicle more expensive, and can lead some people to consider more fuel-efficient or alternative-fuel vehicles, such as electric vehicles (EVs).

After completing this research project, we will be able to answer whether fuel price may influence the sales/investment figure of representative vehicle producers in the US.

The dataset is downloaded from Yahoo Finance and eia.gov. Yahoo Finance is owned by Yahoo, and Yahoo Finance it is generally considered a credible source of financial news and data. The website has a team of experienced journalists and analysts who report on the latest market trends and provide analysis of financial data. It sources data from reputable providers such as Morningstar, S&P Global Market Intelligence, and Refinitiv, among others, which helps to ensure the accuracy and reliability of the information presented on the website. In addition, there are wide range of sampling frequency (daily, weekly, monthly, etc…) and time period available, which can be easily configured to acquire desired data set.

On the other hand, as a government agency, the EIA is generally considered a credible source of information, and its data and analysis are widely used by policymakers, industry professionals, and researchers. The EIA has a rigorous process for data collection and analysis, and its findings are subject to review by internal and external experts to ensure accuracy and reliability. Similar to yahoo finance, EIA also provide options on the sampling frequency and time period when selecting data.

PART III - Data Analysis and Experimental Design

Data Sets Overview

There are 4 data sets in total that are used in this project, which are stock performance of TESLA Inc, TOYOTA Motor Corporation, FORD Motor Company, and the US average fuel price over time. The head portion of the data set will be provided in the appendix.

The stock data sets (TSLA, FORD, TOYOTA) are downloaded finance.yahoo.com, and their sampling frequency are set to weekly basis from July 5th, 2010, to Jan 30th, 2023. Besides, the average fuel price of US is from www.eia.gov (U.S. Energy Information Administration), and the sampling frequency is also set to a weekly basis from July 5th, 2010, to Jan 30th, 2023. Hence, there are 657 observations in each data set, which is sufficient for training and testing data in forecasting process.

Data Analysis

For the stock data sets, we can see there are 4 variables representing stock price. Each of the Open, Close, Low, and High prices provide different information about the performance of a stock during a specific time period, and none of them can be considered more representative than the others on their own. However, for this project, only the close price will be used for forecasting since we are looking at long term price variation, and the price variation in a single day can be neglected.

Here are four graphs showing the raw data as time series objects.

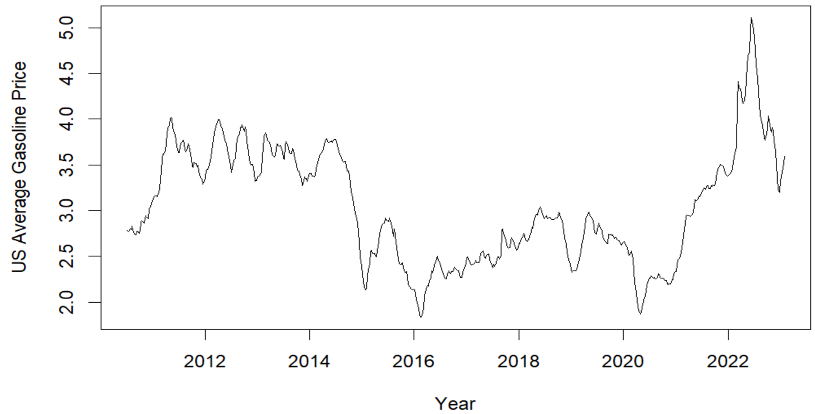


Figure 2.1 – Time Series Plot of Weekly US Average Gasoline Price Since 2010-07-05 to Present

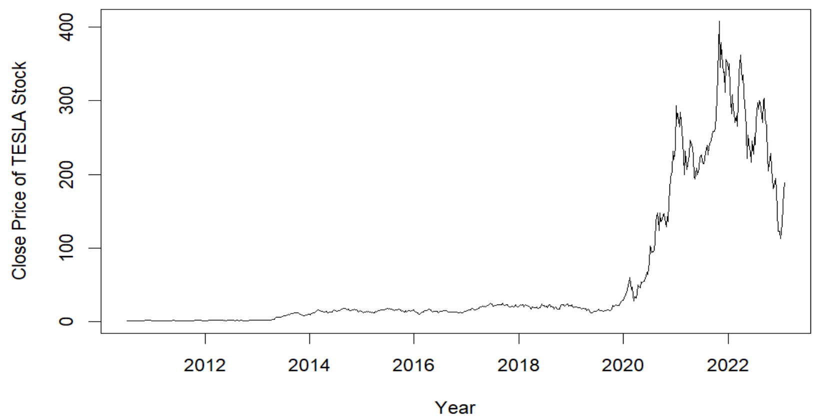


Figure 2.2 – Time Series Plot of Weekly Tesla Stock Closing Price Since 2010-07-05 to Present

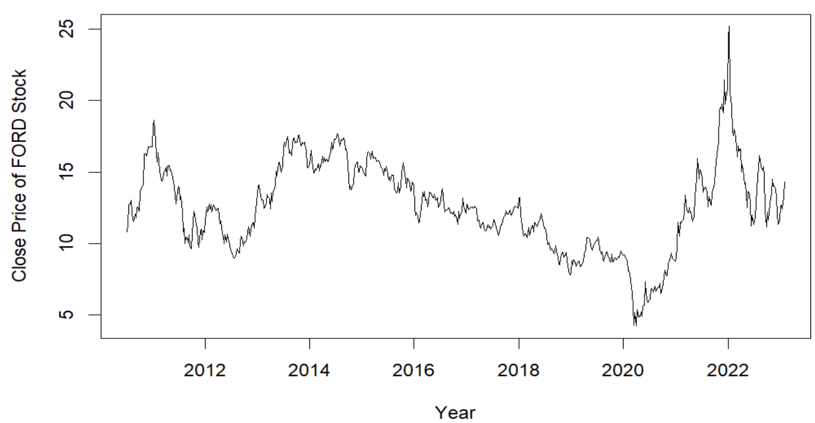


Figure 2.3 – Time Series Plot of Weekly Ford Stock Closing Price Since 2010-07-05 to Present

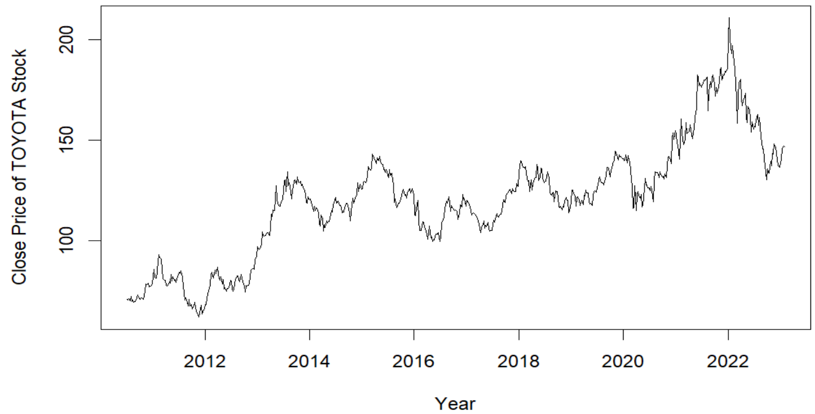


Figure 2.4 – Time Series Plot of Weekly Toyota Stock Closing Price Since 2010-07-05 to Present

Unsurprisingly, most stock data showing an upward trend largely due to economic development. Specifically, Tesla's stock price was showing no significant growth nor decline before the year of 2020, and it suddenly skyrocketed to its peak since 2020 to 2022, cause exceptionally high variation; the stock performance of Ford and Toyota were showing opposite trend where Toyota was experiencing slight growth, but Ford was experiencing diminishment with possible seasonality from 2014 to 2020. Coincidentally, all stock prices and gasoline prices somewhat increase dramatically from 2020 to 2022, and then drop significantly and simultaneously from 2022 to the present. Besides, there are tons of volatility in the data which contribute to random fluctuation. Such high variation will cause difficulty in performing stability analysis and transformation.

Since the data has observable variability, we can perform box-cox transformation to stabilize the data to be closer to normally distributed for our model selection. By applying boxcox() function (explained in appendix 1) to the data, both Ford and Toyota stock data do not require further transformation since their lambda values are close to 1. US gasoline price and tesla stock require further transformation. Here’s a box-cox plot produced by R.

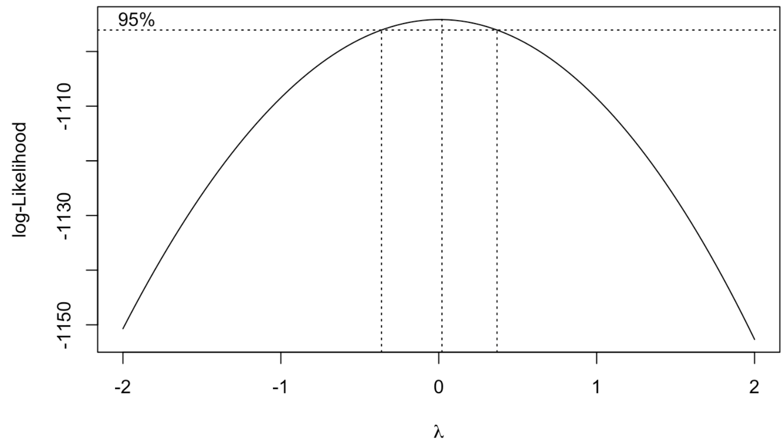


Figure 3.1 – Box-Cox transformation for US Gas Price

As shown in the above plot, the peak lambda value is close to 0, hence we can perform a log transformation to the US Gas Price data.

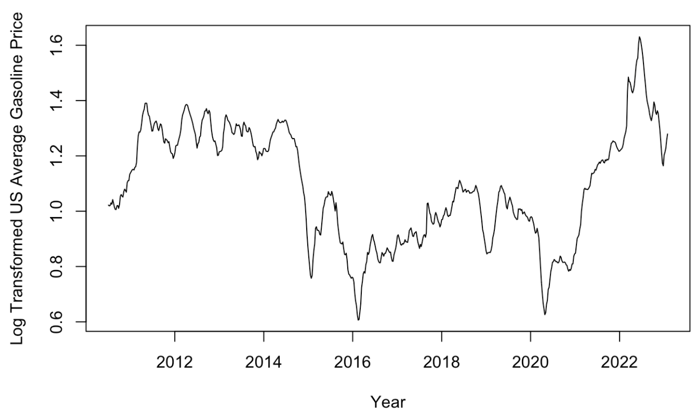


Figure 3.2 – Log-tansformed US Gas Price Time Series Plot

After transformation, the plot does not look too much different from its raw data graph. However, we do see the scale of the variation is reduced.

Now let's look at the box-cox transformation of Tesla's stock close price.

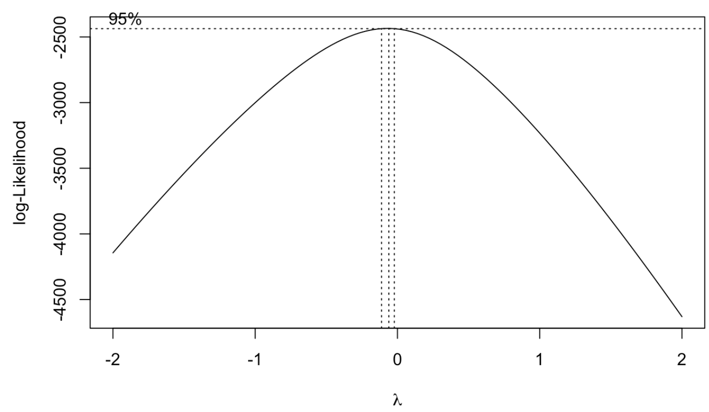


Figure 3.3 – Box-Cox Transformation for TSLA Stock Close Price

Since 0 is not included in the confidence inverval, we are unable to perform a log-transformation as we did for the gas price data, and it is hard to tell what the exact value by merely look at the graph. By extract the lambda value in R, The box-cox plot is suggesting a lambda value of -0.06. We then can input this lambda value to let R do the transformation calculation for us. Here's what the transformed data look like:

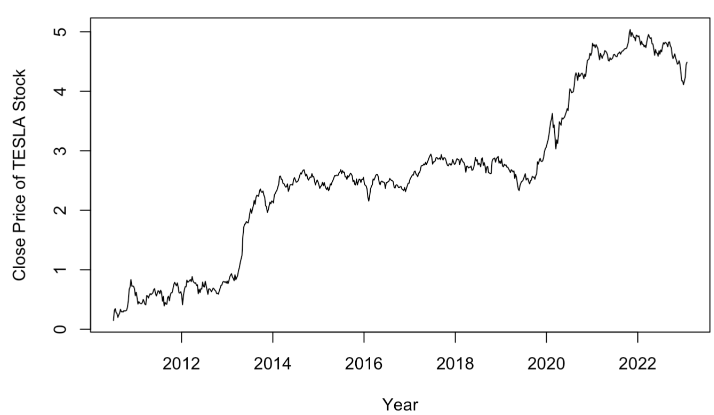


Figure 3.4 – Box-Cox Transformed TSLA Stock Time Series Plot

We can see the trend is clearer after the transformation, and the scale of variation is minimized.

Here are ACF plot for the original and transformed data:

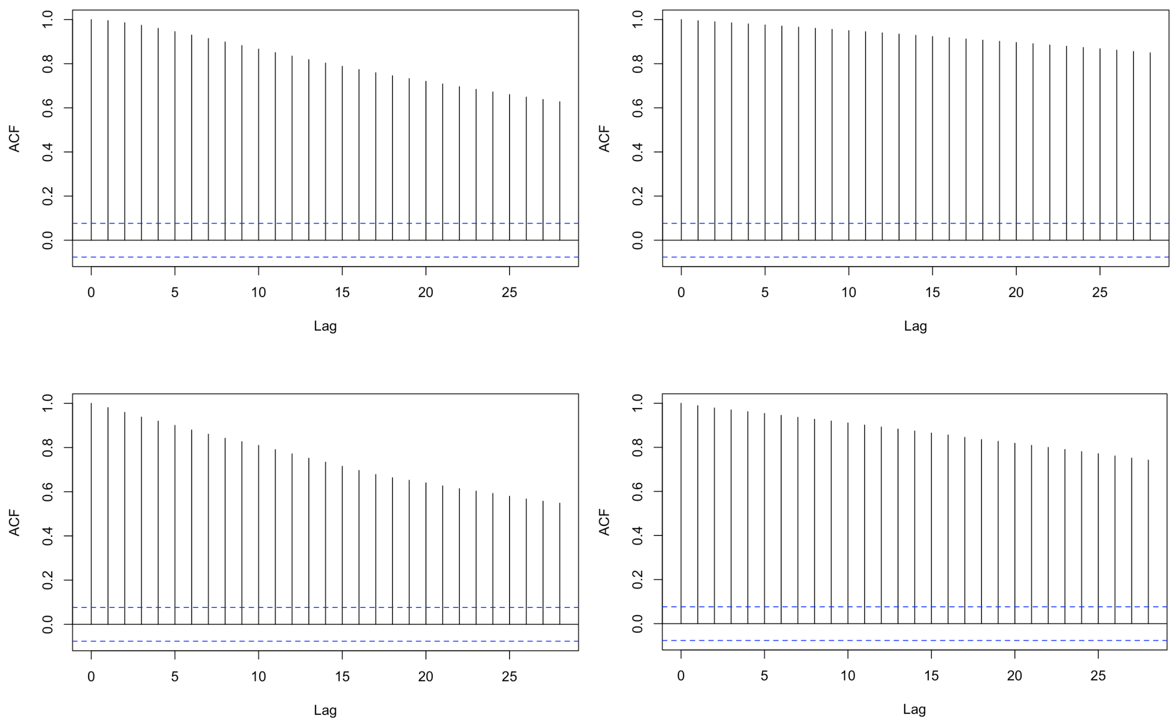


Figure 3.5 – ACF Graphs for US Gas Price, TSLA, FORD, TOYOTA Stock Price Respectively

ACF stands for Autocovariance Function, As shown in the ACF graphs in figure 3.5, the ACFs are very high and decrease slowly as lag value increases. Because of the strong serial dependence and visible trend in the raw graph, the data is definitely not stationary, we might consider differencing the data to remove the trends and seasonality effect to improve stationarity. Using stationary time series data for forecasting is important because stationary time series have stable statistical properties over time, such as a constant mean, variance, and autocorrelation. This stability makes it easier to model and forecast the time series, as the patterns observed in the past are likely to continue in the future.

We will begin with US gas price data. We can use unit-root test to determine whether differencing is necessary objectively. In this test, the null hypothesis is that the data are stationary, and we look for evidence that the null hypothesis is false. Consequently, small p-values (e.g., less than 0.05) suggest that differencing is required. The test can be computed using the ur.kpss() function from the urca package. By applying unit-root test, we obtain a critical value of 1.7055, which is much larger than 0.01 critical value. In this case, we can difference the data, and the result shows a critical value of 0.0867, which is tiny, and well within the range we would expect for stationary data. So we can conclude that the differenced data are stationary. We can do the same test for the rest stock data. and their respective test statistics are 8.14, 1.37, and 6.5449, which are larger than 0.01 critical value. Thus, we are taking first difference on all stock data, we obtain test statistics 0.0884, 0.0473, 0.0552. We now can conclude that all data are stationary.

Let's look at their ACF and PACF graph after differencing.

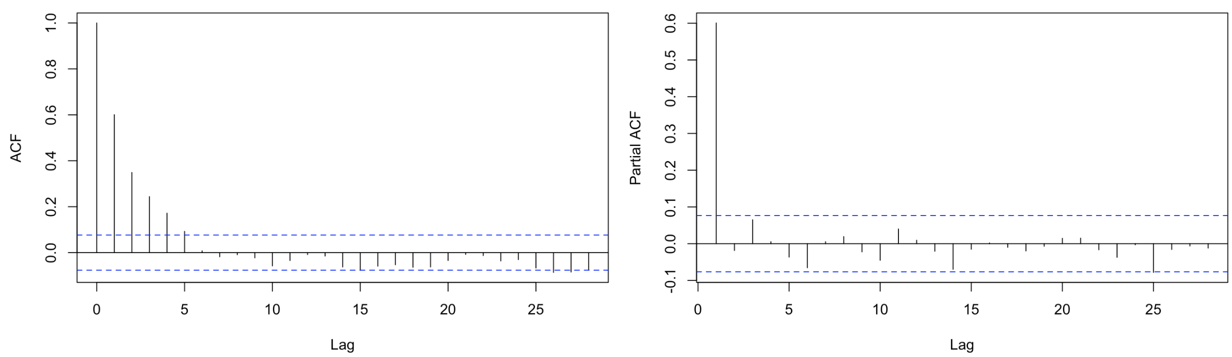


Figure 3.6 – ACF and PACF for Differenced US Gas Price data

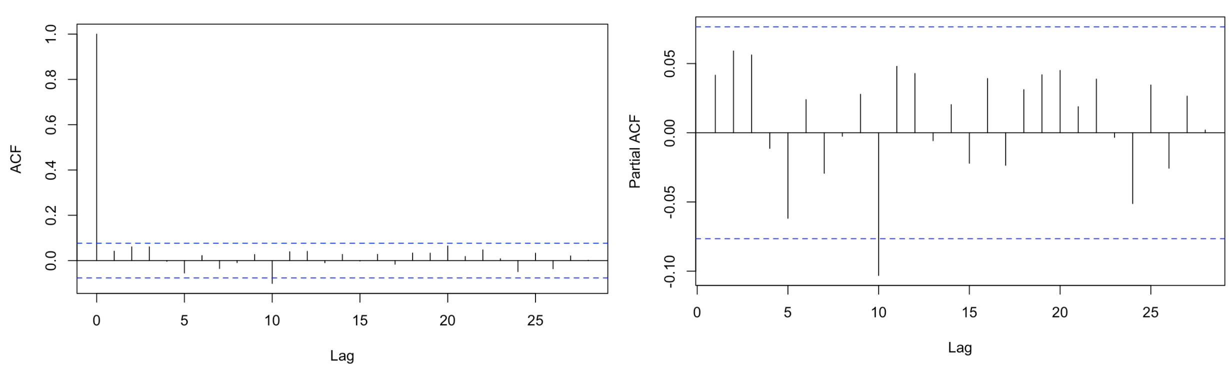


Figure 3.7 – ACF and PACF for Differenced TSLA Stock Price data

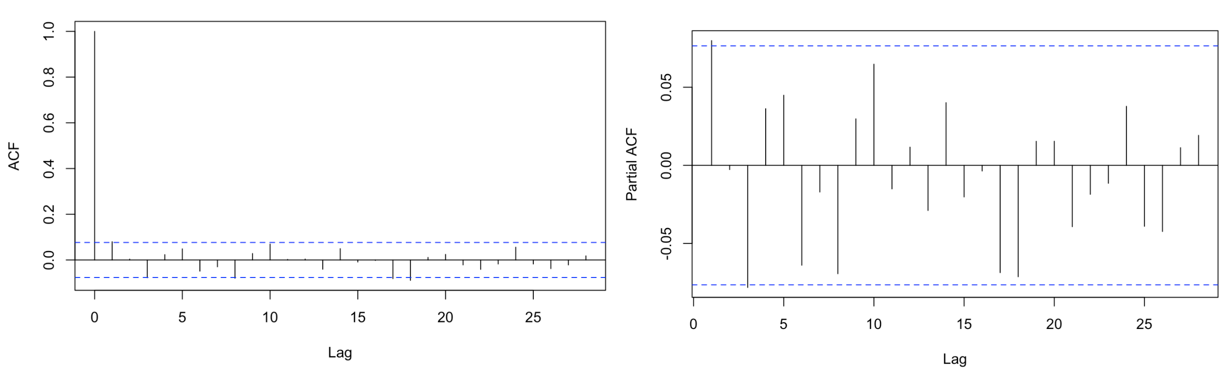


Figure 3.8 – ACF and PACF for Differenced FORD Stock Price Data

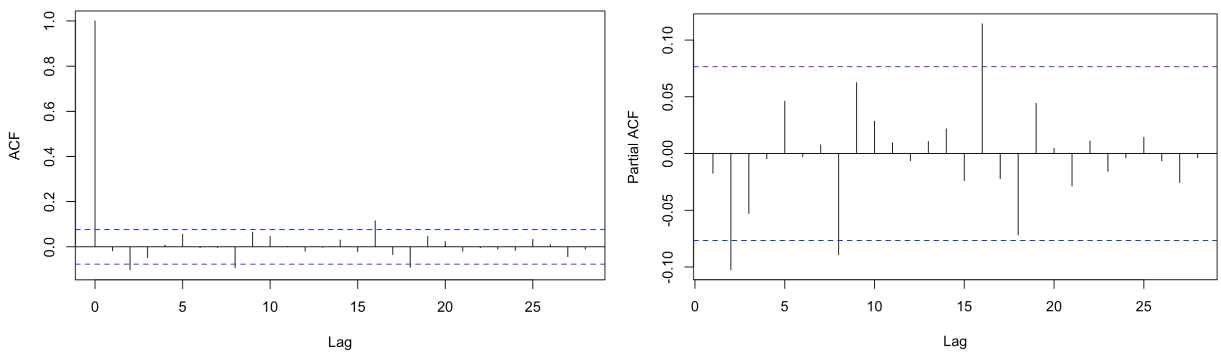


Figure 3.9 – ACF and PACF for Differenced TOYOTA Stock Price Data

The plots look much better now for model selection. The plots of differenced data are included in the appendix.

PART IV - Time Series Modelling and Forecasting

Model Selection

For this project, we will be primarily use ARIMA() and SARIMA() models for our data set. Detail about those models will be illustrated in the appendix section

Average Weekly US Gas Price

We will begin with the data of US Gasoline Price. By looking at the original log-transformed plot in figure 3.2, there is a price jump halfway from year 2014 to 2015, and it climbed back after 2021. This price variation usually led to large outliers in subsequent analysis. Besides, in figure 3.6, both ACFs and PACFs decay quickly, confirming that the series is weakly stationary. In the ACF plot, the first five lags are significantly different from 0; while only the first lag of PACF are significantly different from 0. Without considering the seasonal effect, the PACF suggest a AR(1) components, and ACF shows a decaying pattern, which suggesting there's no MA(q) components. We will fit an ARIMA(1,1,0) model along with the variations including ARIMA(2,1,0), and ARIMA(1,1,1) model. In addition, we will implement auto.arima() function in R to help us to determine the best possible model.

We will also use AIC (Akaike's Information Criterion) to help us to determine the best model fit. In short, good models are obtained by minimizing the AIC or BIC. Our preference is to use the AIC.

The auto.arima() function also provide us an ARIMA(1,1,0) model. Based on the AIC value, ARIMA(1,1,0) gives the lowest AIC value of -3668.63. The rest models can be neglected. We will then check the residuals.

The fitted model is:

$$(1 - 0.6012B)(1 - B)G_t = a_t, \sigma_a^2 = 0.000217$$

Where $Y_t = (1 - B)G_t$ is the differenced data.

Here's its residual plots:

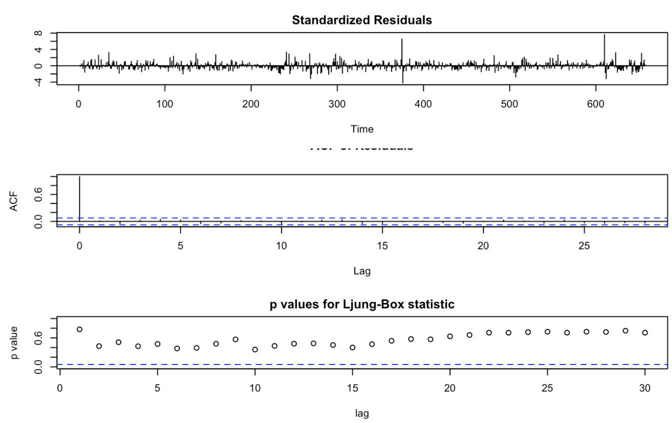


Figure 4.1 – Residual Plot of ARIMA(1,1,0) Model

As shown in Figure 4.1, our model has normal distributed residuals. Within each correlogram, the autocorrelations of all lags are within the significant level. Its Ljung-Box test gives a p-value of 0.2766, suggesting normally distributed residuals.

Tesla

Now let's look at Tesla's stock price data. In figure 3.7, by examining the ACF and PACF plot, it suggests that the differenced data are weakly-stationary as most autocovariance/partial autocovariance are within the significant level except lag 10, which surpasses the significance level in both ACF and PACF plot; also, lag 20 in ACF is almost significant. Can it be a ARIMA(10,1,10) model? As we try to fit ARIMA(10,1,10) model to the data in R, a warning appears telling us there's possible convergence problem. Let's take a different approach. There may be seasonal component exist in the data. We might be better off to check the seasonally differenced data at lag 10.

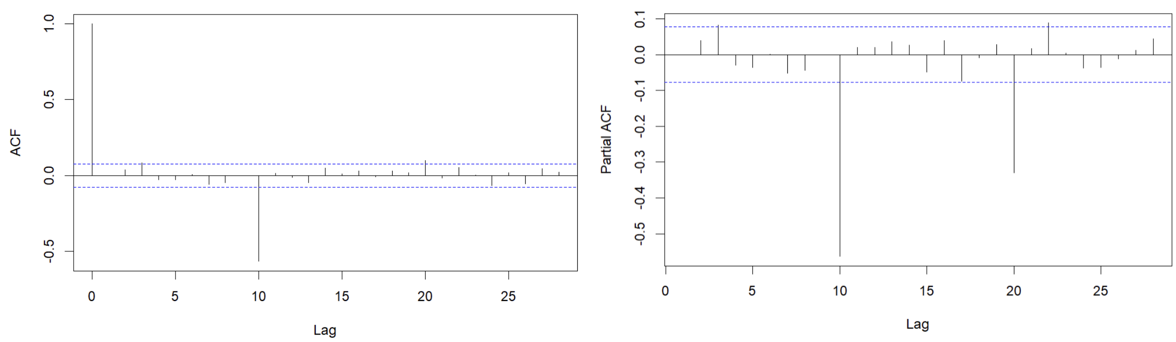


Figure 4.2 – ACF and PACF of Seasonally differenced TSLA Stock data

As shown in figure 4.2, we can see the lag-3 PACF and ACF values are significant, indicating an AR(3) and a MA(3) component exist in the model; in addition, as the ACF and PACF exhibit strong seasonal pattern in lag 10, 20, …, we will then consider adding seasonal AR(1) and MA(1) components. The fitted model will then be ARIMA(3,1,3)[1,1,1] with period 10. In addition, variations including ARIMA(0,1,3)[0,1,1] and ARIMA(3,1,0)[1,1,0] will also be tested.

The fitted model is:

$$(1 - 2.4044B + 2.3121B^2 - 0.8908B^3)(1 + 0.1559B^{10})(1 - B)(1 - B^{10})G_t$$
$$= (1 - 2.3528B + 2.2526B^2 - 0.8710B^3)(1 - 0.9999B^{10})a_t, \sigma_a^2 = 0.003865$$

Lastly, we will also implement auto.arima() function with seasonal component enabled to let R help us find a simpler model. The auto.arima() provide us with ARIMA(2,1,2) with drift 0.0067, which does not suggest a seasonal component.

As expected, the model approximated by auto.arima() has a lower AIC value. Let's look at their residual plots.

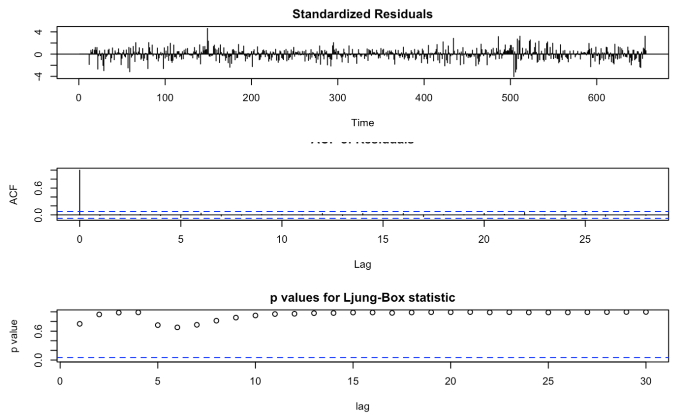


Figure 4.3 – Residual plot of ARIMA(3,1,3)[1,1,1]10 model

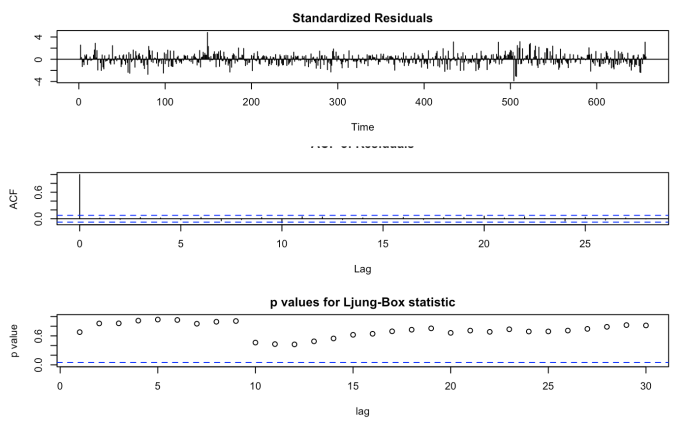


Figure 4.4 – Residual plot of ARIMA(2,1,2) with drift model

As shown in the residual plots, the residuals of ARIMA(3,1,3)[1,1,1]10 model are perfectly contained within the significance level, indicating that the residuals are normally distributed with not lag effect. Besides, the Ljung-Box test gives a p-value of 0.2152, suggesting the residuals are indeed normally distributed (detail on this test is provided in the appendix). The residuals of ARIMA(0,1,3) are somewhat normally distributed but there's still a significant spike at lag-10 in the ACF plot, which is similar to what we got in the original ACF plots. Hence, we will be using ARIMA(3,1,3)[1,1,1]10 model for further interpretation on Tesla's stock performance.

Ford

Moving on, we will then look at Ford's stock data. As shown in figure 3.8, there are several significant spikes at lag 1, lag 3, lag 8, and beyond lag 10 that may be considered as noise in ACF plot; on the other hand, the PACF plot only shows significant spikes at lag 1 and lag 3. Since there are no other obvious pattern in the plots, we may consider fit a ARIMA (3,1,0) model for a trial. As always, auto.arima() will be used to let R help us to determine a potential suitable model. Based on the output, auto.arima() provide us with an ARIMA(2,1,4) model, which is quite the contrary with what we guessed. Let us check their residual plots.

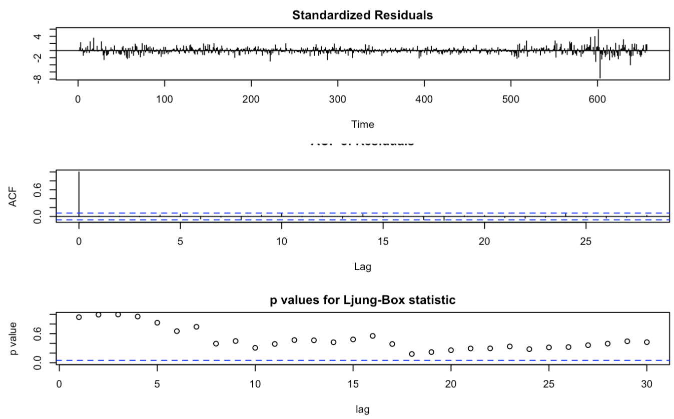


Figure 4.5 – Residual plot of ARIMA(3,1,0) model on Ford's Stock Data

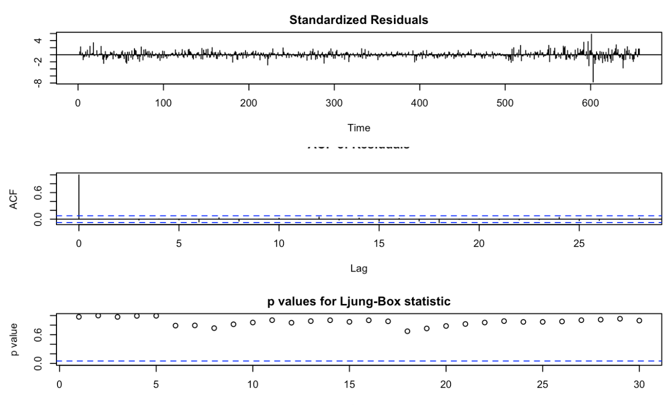


Figure 4.6 – Residual plot of ARIMA(2,1,4) model on Ford's Stock Data

As shown in figure 4.5 and figure 4.6, the residual plot of ARIMA(2,1,4) shows a bit more promising result with only a significant ACF at lag 18, which may be considered as noise. Nevertheless, based on the residual distribution of both models, we can see the variance starts to increase since the year of 2020. This can be seen in the raw time series plot of

Ford's stock data where the closing price reaches lowest at the beginning of 2020 and then suddenly rocked to a new peak around 2022. Based on the residual ACF plot, what if we add a seasonal component to the model? According the ACF and PACF plot, we can see a significant spike at lag 18 of ACF but not PACF, so we can try add MA(1) seasonal component to the model. Let's try modify the model to ARIMA(2,1,4)[0,0,1]$_{18}$ and check its residual plot. However, given that period of 18 is roughly 4 month in a year, the real implication of this period is unknown.

The fitted model is:

$$(1-0.4279B+0.9830B^2)(1-B)G_t$$
$$=(1-0.3558B+0.984B^2+0.0274B^3+0.0283B^4)(1-0.1034B^{18})a_t,$$

$$\sigma_a^2=0.3431$$

Here's its residual plot:

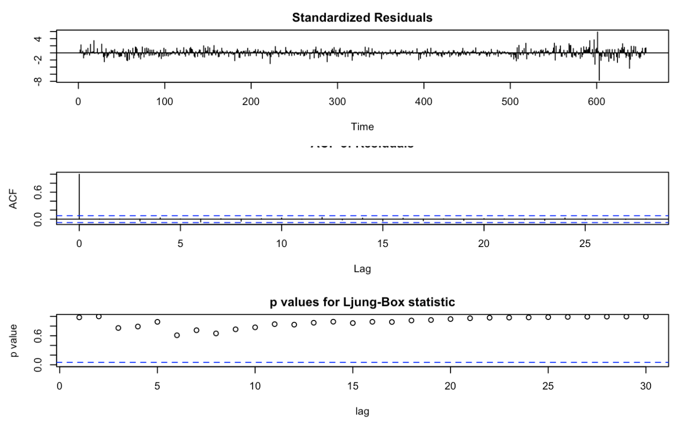


Figure 4.8 – Residual plot of ARIMA(2,1,4)[0,0,1]$_{18}$ model on Ford's Stock Data

Surprisingly, the residual behaves like normal distributed as all autocorrelation value are contained within the significance level. The p-value from its Ljung-Box test is 0.09071, which is slightly bigger than the reject value of 0.05, indicating the residual is normally distributed.

Toyota

Finally, let's move on the Toyota's stock data. Based on the raw data time series plot, the pattern of its closing price is somewhat similar to ford's. By examining the ACF/PACF plot, there are several noticeable spike in both the ACF and PACF plot. In the ACF plot, we can see spikes at lag 2, lag 8, lag 16 and lag 18; in the PACF plot, we can see spikes at lag 2, lag 8 and lag 16, which resembles the pattern in ACF. Let us have a trial run on ARIMA(2,1,2) to test the residual response. Interestingly, our auto.arima() function gives a same result.

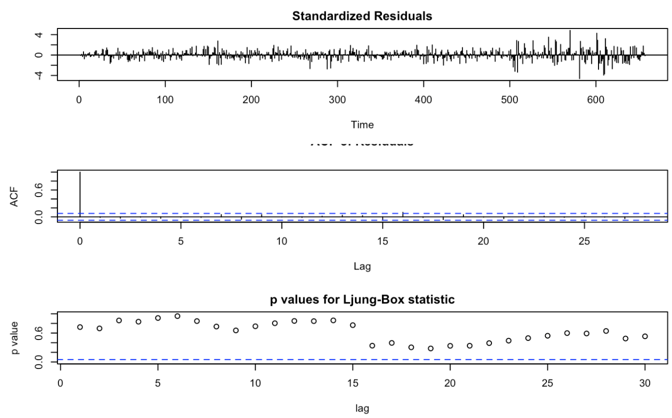


Figure 4.7 – Residual plot of ARIMA(2,1,2) model on Toyota's Stock Data

As shown in the residual plot, we can see a significance spike of ACF at lag 16. We can try fix this problem with a same approach as we did in Ford's model by adding a seasonal component AR(1) and MA(1) since both ACF and PACF plot shows a significance spike at lag 16. Let's try ARIMA(2,1,2)$[1,0,1]_{16}$ and check its residual plot.

The fitted model is:

$$(1 - 0.554B + 0.8543B^2)(1 - 0.2986B^{16})(1 - B)G_t = (1 + B + B^2)(1 - 0.181B^{16})a_t,$$

$$\sigma_a^2 = 12.5$$

And its residual plot:

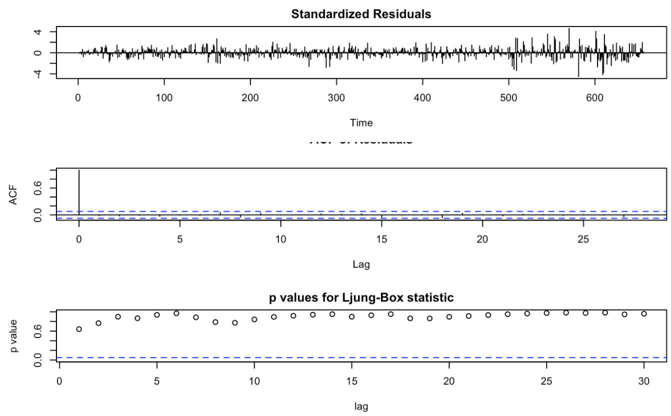


Figure 4.8 – Residual plot of ARIMA(2,1,2)[1,0,1]$_{18}$ model on Toyota's Stock Data

As shown in the ACF plot of the residual analysis, there is no more significance spikes indicating the residuals are now normally distributed, and the Ljung-Box test has a p-value of 0.2253, indicating the residuals are normally distributed.

Summarize

Here's a summarize of our model choice for each data set:

Dataset	Model
Transformed US Weekly Gas Average	ARIMA(1,1,0)
Transformed TESLA Stock Close Price	ARIMA(3,1,3)[1,1,1]$_{10}$
FORD Stock Close Price	ARIMA(2,1,4)[0,0,1]$_{18}$
TOYOTA Stock Close Price	ARIMA(2,1,2)[1,0,1]$_{16}$

Model Prediction/Forecast

To have a overlook of what our model prediction may look like, we can use the forecast() function. We set h=104 to predict the growth in roughly next two years.

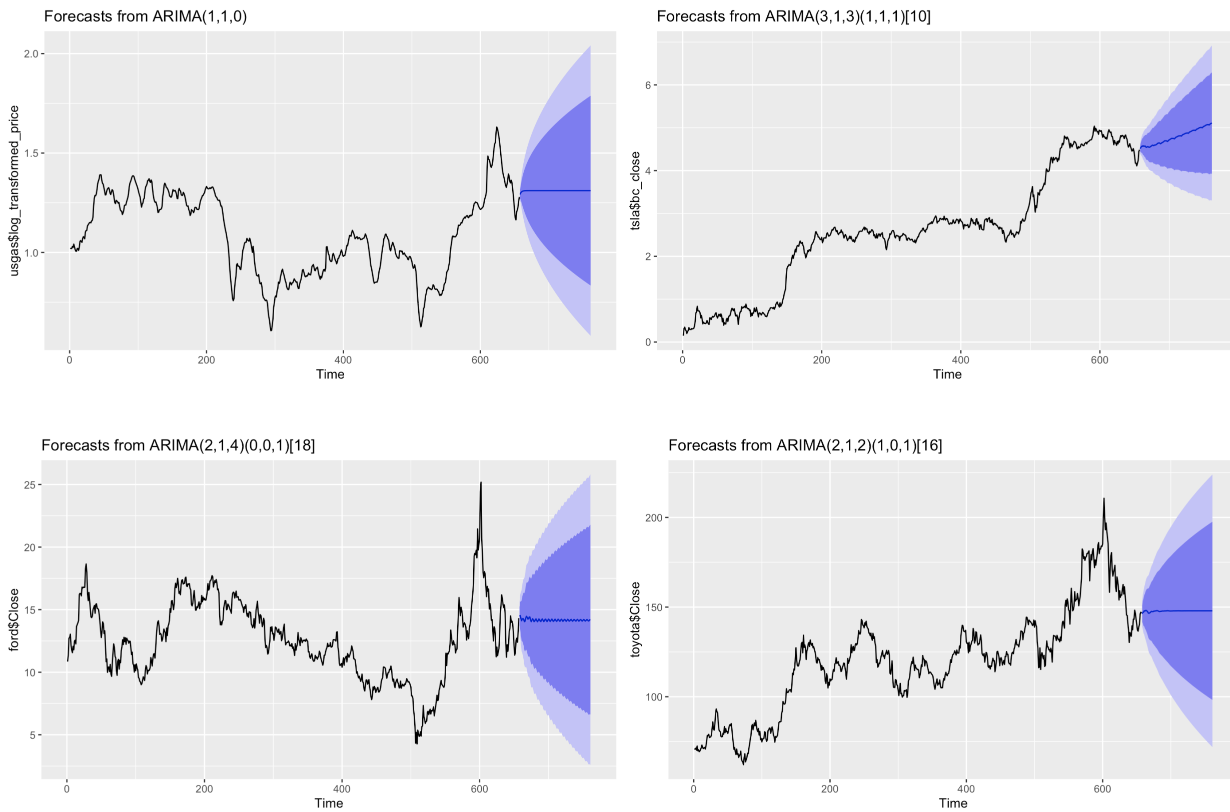


Figure 5.1 – initial forecast made by forecast() function

Based on this initial prediction, the prediction intervals are getting wider in most models due the the huge variability of data; except for Tesla's prediction interval, which is comparatively narrower than other's. Solely based on the trend line in each forecast, we can see a upward trend for all car manufacturers' stock performance with variaed slope.

We can test out model prediction by fitting past value then compare it to our test data. we can fit the model by removing some value from our training data set and use the removed value as our testing data points. Here are the fitting results:

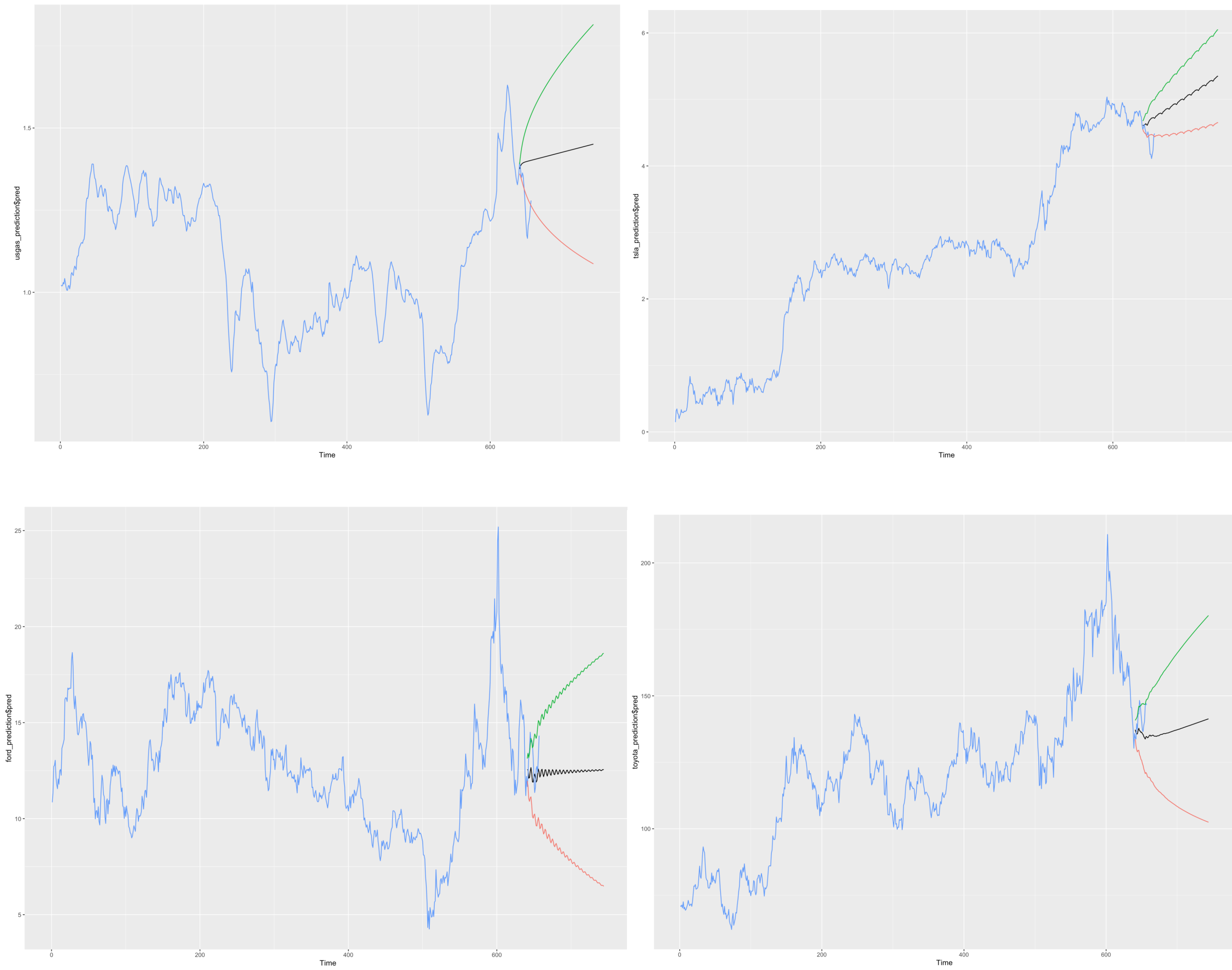


Figure 5.2 – Forecast VS Part of Actual Value with Future Values

In the above plots, the blue line is the actual stock/gas price data, and the black line is the prediction made by our model, where the green and red line represents the upper bound and lower bound of prediction interval respectively. As we can see in the plot, the prediction of stock data kind of resembles the past value, but the us gas data seems different from our prediction. However, the standard error must be significantly large considering the prediction interval shown in figure 5.1 and figure 5.2 almost suggests a random trend except for the tesla's stock data that promises a relatively stable trend and variance. Nevertheless, the trend of the rest stock data seems to capture the overall trend of the past stock performance.

PART V - Discussion and Conclusions

To answer our research question, as shown in figure 5.2, based on what our model predicted, all stock close price shows a upward trend with varied slope, where the growth of Tesla’s stock looks more positive compared to that of Ford and Toyota. With the frequent fluctuation of average gas price in the United States, the growth of traditional automotive company is steady but without prominanent progress. This may be caused by the fact that more customers are considering switching to envionmental friendly energy driven vehicle such as electric car. Tesla, which is the representative of new energy vehicle manufactuer, possess the momentum of a more promising growth for investors, as we can see in its stock performance and model prediction. While our prediction may have provide us with some insight into the relationship between gas price and the stock performance of representative car manufacturers, again, causing by the high variance and fluctuation in the stock market, it is hard to tell the future trend, which lead to the exceptionally wide prediction interval we encountered in the forecasting plot.

There are potential limitations in this project. Firstly, the data we are using for our model constuction is the close price of each stock data. However, we ignored open, high, low, and volume components in the stock data. Together, open, close, high, low, and volume provide a complete picture of the price movements and trading activity in a particular market or security. They are used by traders and analysts to make informed trading decisions and to identify trends and patterns in the market. While the closing price of a stock is an important factor in analyzing stock data, it is generally not suitable to solely rely on the close price for analysis since the closing price doesn't provide context about the overall trend of the stock or the broader market. For example, a stock may have closed higher on a particular day, but if the market experienced a significant decline, the stock's increase in price may not be significant.

In addition, we are only implementing ARIMA model for fitting our data and make forecasts. Even though ARIMA model are popular statistical method used for time series forecasting, when it may not perform ideally when it encounters non-stationary data or data with complex patterns, such as stock price. Hence, while ARIMA models are flexible, they may not be able to capture all the complexities of some time series data. For improvements, we can try fitting more model options such as ARDL (time series regression) to see the response.

To summrize, in this project, we investigates the relationship between gas price and the stock performance of three representative car manufacturers in the United States under time series context. After analysis of basic properties of data sets, and applying differcing to improve the stationality of the data, we are able to carry out basic predictions using ARIMA models. With some modification and tweak to the models such as adding higher orders and seasonal effect, we are able to control the residuals under normal assumption. However, the final prediction is somewhat unsatisfactory because the wide predcition interval and high standard error merely suggests any significant result due to the huge variance presents in the nature of gas and stock data.

PART VI – Bibliography and Appendix

Bibliography

(no date) Weekly U.S. regular all formulations retail gasoline prices (dollars per gallon). U.S. Energy Information Administration (EIA). Available at: https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=EMM_EPMR_PTE_NUS_DPG&f=W (Accessed: March 10, 2023).

Tesla, Inc. (TSLA) stock price, news, Quote & History (2023) Yahoo! Finance. Yahoo! Available at: https://finance.yahoo.com/quote/TSLA?p=TSLA&.tsrc=fin-srch (Accessed: March 10, 2023).

Ford Motor Company (F) stock price, news, Quote & History (2023) Yahoo! Finance. Yahoo! Available at: https://finance.yahoo.com/quote/F?p=F&.tsrc=fin-srch (Accessed: March 10, 2023).

Toyota Motor Corporation (TM) stock price, news, Quote & History (2023) Yahoo! Finance. Yahoo! Available at: https://finance.yahoo.com/quote/TM?p=TM&.tsrc=fin-srch (Accessed: March 10, 2023).

Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. (Accessed: March 10, 2023).

Appendix

1.

	Date	Open	High	Low	Close	Adj.Close	Volume
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	2010-07-05	1.333333	1.333333	0.998667	1.160000	1.160000	383259000
2	2010-07-12	1.196667	1.433333	1.126667	1.376000	1.376000	231583500
3	2010-07-19	1.424667	1.483333	1.300000	1.419333	1.419333	107635500
4	2010-07-26	1.433333	1.433333	1.303333	1.329333	1.329333	45780000
5	2010-08-02	1.366667	1.478667	1.301333	1.306000	1.306000	65995500
6	2010-08-09	1.326667	1.332000	1.159333	1.221333	1.221333	63249000

Appendix 1.1 – Head Portion of Tesla’s weekly stock performance

	Date	Open	High	Low	Close	Adj.Close	Volume
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>
1	2010-07-05	10.42	10.94	10.04	10.85	7.087034	266900400
2	2010-07-12	10.84	11.92	10.83	11.34	7.407092	363254600
3	2010-07-19	11.44	12.75	11.23	12.72	8.308484	434329700
4	2010-07-26	12.82	13.18	12.58	12.77	8.341143	450845500
5	2010-08-02	13.06	13.24	12.77	13.04	8.517502	359481700
6	2010-08-09	13.12	13.12	11.90	12.15	7.936166	267122600

Appendix 1.2 - Head Portion of Ford’s weekly stock performance

	Date	Open	High	Low	Close	Adj.Close	Volume
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>
1	2010-07-05	70.20	71.21	69.42	71.07	60.93841	2138700
2	2010-07-12	70.80	73.34	70.00	70.75	60.66403	3396900
3	2010-07-19	70.88	71.72	68.91	71.31	61.14419	3834700
4	2010-07-26	70.91	71.89	68.70	70.23	60.21816	4050400
5	2010-08-02	71.46	74.67	71.35	72.52	62.18171	6629700
6	2010-08-09	72.05	72.25	69.60	69.87	59.90949	3414200

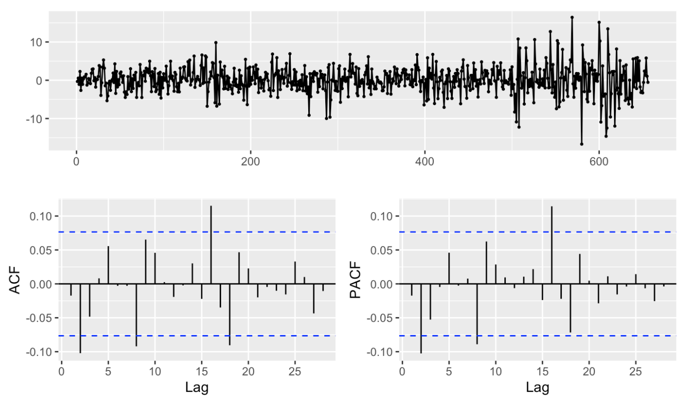
Appendix 1.3 - Head Portion of Toyota’s weekly stock performance

	Week.of <chr>	Weekly.U.S..All.Grades.All.Formulations.Retail.Gasoline.Prices.Dollars.per.Gallon <dbl>
1	01/30/2023	3.594
2	01/23/2023	3.519
3	01/16/2023	3.416
4	01/9/2023	3.366
5	01/2/2023	3.331
6	12/26/2022	3.203

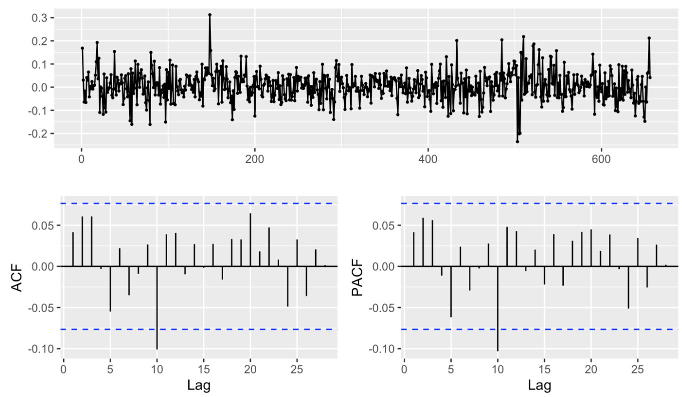
Appendix 1.4 - Head Portion of the weekly average gasoline price in US

2. Box-Cox transformaiton is a statistical technique used to transform a non-normal dependent variable into a normal distribution. The transformation involves raising the variable to a power, λ, which is determined using maximum likelihood estimation. The formula for the Box-Cox transformation is: $y(\lambda) = [(y^\lambda)-1]/\lambda$ where y is the original variable and λ is the parameter that determines the transformation. More detail regarding this transformation is presented in the .Rmd file

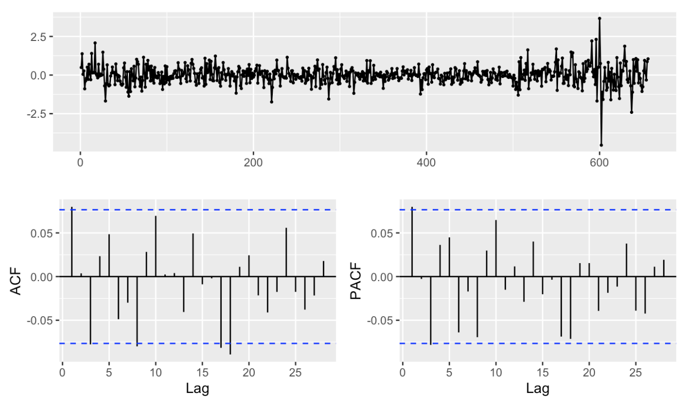
3.



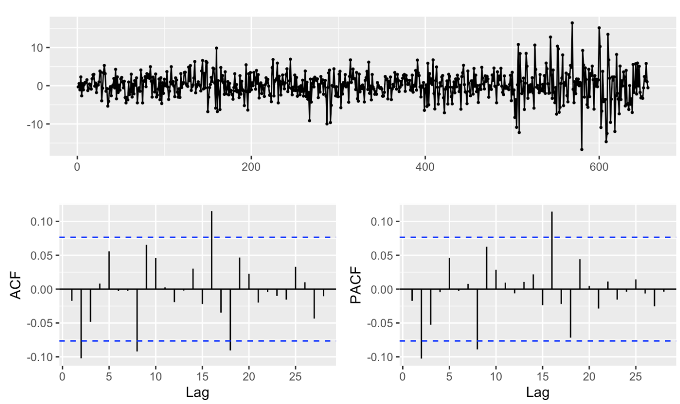
Appendix 3.1: First difference of log-transformed Average US Gas price data



Appendix 3.2: First difference of box-cox transformed Tesla Stock Close Price data



Appendix 3.3: First difference of Ford Stock Close Price data



Appendix 3.4: First difference of Toyota Stock Close Price data

4. Ljung-Box test is a statistical test used to check for the presence of autocorrelation in a time series. The test is a type of goodness-of-fit test that checks whether the autocorrelation coefficients of a time series are significantly different from zero at various lags. The null hypothesis of the test is that there is no autocorrelation in the

time series. The test statistic is calculated as Q = n(n+2)∑(r^2)/(n-i), where n is the number of observations, i is the number of lags being tested, and r is the autocorrelation coefficient at lag i. The test statistic follows a chi-square distribution with i degrees of freedom under the null hypothesis. If the calculated test statistic is greater than the critical value from the chi-square distribution at a given significance level, the null hypothesis is rejected, indicating that there is significant autocorrelation in the time series. If the test statistic is less than the critical value, the null hypothesis is not rejected, indicating that there is no evidence of significant autocorrelation.

5. ARIMA(p,d,q) model is the combination of autoregression (AR) and moving-average(MA) models. It can be written in the following format:

$$\left(1-\phi_1 B-\cdots-\phi_p B^p\right)(1-B)^d y_t=c+\left(1+\theta_1 B+\cdots+\theta_q B^q\right)\varepsilon_t$$

Where p is the order of autoregressive part, d is the order of differencing, and p is the order of moving average part. The same stationarity and invertibility conditions that are used for autoregressive and moving average models also apply to an ARIMA model.

For seasonal ARIMA model, we can write it as ARIMA(p,d,q)[P, D, Q]$_m$, . For a ARIMA(1,1,1)[1,1,1]$_4$ model, we can write it as:

$$(1-\phi_1 B)(1-B)(1-\Phi_1 B^4)(1-B^4)y_t=(1+\theta_1 B)(1-\Theta_1 B^4)\varepsilon_t$$

Where P, D, M represent the order of the seasonal ARMA components.