1. Abstract

The study fitted two models, the harmonic regression with trend, and the ARIMA model, to supplies of US finished motor gasoline product from 1991 to 2017. The study made a prediction of year 2018, provided theoretical and practical explanations, and gave realistic suggestions. The result of the study is that gasoline supplies are likely to experience a smooth increase followed by a fluctuated decrease in 2018. The beginning and ending amount is 90% likely to be at least 5% higher than the counter part of 2017, but the peak is 80% likely to be about 5% around the value of the same period in 2017.

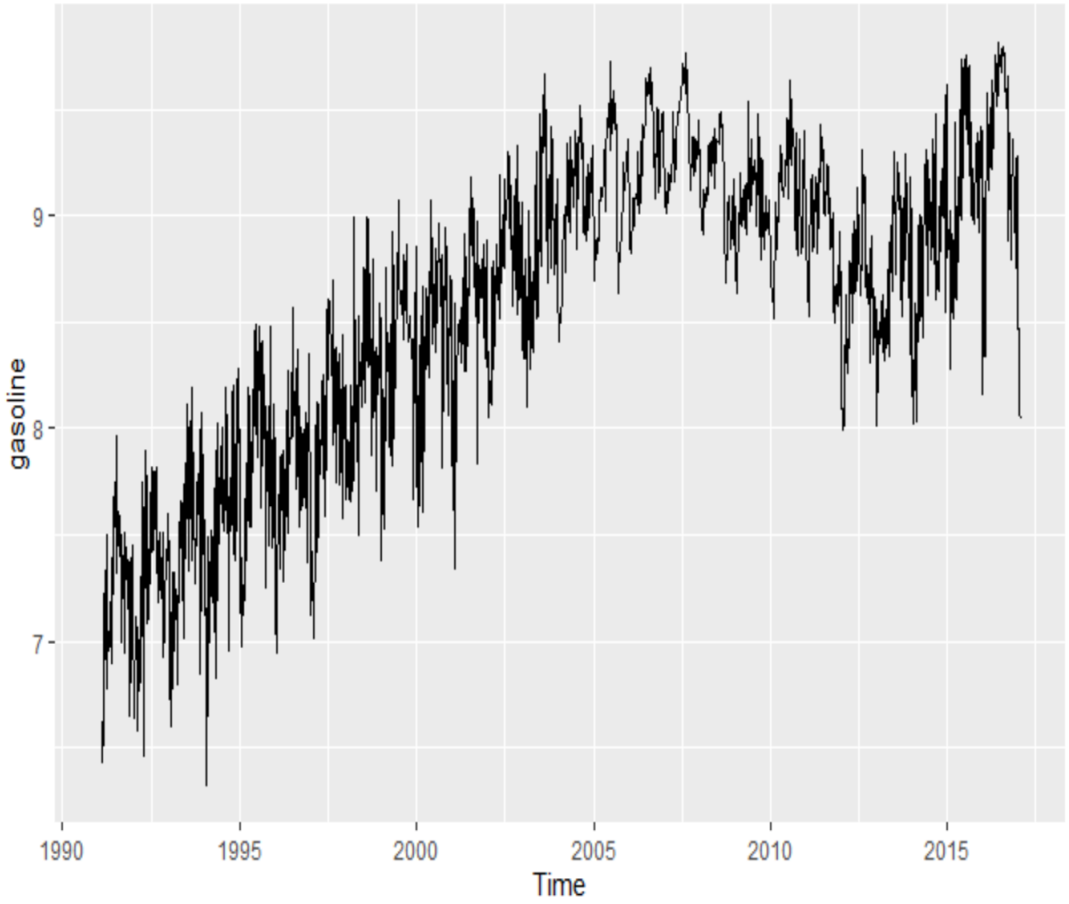
2. Introduction

Nowadays, environmental protection has arosen people's attention, especially towards supply and demand of gasoline. Therefore, the original motivation of the study is to analyze from past data and try to find a useful model to predict the future usage for gasoline one year ahead.

The dataset used in the paper is a time series data consists of weekly data for supplies of US finished motor gasoline product, from 2 February 1991 to 20 January 2017. The units are in "million barrels per day."

With respect to additional relevant scientific information, during the time period, oil price had dropped due to the drivers' successful boycott in 1997 (Erin O'Neill, 2012) and decreased despite tensions in oil-producing regions such as the Middle East and South America (L.A. TIMES ARCHIVES, 2002). However, the US oil price began to rise since 2007 (U.S. News, 2007) and after one decade, the improving economy and OPEC production cuts caused oil prices to rise for 10% over only two months in 2013 (Steve Hargreaves, 2013). Since the data during the period had experienced large-scale fluctuations, it is plausible to suppose that if the created model can well handle the task, the model-creating process may be widely applied to predict the further future with more adequate information.

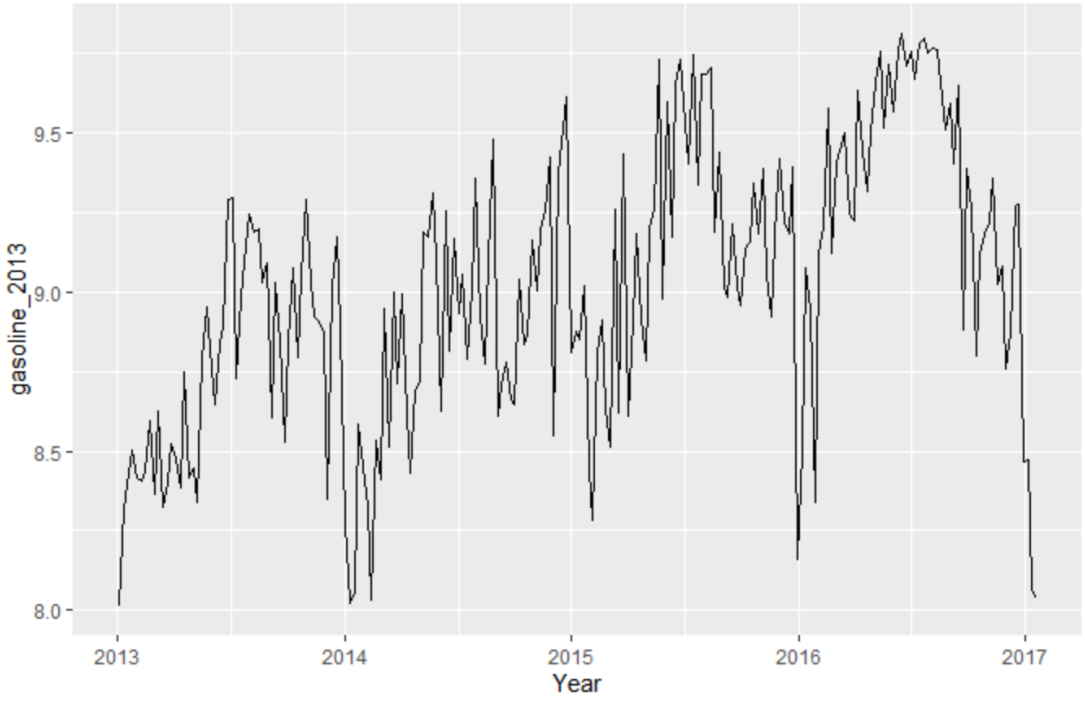
3. Statistical Methods: A discussion and justification of the methods you have used to analyze



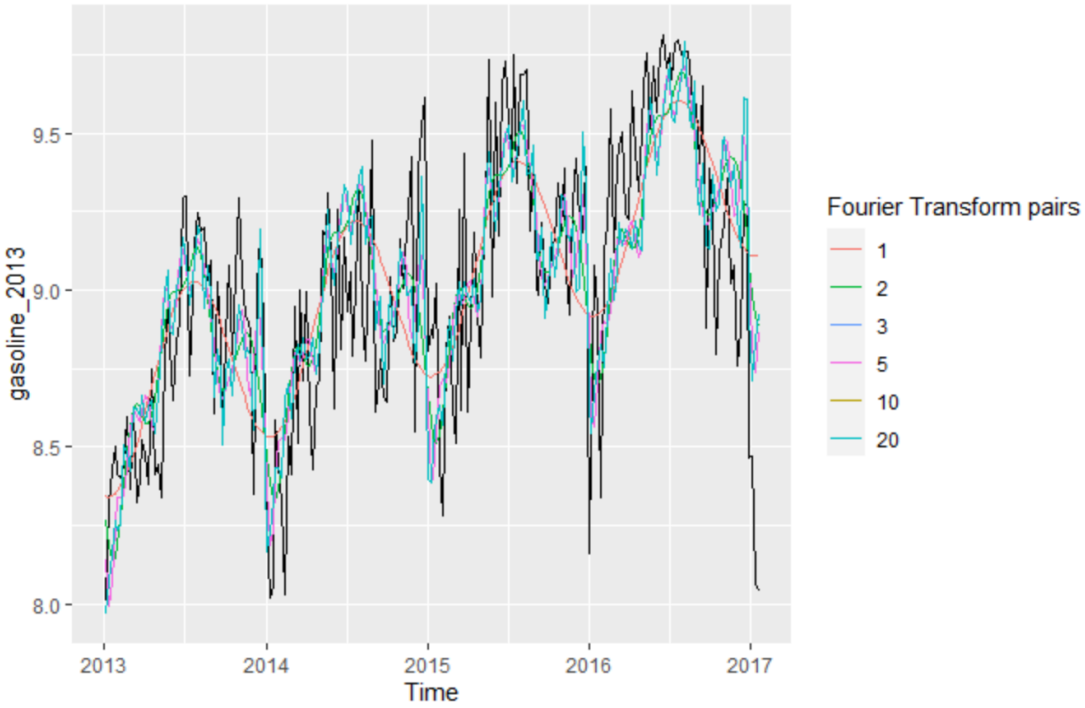
It looked like this data can be divided into 3 parts. First part is increase from the 1991(start) to June, 2007. Second part is decrease from July, 2007 to the end of 2012. Third part is increase from 2013 to 2017(end).

Therefore, two models, the harmonic regression method, and the ARIMA method can be applied. The harmonic regression method can mimic the last part of the data for its trend seems to be smooth, and the number of fourier terms is selected by minimizing AICc and CV. While the ARIMA model can capture the systematic fluctuations and records the accidental ones by knots, the model is fitted automatically but adjusted manually aiming at making the residuals closer to a white noise.

1) The first method used in analyzing the data is harmonic regression with trend to the last part of the data (2013 to 2017). The observed values are plotted as follows.



Firstly, from experiment with changing the number of Fourier terms as follows,



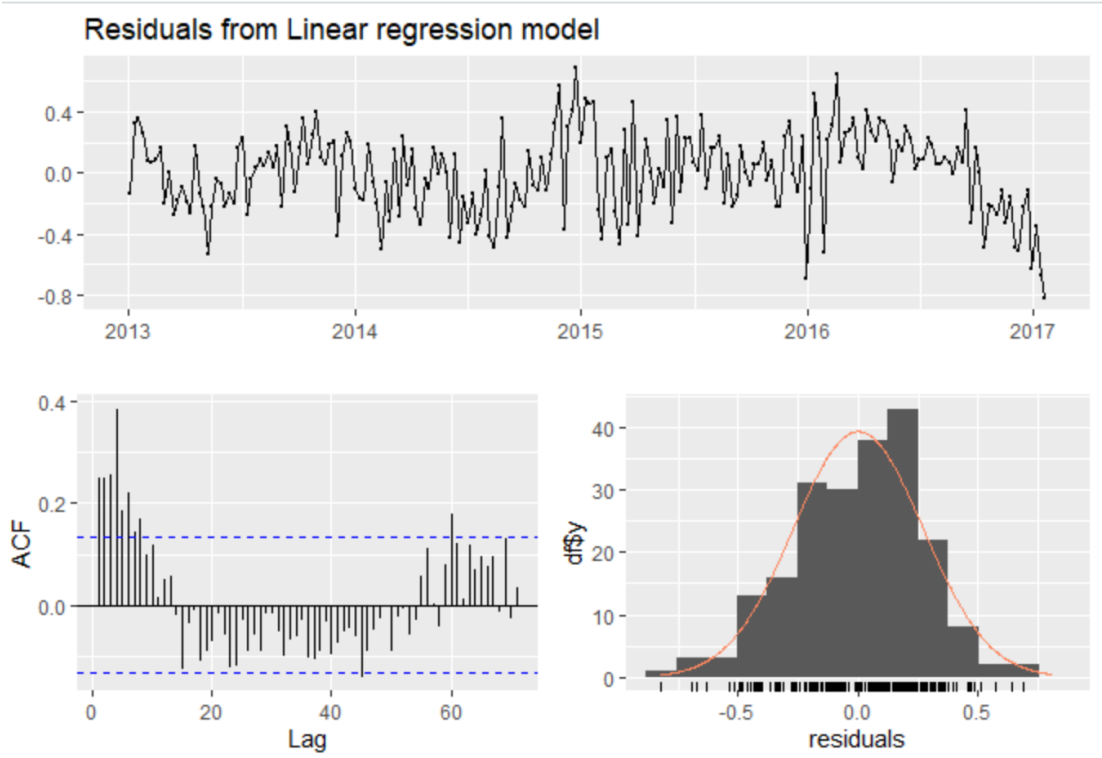
The more number of Fourier pairs used, the fitted line looks more like the original data.

Secondly, the appropriate number of Fourier ters to include is selected by minimising the AICc or CV value.

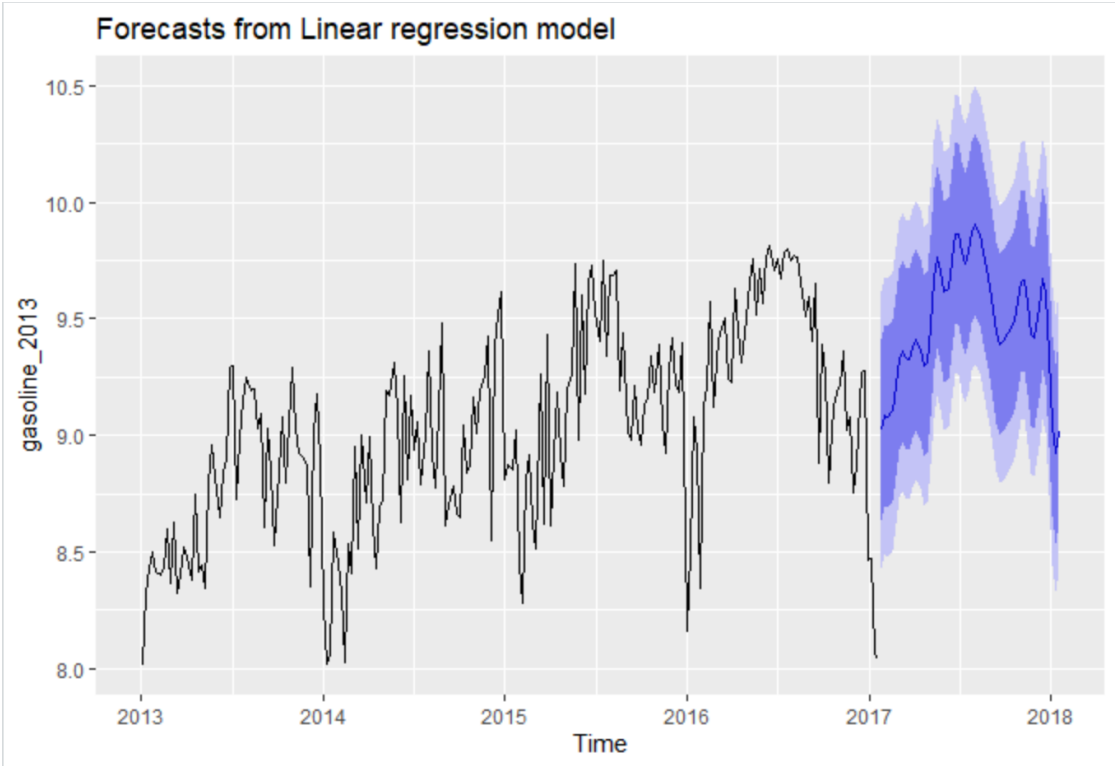
```
fit1_gasoline_2013

          CV           AIC           AICc           BIC         AdjR2
  0.09918017 -488.12242830 -487.83116616 -471.33949692    0.49414105

fit2_gasoline_2013

          CV           AIC           AICc           BIC         AdjR2
  0.09846553 -489.85210962 -489.30309001 -466.35600570    0.50284941

fit3_gasoline_2013

          CV           AIC           AICc           BIC         AdjR2
  0.09162275 -505.10609883 -504.21500972 -474.89682236    0.54156045

fit5_gasoline_2013

          CV           AIC           AICc           BIC         AdjR2
  0.09034925 -508.25325394 -506.41487010 -464.61763237    0.55634370

fit10_gasoline_2013

          CV           AIC           AICc           BIC         AdjR2
  0.09043565 -509.23796544 -503.36562502 -432.03648112    0.57700559

fit20_gasoline_2013

          CV           AIC           AICc           BIC         AdjR2
  0.09958119 -494.84621265 -472.32240312 -350.51300284    0.58103349
```

Since the fourier term of 10 minimize AICc and has a second smallest CV, the final model is a harmonic regression with 10 fourier terms. Also, by checking the results, Even though the residuals fail the correlation tests, the results are probably not severe enough to make much difference to the forecasts and prediction intervals. (Note that the correlations are relatively small, even though they are significant.)

```
> checkresiduals(fit10_gasoline_2013)

        Breusch-Godfrey test for serial correlation of order up to 42

data:  Residuals from Linear regression model
LM test = 68.496, df = 42, p-value = 0.006043
```



Thirdly, the harmonic regression model is used to produce a prediction.

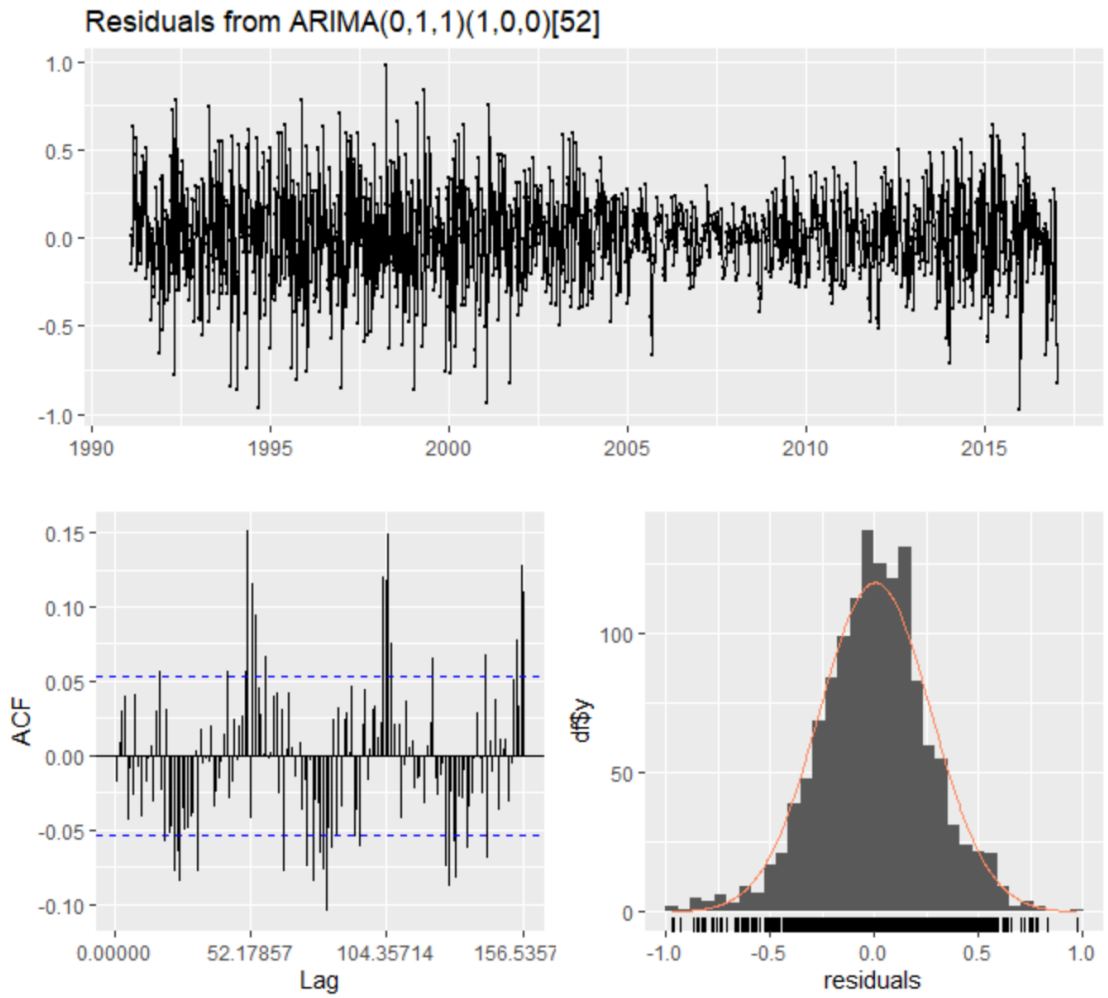


2) The second method to be used is ARIMA modeling.

Firstly, the first knot is set as 2007.5 and second knot as 2013.

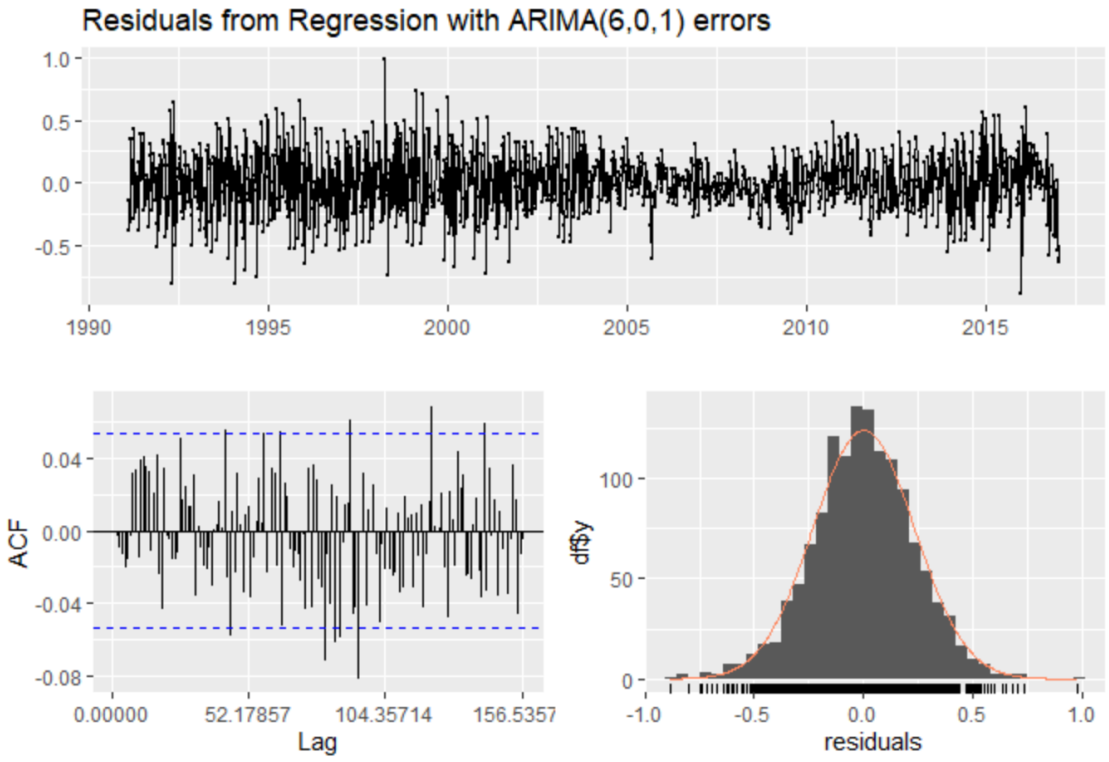
fitted values are somewhat similar to the data.

Secondly, an ARIMA model is automatically fitted and the residuals are checked. However, the residuals aren't like white noise.

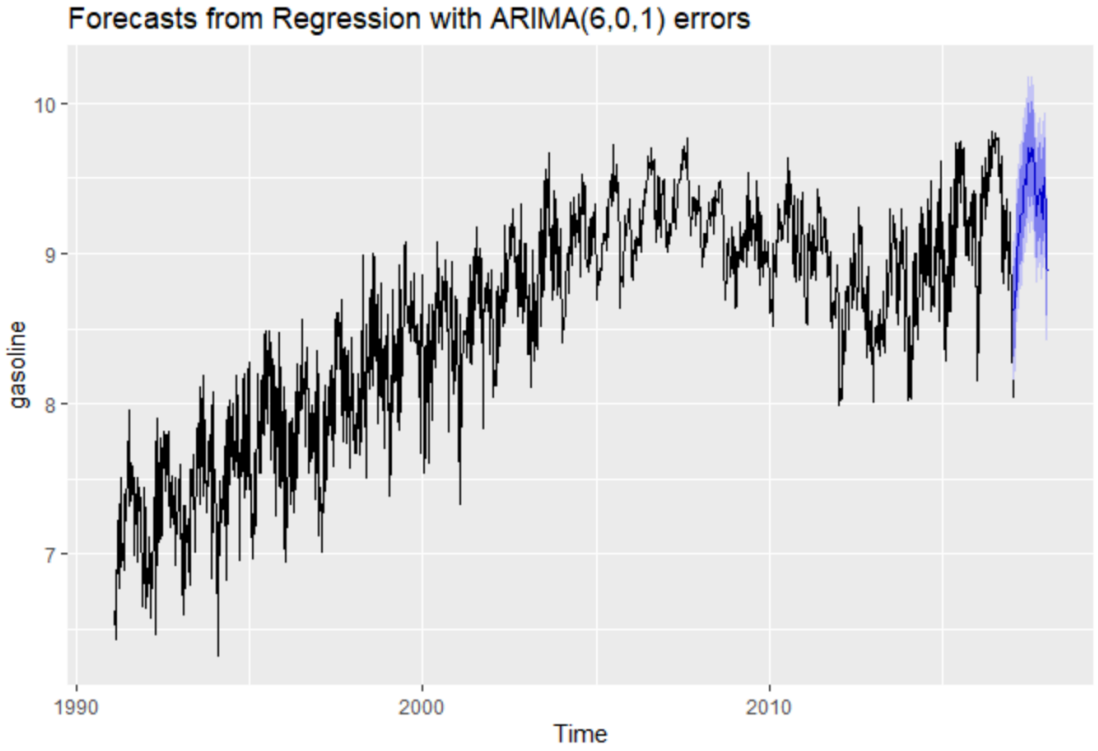


And by using more autoregressive terms, the residuals like white noise (even if the likelihood will be worse).

```
> checkresiduals(gasoline_arima.6.0.1)

        Ljung-Box test

data:  Residuals from Regression with ARIMA(6,0,1) errors
Q* = 137.23, df = 71.357, p-value = 4.597e-06

Model df: 33.   Total lags used: 104.357142857143
```



Thirdly, the harmonic regression model is used to produce a prediction.



4. Results:

   From a statistical perspective, the harmonic regression model and ARIMA model seems to produce similar results, specifically, increasing-decreasing trend, smooth increase and fluctuating decrease. Therefore, practically, both harmonic regression model on part of the

data and ARIMA model on the whole data reveals that, during the first half of the coming year, there is supposed to be a rise of gasoline supply, and the increasing trend can be smooth or with some slight jumps without fluctuations. Also, during the following second half of the year, both models show that gasoline supply is likely to fall, although it tends to be some ups and downs in the last quarter, the overall trend is decreasing at large.

However, from the statistical results, the supplies at the beginning of the year 2018 derived from the ARIMA model seem to be much lower than that of the harmonic regression model. Moreover, according to the harmonic regression model, the peak of gasoline supplies in 2018 may exceed the highest amount of the same period in 2017, while the ARIMA model provides a top value that is more likely to be smaller than that of 2017. Lastly, with respect to the ending value, the harmonic regression predicts that it will fall back to the amount at the beginning of the year 2018, but the result from the ARIMA models is to be lower. The contradictions shall be paid attention to in a practical manner.

5. Discussion

On model fitness, the harmonic regression is restrained to only the last part of the data (2013-2017), although it may capture the recent trend, but the residuals fail the correlation tests, which means that they are significant. The problem may be solved by making more adjustments to the original harmonic regression model.

On the other hand, the ARIMA regression tries to mimic the situation with more data at hand, but may still suffer from overwhelmingly aggressiveness. The problem may be solved by including more knots or using less autoregressive terms, to keep a balance between white noise-liked residuals and number of autoregressive terms.

On the perspective of reality, as the oil supplies is determined by not only short-term fluctuations but also long-time impacts. As a result, when predicting future values, past factors that may leave profound effects shall be taken into consideration.

With respect to final conclusions, quantitatively, gasoline supplies are likely to increase smoothly with no more than some small jumps, until the peak in the middle of 2018. Afterwards, the supplied amount likely to fall with fluctuations, probably occurs in the last quarter, back to the amount at the beginning of the year or even lower.

Qualitatively, the supplies at the beginning of 2018 is 90% likely to have a sharp jump from the ending of 2017 for at least larger than 5%. Also, the peak of the gasoline supplies in 2018 is 80% likely to be only about 5% larger or smaller than the same period of 2017. Lastly, the supplied amount of gasoline at the ending of 2018 is also 90% likely to be at least 5%higher than the counterpart of 2017.

Reference:

Erin O'Neill, (2012). *'Gas out' boycott in 1997 pushed gasoline prices down 30 cents overnight,*

according to Facebook posts. Retrieved December 22, 2021. From https://www.politifact.com/factchecks/2012/apr/15/facebook-posts/gas-out-boycott-1997-pushed-gasoline-prices-down-3/

L.A. TIMES ARCHIVES, (2002). *Gas Prices Drop Half a Cent Despite Global Tensions.* Retrieved December 22, 2021. From https://www.latimes.com/archives/la-xpm-2002-apr-22-fi-gas22-story.html

U.S. News Reuters Staff, (2007). *U.S. gasoline prices rose in 2007: survey.* Retrieved December 22, 2021. From https://www.reuters.com/article/us-energy-gasoline-retail-idUSN0612178120071223

Steve Hargreaves, (2013). *Why gas prices are rising.* Retrieved December 22, 2021. From https://money.cnn.com/2013/02/06/news/economy/gas-prices/index.html