Gas Price Forecast with Time Series Models

Rui Cao, Xiaoyun Wang, Rong Zhang

May 2019

Abstract

In this project, we built times series models to forecast weekly regular gasoline prices in East Coast. We used backtest method to compute the mean squared error (mse) to assess the accuracy of prediction of the model. We applied different window length to find the optimal amount of data to be fit in each model. We also implemented Monte-Carlo simulation to answer the statistical question that if correct order specification is important in forecast accuracy when we fit the model with different amount of data. Finally, we selected the best model among AR, ARIMA and GARCH to forecast the gas price.

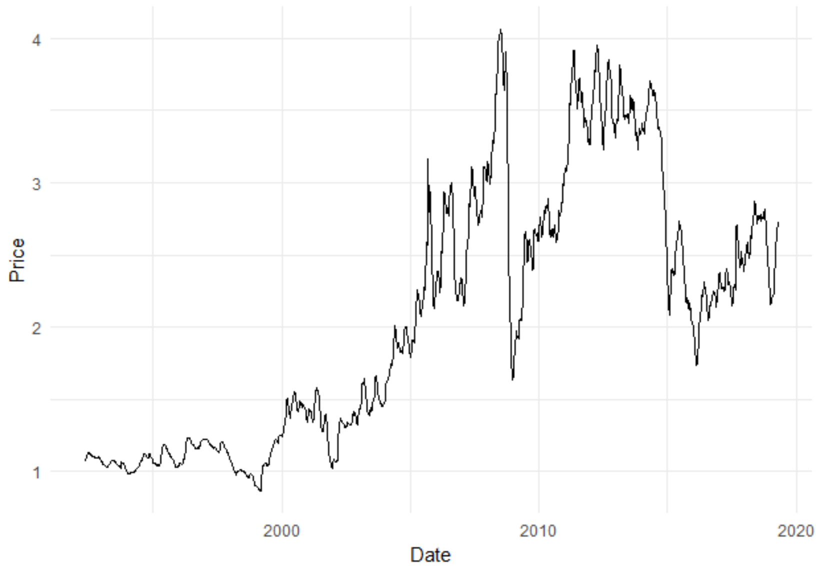
1 Introduction

Various statistical methods are often used in the business world in order to forecast future development. Companies today use everything from simple spreadsheets to complex financial planning software to attempt to accurately forecast future business outcomes such as product demand, resource needs, or financial performance. These tools build forecasts by looking at a historical series of data, which is called time series data. For example, such tools may try to predict the future sales of a boots by looking only at its previous sales data with the underlying assumption that the future is determined by the past. In our project, we used time series models to predict future gas price based on previously observed gas price.

2 Data Description and Assumptions

2.1 Data Description

The data we used is the weekly retail gasoline and diesel prices downloaded from the website of U.S. Energy Information Administration. The website has all the data of gas price from 1992 to 2019, and the plot of the prices is shown below.



year.png

Figure 1: Gas Price of year from 1992 to 2019

Since there is a big change before the year 2010, and we believe that the big change is a rare event, we used only the data after 2010 to make prediction. The plot of the price after 2010 is shown below.

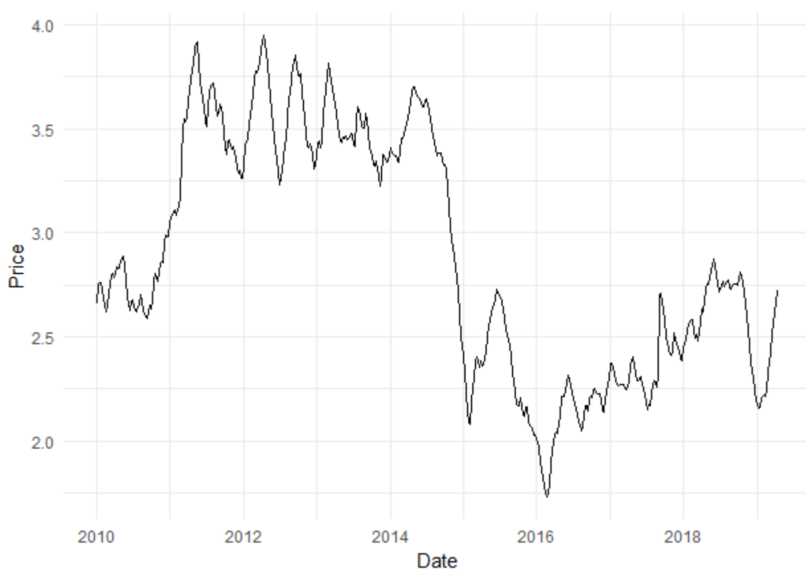


Figure 2: Gas Price of year from 2010 to 2019

Also, we considered the inflation effect through the years. We adjusted the prices based on the inflation rate and the plot of the adjusted price after 2010 is shown below. This is the final data we used in our analysis.

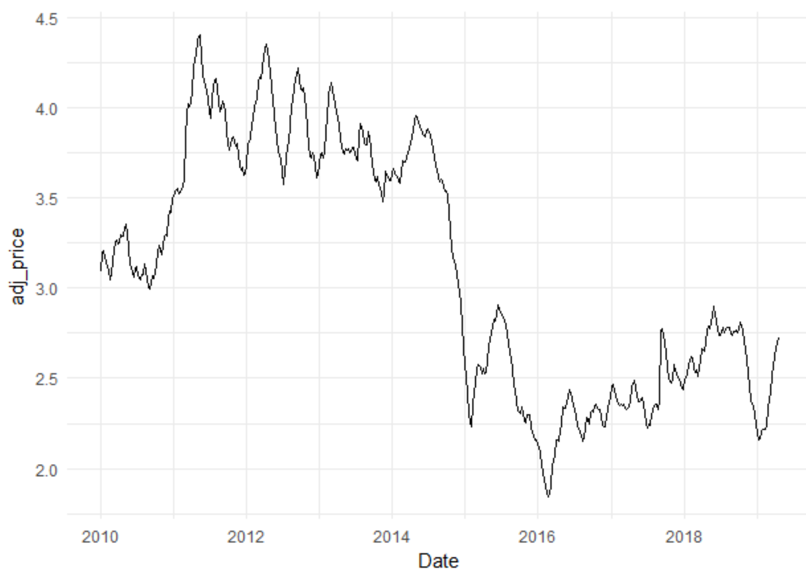


Figure 3: Adjusted Price from 2010 to 2019

2.2 Assumptions

2.2.1 Normality

First, we want to check the normality of the gas price. We used Q-Q (quantile-quantile) plot to compare the quantiles of the gas price to the normal distribution. The plot is shown below.

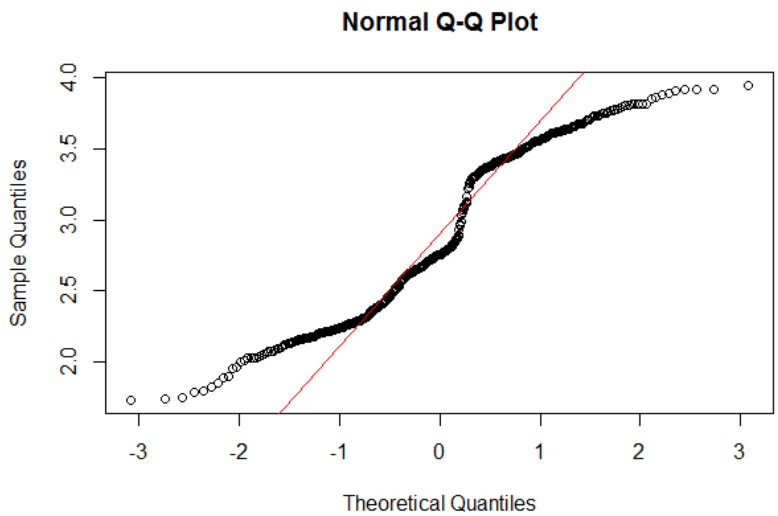


Figure 4: Q-Q Plot of Gas Price againt Normal Distribution

From figure 4, we can see that the quantiles of the gas price are not in consistency with the line of normal distribution. Therefore, the gas price does not strictly follow the normal distribution.

2.3 Autocorrelation

Next, we need to check if the data has autocorrelation, because it is an important assumption for us to apply time series models. We plot the autocorrelation function (ACF) and partial autocorrelation function (PACF) to check the significance of autocorrelation.

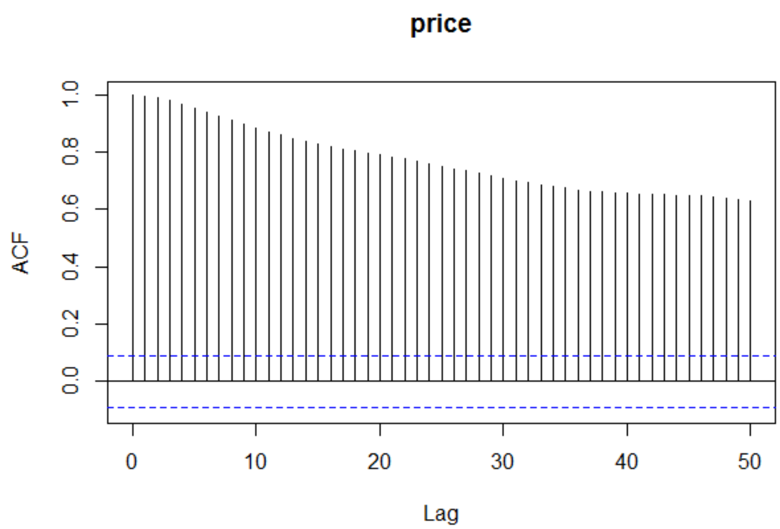


Figure 5: ACF Plot

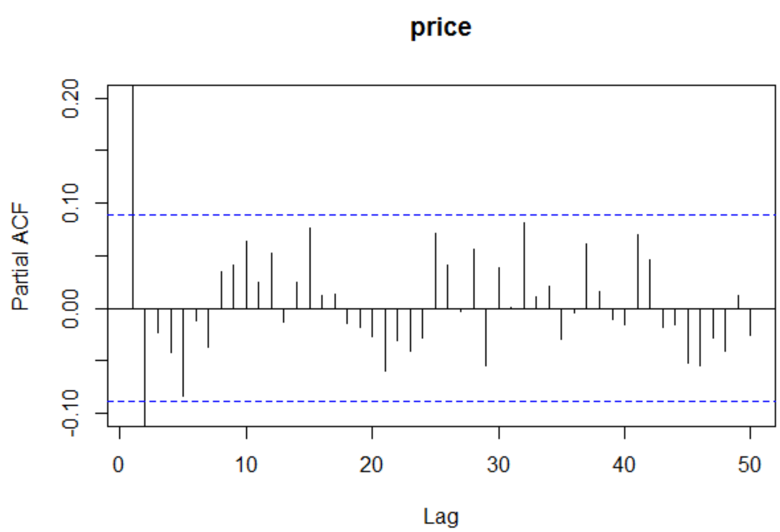


Figure 6: PACF Plot

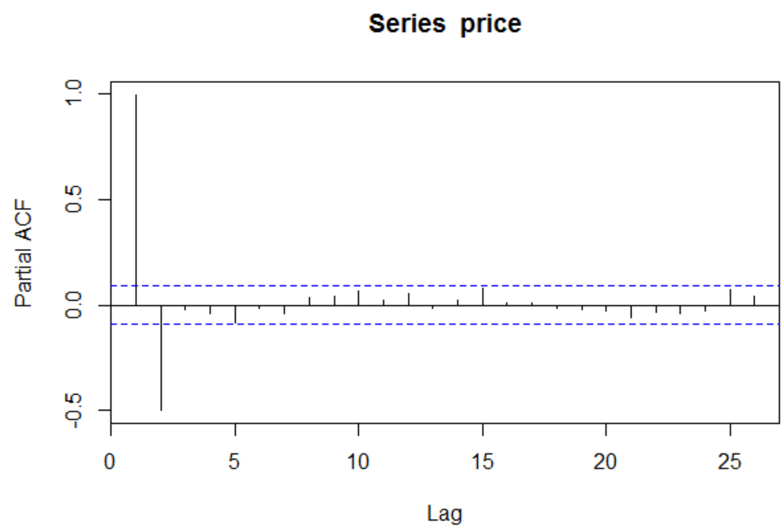


Figure 7: PACF Plot on a larger scale

From figure 5, we can see that up to lag 50, the autocorrelation is still strong. Figure 6 and figure 7 show partial autocorrelation function (PACF) of the data on different scales. Note that in the PACF plots, there are significant spikes at lag 1 and lag 2, meaning that all the higher-order autocorrelations are effectively explained by the lag 1 and lag 2 autocorrelation. The plots of ACF and PACF can help us identify the order of time series models that are most suitable for the data.

3 Models

3.1 ARMA Model

An autoregressive (AR) model is a representation of a type of random process.The autoregressive model specifies that the output variable depends linearly on its own previous values and on a stochastic term (an imperfectly predictable term); thus the model is in the form of a stochastic difference equation. While the AR model involves regressing the variable on its own lagged (i.e., past) values, the moving average (MA) model involves modeling the error term as a linear combination of error terms occurring contemporaneously and at various times in the past.

The notation AR(p) refers to the autoregressive model of order p. The AR(p) model is written

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-1} + \epsilon_t$$

where ϕ_1,...,ϕ_p are parameters, c is constant, and the random variable ϵ_t is white noise.

The notation MA(q) refers to the moving average model of order q:

$$X_t = \mu + \sum_{i=1}^{q} \theta_i \phi_{t-i} + \epsilon_t$$

where θ_1,...,θ_q are parameters, μ is the expectation of X_t (often assumed to equal 0), and the ϵ_t,ϵ_{t-1}... are white noise error terms.

Adding AR(p) and MA(q) together will give us ARMA(p,q) model:

$$X_t = c + \epsilon_t + \sum_{i=1}^{p} \phi_i X_{t-1} + \sum_{i=1}^{q} \theta_i \phi_{t-i}$$

3.2 ARIMA Model

An autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. It includes the technique of differencing. Differencing removes the changes in the level of a time series, eliminating trend and seasonality and consequently stabilizing the mean of the time series. In order to difference the data, the difference between consecutive observations is computed. The differenced data is then used for the estimation of an ARMA model. In ARIMA (p,d,q) model, the order of differencing is specified as d.

3.3 GARCH Model

The autoregressive conditional heteroskedasticity (ARCH) model describes the variance of the current error term or innovation as a function of the actual sizes of the previous time periods' error terms. If an autoregressive moving average (ARMA) model is assumed for the error variance, the model is a generalized autoregressive conditional heteroskedasticity (GARCH) model.

Let ϵ_t denote the error terms. These ϵ_t are split into a stochastic piece z_t and a time-dependent standard deviation σ_t characterizing the typical size of the terms so that

$$\epsilon_t = \sigma_t z_t$$

The random variable z_t is a strong white noise process. In GARCH(p,q) model, the series σ_t^2 is modeled by

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

where $\omega > 0$, $\alpha_i \geqslant 0$ and $\beta_i \geqslant 0$.

4 Simulation Study

Monte Carlo are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results. Monte-Carlo simulation is also called random sampling or statistical testing method.

In this project, we simulated a set of data that looks similar to the actual gas price data. The parameters were selected according to the parameters estimated in ARIMA (2,1,1) model fit with the actual gas price. The function used and the parameters are shown below:

arima.sim(order=c(2,1,1),ar=c(0.319,0.0741),ma=0.2749), sd=0.05,n=499)

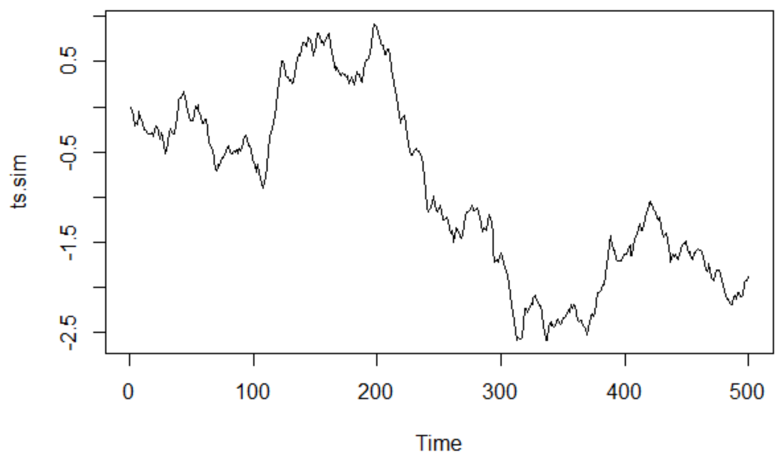


Figure 8: Simulated Data Example 1

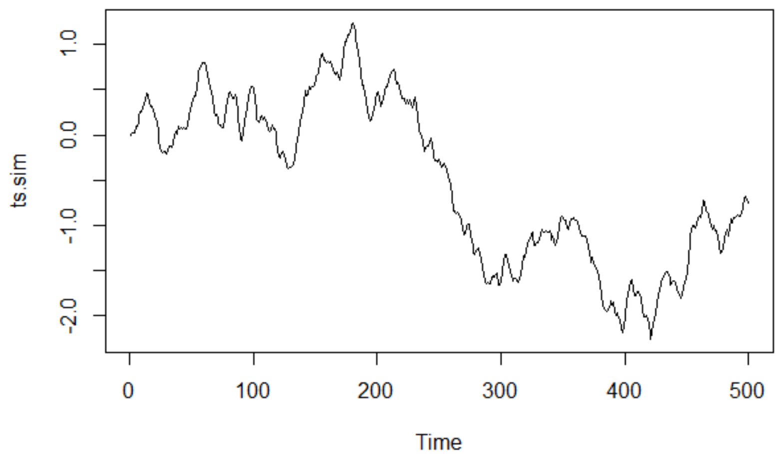


Figure 9: Simulated Data Example 2

In Monte-Carlo simulation, we simulated 1000 samples of sample size 500. Figure 8 and figure 9 show two examples of our simulated data which look similar to the actual price in terms of the trend and range. By taking Monte-Carlo simulation, we want to find suitable window length of data and model parameters.

For time series model, it is important to choose an appropriate window length, which ensures that the data distribution is stable. In figure 10, we vary the window length from 10 to 400 and compare the AR model with different orders. The result shows that with order = 1, the mean squared error between predicted data and simulated data is bigger comparing with higher orders. It means that the prediction of next point should not be only dependent on single one past point but at least 2 points. Another argument from Figure 10 is that the mean squared error decreases with window length for all orders. Statistically more data training can make the model more robust, so the model can be more generative to all situations. This explains why longer window length is preferable.

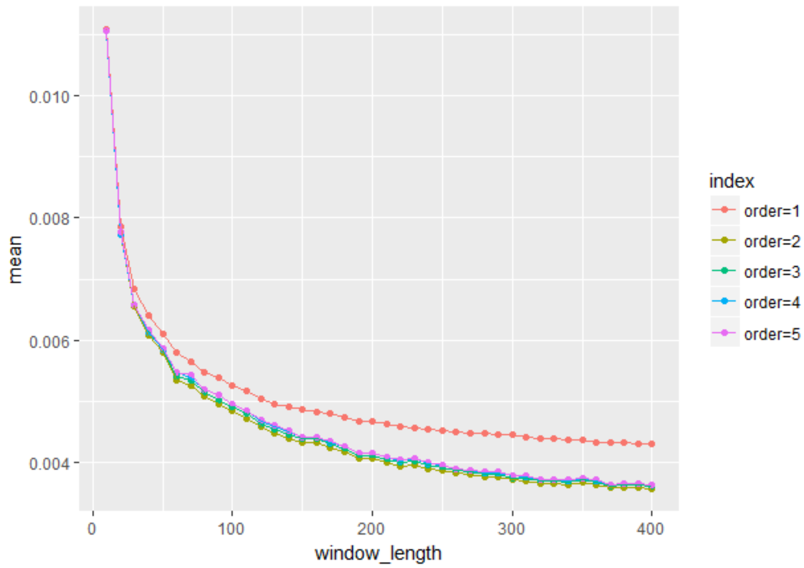


Figure 10: MSE Comparison among different AR orders

With the tuned parameters of AR, we would like to compare the performance between AR and auto.arima. Figure 11 shows the comparison of the mean squared error between AR(2) and auto.arima. The AR(2) performance is consistently with that of figure 10, longer window-length has better performance. But it is interesting to see that auto.arima always has a stable mean squared error which is lower than AR(2). Therefore, we can say that auto.arima is better than simple AR. The reason we conclude is that auto.arima is flexible to modify its parameters to match the original data better.

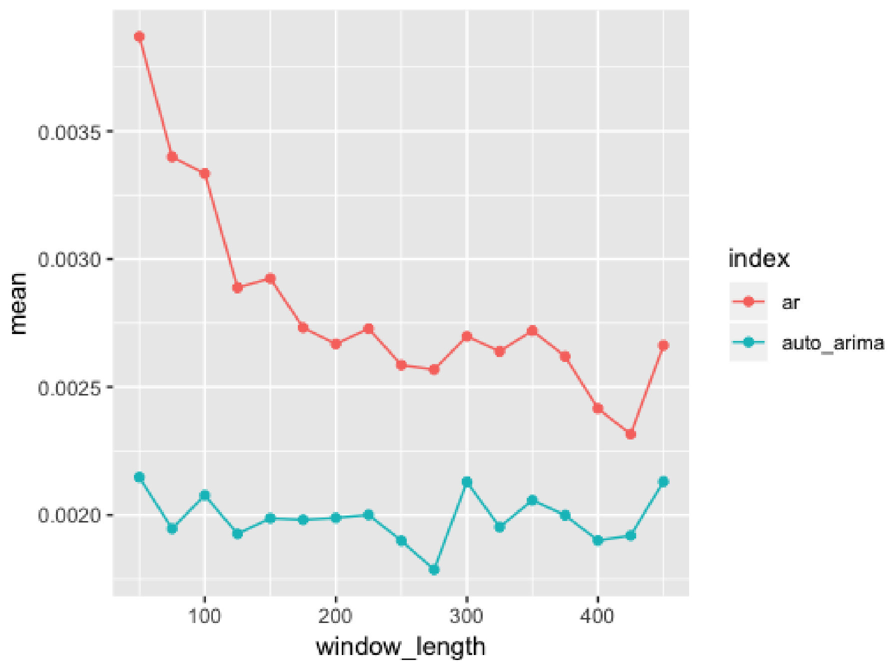


Figure 11: MSE Comparison between AR(2) and auto.arima

To further verify if auto.arima can grab the true distribution of data. We compare the parameters produced by auto.arima with the true parameters (2,1,1). During simulation, we count the frequency when the parameters produced by auto.arima is equal the the true parameters. The result is shown in figure 12. It is easy to see that auto.arima can hardly produce parameters which are equal to true parameters. But auto.arima can still get very good predictions, the reason we conclude is that correct order specification is not really important.

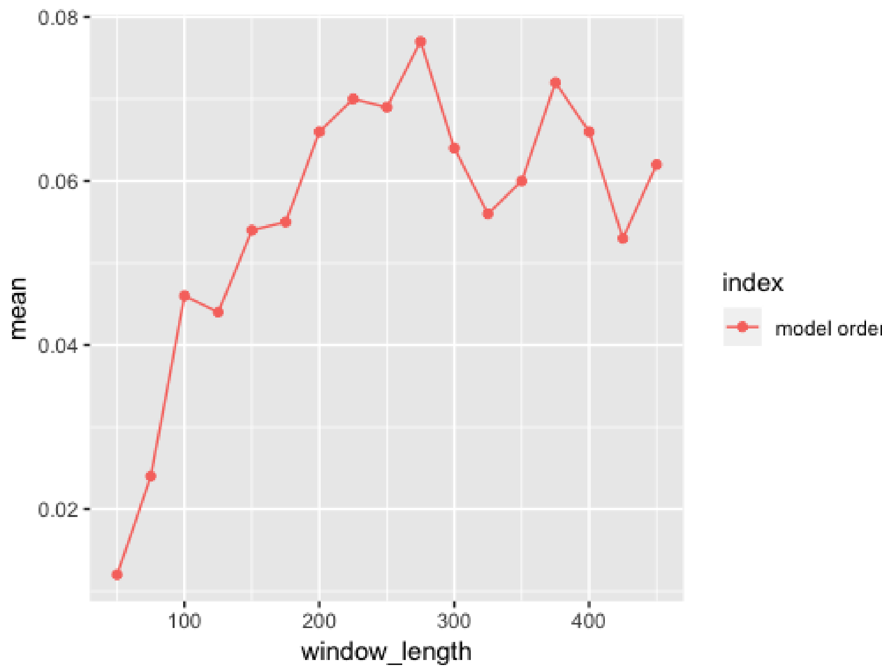


Figure 12: Order Consistency of auto.arima with ARIMA(2,1,1)

5 Comparison between Models

In the GARCH specification, we used ARMA to model the conditional mean of the error terms and used GARCH to model the conditional variance of the error terms. In order to find the optimal order in the specification, we first fixed the order of GARCH to be (1,1) and changed the order in ARMA to make comparison. Figure 13 shows the mean squared error of the AR models of order 1 to 4 with window length from 400 to 480. We can see that AR(2) gives the lowest MSE and 440 is the optimal window length.

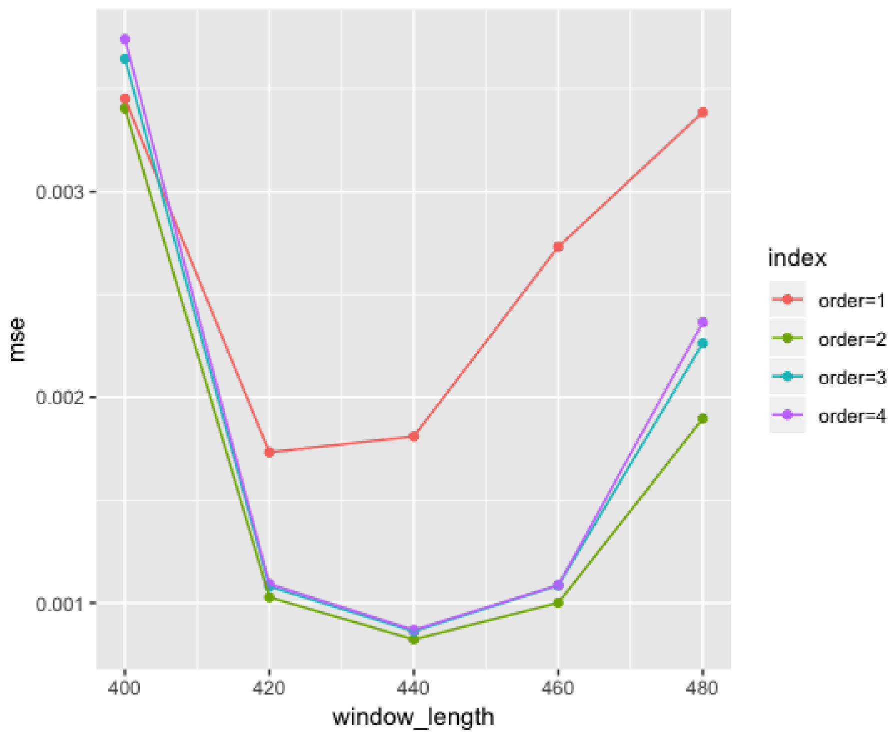


Figure 13: MSE Comparison among different AR orders with GARCH(1,1)

Then we fixed the ARMA order to be (2,0) and changed the GARCH order to find the optimal GARCH order. The GARCH orders we have tried are listed in the legend. However, from figure 14, we can see that the lines superimpose on one another, which means that, changing the GARCH order does not make much difference in the accuracy of prediction. Therefore, the basic model GARCH(1,1) is our choice because it is simple and gives good prediction.

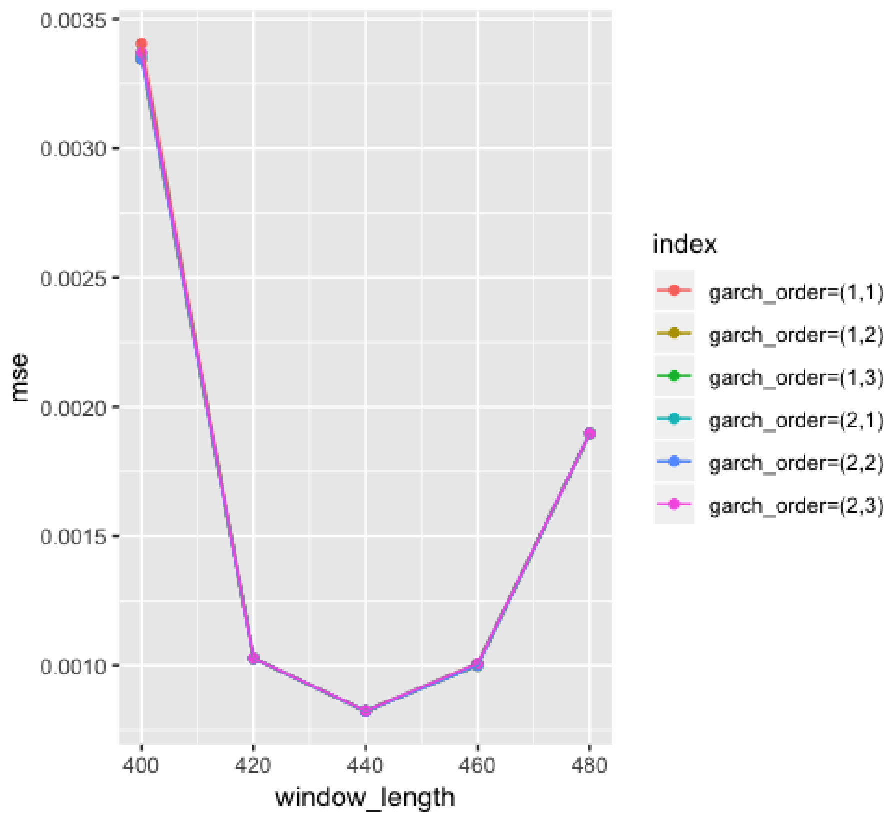


Figure 14: MSE Comparison among different GARCH orders with AR(2)

We also did a brief comparison between ARIMA(2,1,1) and ARIMA(2,1,2) to check the effect of MA order. The MSE of the two models are shown in figure 15 with window length from 400 to 480. We can see that the difference in MSE between the two models is negligible, so it is okay to keep the MA order as 0 or 1 for simplicity.

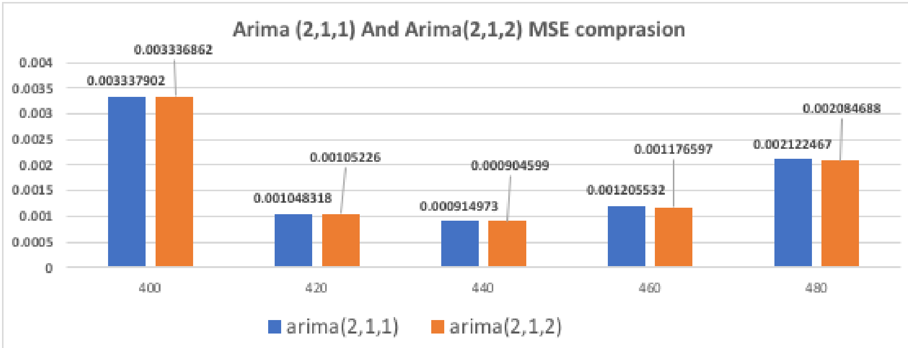


Figure 15: MSE Comparison between ARIMA(2,1,1) and ARIMA(2,1,2).

Finally, we compared the models of GARCH(1,1) with mean model of AR(2), AR(2) and ARIMA(2,1,1). The MSE are shown in figure 16 with window length from 400 to 480. From the graph, we can see that GARCH model gives the lowest MSE at almost all window lengths and achieves its lowest point at window length 440. Generally, ARIMA performs better than AR and GARCH performs better than ARIMA.

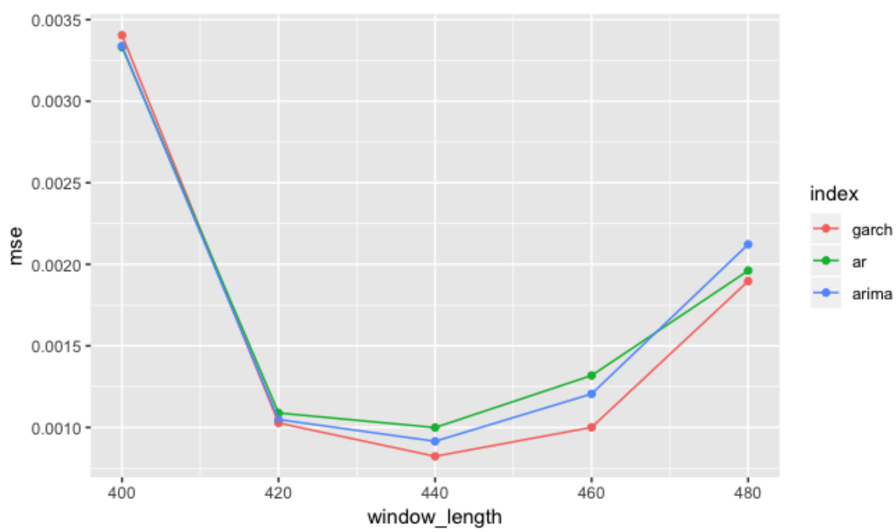


Figure 16: MSE Comparison between GARCH(1,1), AR(2) and ARIMA(2,1,1)

Garch model gives the most accurate prediction as expected. The result matches the methodology, because GARCH model contains the prediction of AR model while it allows the variability of the error terms. However, AR and ARIMA model assume a constant variance of the error terms. Therefore, GARCH model gives more flexibility in fitting the data. ARIMA and GARCH both need large window length to get results (at least 400). AR order has the largest effect, but when fixing AR order, ARIMA and GARCH do improve the result. But it does not show much difference changing the other parameters when fixing AR part. Therefore, ARIMA(2,1,1) and GARCH(1,1) are good enough basic models to be used, and between them, GARCH(1,1) is the better choice.

6 Conclusion

The report describes the utilization of different time series models for prediction of the volatility of future gas prices. We used PACF as well as backtest method to determine the optimal order of the time series models. We explored AR, ARIMA and GARCH models and compared the accuracy of prediction of these models by computing the MSE in rolling windows. The best window length to fit the model is 440. After comparisons, we found that GARCH model is the best model that gives the least mean squared error.

References

[1] R.Dalinina, *Introduction to Forecasting with ARIMA in R*, 2017.

[2] M. Bohdalova, R. Bohdal, V. Valach, *Short Term Prediction of Gas Prices Using Time Series Analysis*, 2016

[3] C. Chatfield, *The Analysis of Time Series*

[4] J.D. Cryer,K. Chan, *Time Series Analysis with Application in R*

[5] R. F. Engle, *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation*, Econometrica, Vol. 50(4), pp. 987-1007, Jul. 1982

[6] T. Bollerslev, *Generalized Autoregressive Conditional Heteroskedasticity*, Journal of Econometrics, 31, 307-327, 1986

[7] V. Onado, *Application of ARIMA models to forecast the quality of a french cheese: The 'Comte*, International Journal of Bio-Medical Computing Volume 28, Issue 4, September 1991, Pages 249-267

[8] E. J. Hannan B. G. Quinn, *The Determination of the Order of an Autoregression*,1979

[9] C. G. Lamoureux, W. D. Lastrapes *Persistence in Variance, Structural Change, and the GARCH Model*, Journal of Business Economic Statistics Volume 8, 1990 - Issue 2