STA 137 Final Project

Huanjie Dong, Zhiye Jiang

06 December, 2023

Abstract

This study presents a detailed time series analysis of oil prices from 2004 to 2023. Utilizing time series methodology we learned in STA 137, we initially transformed the data to stabilize variance and achieve symmetry. The analysis involved decomposing the time series into trend, seasonality, and stationary errors. We extensively analyzed the smooth component for trend and seasonality, ensuring the residuals were stationary. These residuals were further checked for whiteness, remaining trends, and normality. For the rough component, we fitted stationary ARMA models. Our predictive model combines forecasts of both smooth and rough components, aiming to provide accurate predictions of future oil prices. This comprehensive approach is designed to enhance the understanding of the oil market dynamics and provide reliable forecasts.

Contents

1 Introduction

In contemporary society, the daily commute of individuals heavily relies on a variety of vehicular modes. Among these, gasoline-powered vehicles are predominantly favored by the populace of the United States. Consequently, the cost of gasoline bears a significant correlation with the routine life of the average citizen.

This study endeavors to conduct a comprehensive analysis of the time series data concerning oil prices spanning from 2004 to 2023. The primary objective is to gain a profound understanding of the dynamics of the oil market in relation to temporal factors. Additionally, this analysis aims to develop a viable predictive model for future oil prices, based on the insights derived from historical data trends and patterns.

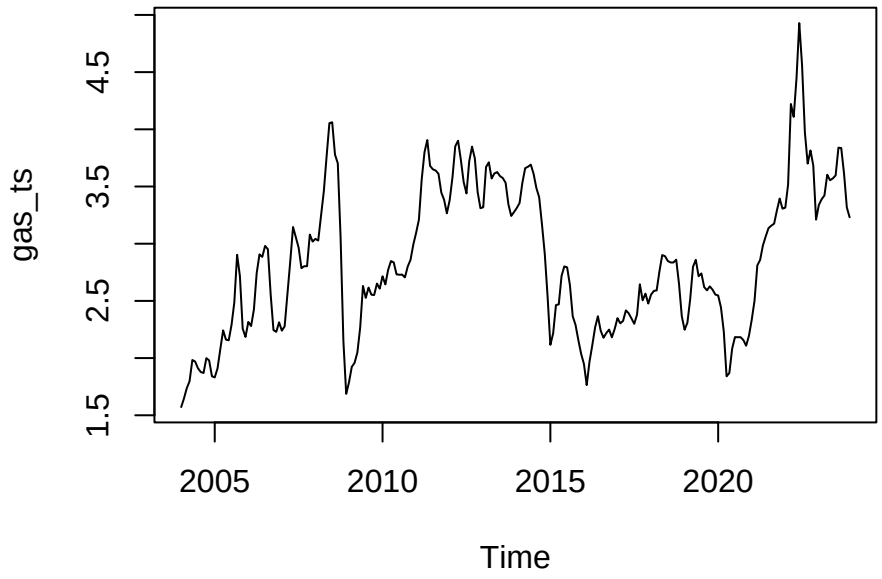
Subsequently, we focus on the rough component, applying ARMA models to the refined residuals and ensuring that all dependencies are captured for accurate modeling. This dual analysis of smooth and rough components allows for a comprehensive understanding of the time series dynamics.

Furthermore, this paper incorporates spectral analysis to investigate the presence of periodic behavior in the residuals of the time series data. This analytical approach is employed with the intention of identifying any cyclical patterns that may exist within the fluctuations of oil prices, thereby enhancing the robustness and accuracy of the proposed predictive model.

The final phase of our study involves predicting future values of oil prices, a task of significant practical importance. By combining predictions from both smooth and rough components, we aim to provide a reliable forecast that encapsulates both the long-term trends and short-term fluctuations in oil prices.

We hope this paper provides our audience a better understanding of what drives oil prices. By carefully analyzing data and using STA 137 methods, we're trying to explain how the oil market works and what might happen with oil prices in the future.

2 Data Description



The data contains monthly average oil price (U.S. Energy Information Administrative) in United States from January 2004 to December 2023.The interesting fact is that the gas price follows the economic cycle, when the economy is going down, the gas price is going down too, and when the economy is blooming, the gas price will go up. And during the covid, the gas price dropped a lot at first after a long time increasing but then it started to increase again. We can see from the plot it do follows a large economic cycle and a relatively narrower seasonal cycle like when during the sring and summer the gas price is higher.

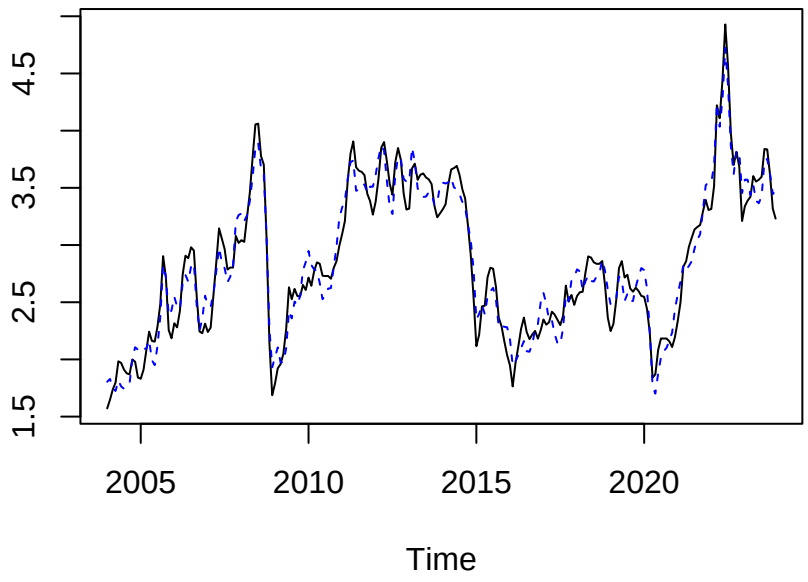
3 Data Analysis

3.1 Deseasonalization

First, we consider to use the 2-sided MA process to deseasonalize the data.

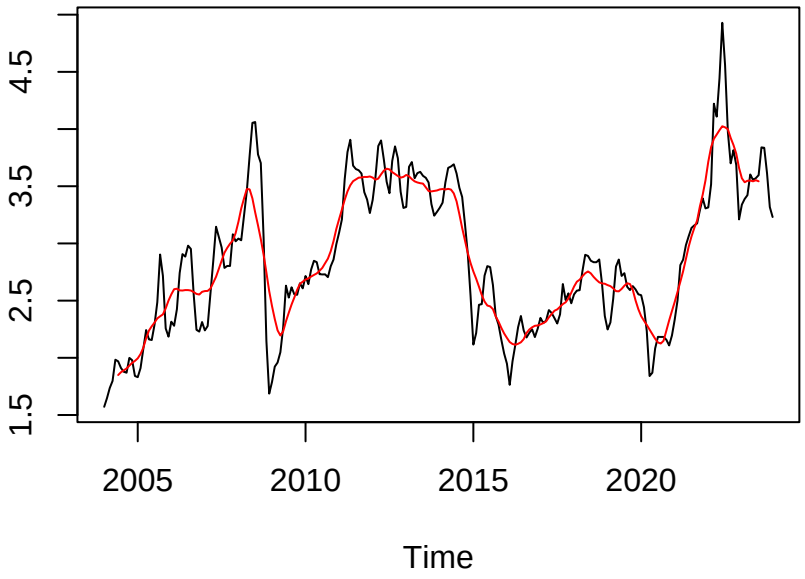
```
##  [1] -0.23278351 -0.18191662 -0.01455697  0.07385103  0.16833020  0.20531222
##  [7]  0.16770355  0.11049588  0.08152472 -0.00697627 -0.12704173 -0.24394250
```

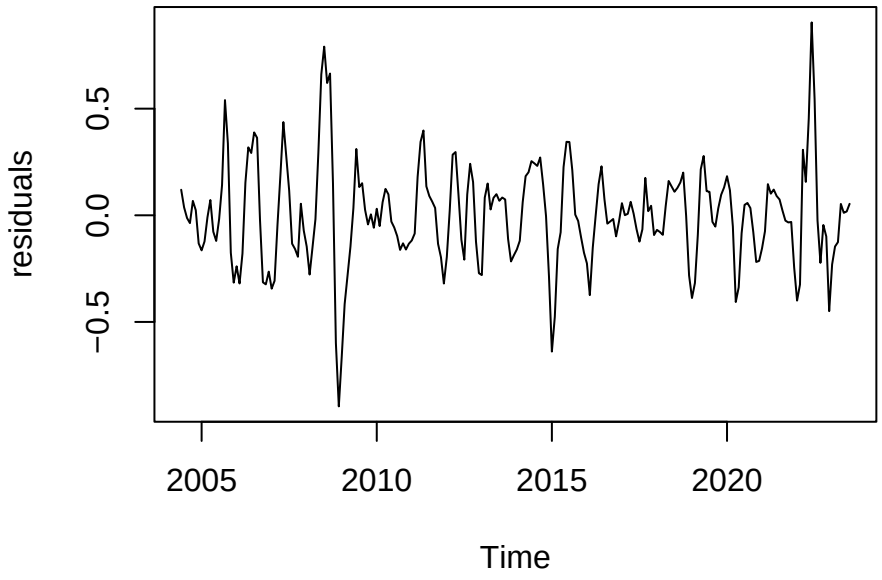
Therefore, we got the seasonality for each month under 2 sided MA process.



3.2 Detrend

Since our dataset contains 240 data points for 20 years monthly data and the gas price is affected by the economic cycle, we want to detrend it by 2 side moving average.





3.3 Augmented Dickey-Fuller (ADF) Test (Stationarity Check)

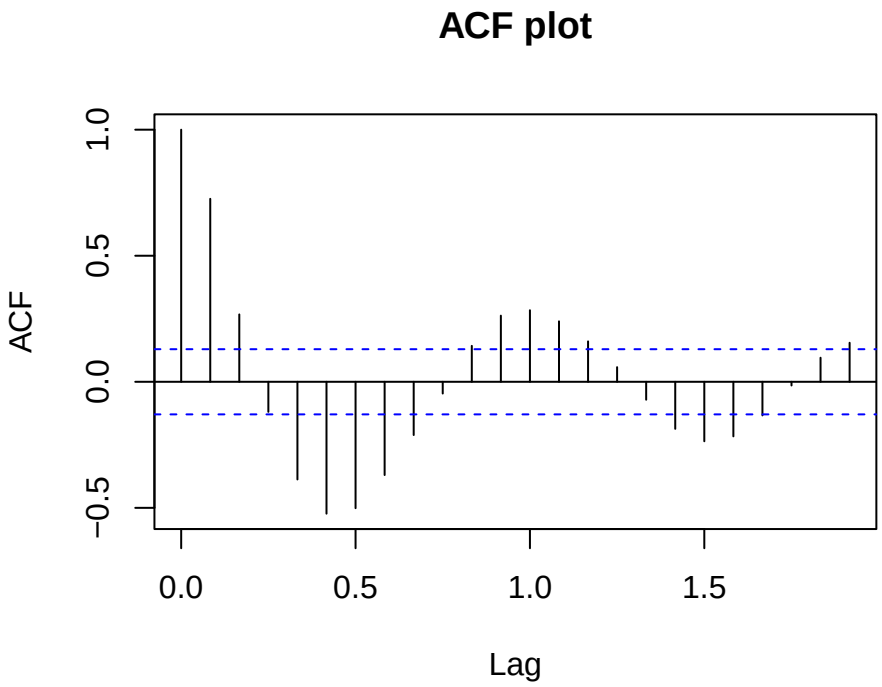
Then we want to check if the residual is stationary or not so we do a dick fuller test on it.

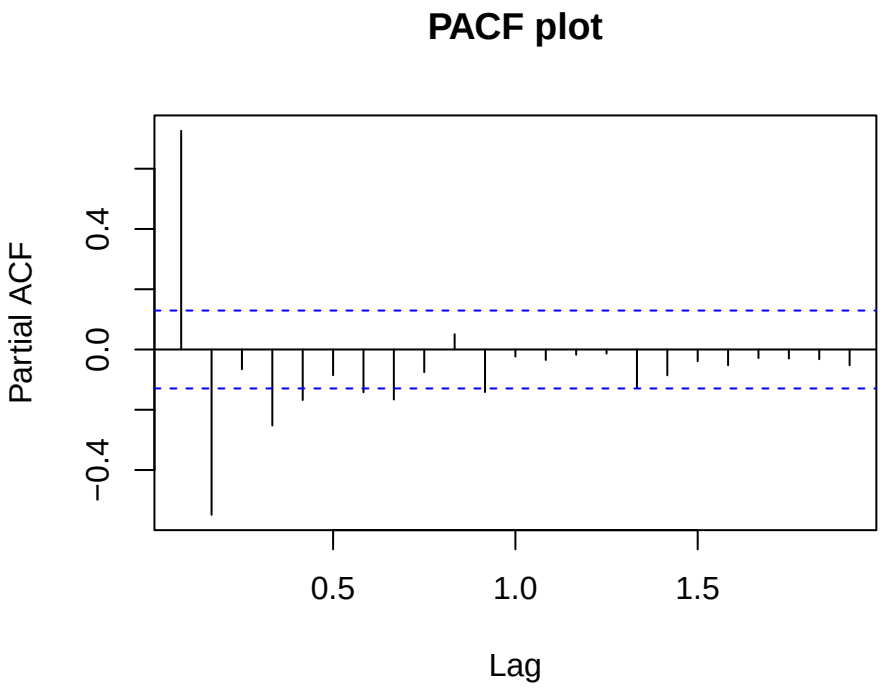
```
## Warning in adf.test(detrended_gas_ts, alternative = "stationary"): p-value
## smaller than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  detrended_gas_ts
## Dickey-Fuller = -8.7146, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
```

Therefore, it it indicates the residual is stationary.

3.4 ACF and PACF Analysis





With the ACF and PACF plot, we choose to use ARMA(2,1) as our best fit model. We will compare it with the best model given by the package as well.

3.5 ARIMA Model

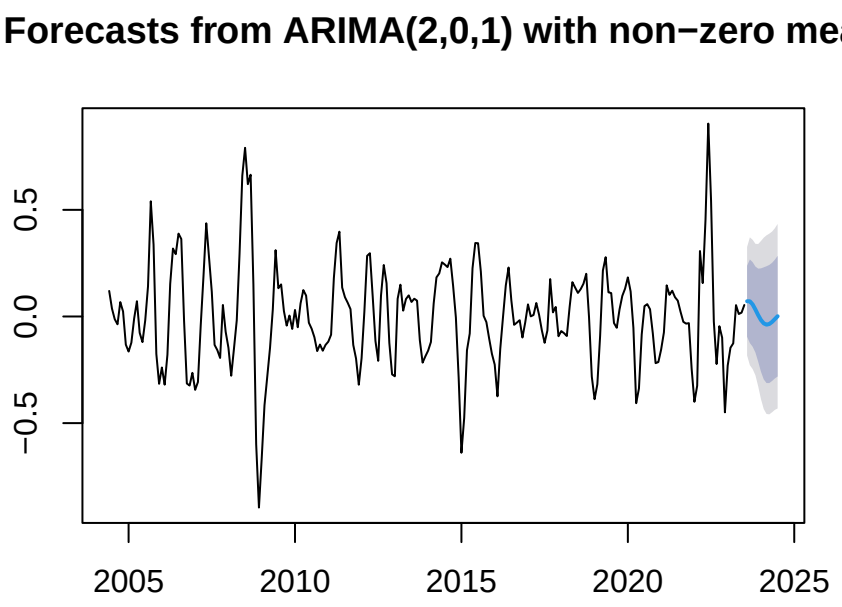
```
## Series: detrended_gas_ts
## ARIMA(2,0,1) with non-zero mean
##
## Coefficients:
##          ar1      ar2      ma1   mean
##       1.6171  -0.8236  -1.0000  2e-04
## s.e.  0.0362   0.0360   0.0219  6e-04
##
## sigma^2 = 0.01679:  log likelihood = 142.63
## AIC=-275.27   AICc=-275   BIC=-258.08
##
## Training set error measures:
##                       ME      RMSE        MAE       MPE     MAPE      MASE
## Training set 0.004902826 0.1284356 0.09932237 -33.84227 180.6819 0.4530804
##                    ACF1
## Training set 0.2307655

## Series: gas_ts
## ARIMA(0,1,1)
##
## Coefficients:
##          ma1
##       0.4854
```

```
## s.e.  0.0529
##
## sigma^2 = 0.02805:  log likelihood = 88.29
## AIC=-172.58   AICc=-172.53   BIC=-165.63
##
## Training set error measures:
##                       ME      RMSE       MAE       MPE     MAPE      MASE
## Training set 0.004670614 0.1667916 0.1237931 0.1154932 4.425351 0.2435614
##                     ACF1
## Training set 0.03118218
```

Therefore, we compared the 2 models, the previous one is based on our preprocessing (deseasonalization and detrend), and the second one is the function provided to automatically choose the model. We found that our model explain the data better and provided a more comprehensive outlook to it. However, the MAPE is much more higher in our ARIMA model, which might be a concern.

3.6 Forecasting the Rough Component



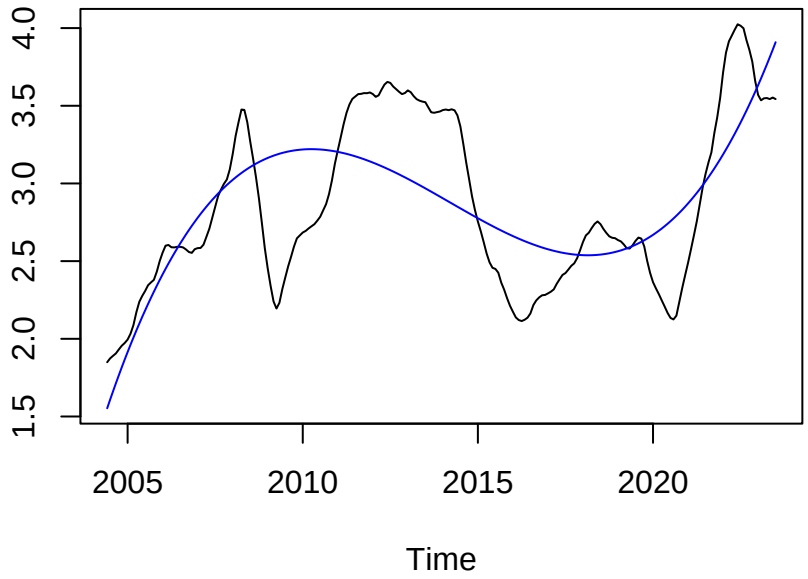
Therefore, we get our predicted value of residuals, we want to check if the price predicted follows the same pattern with the true price. So we need to use the trend and seasonal factors.

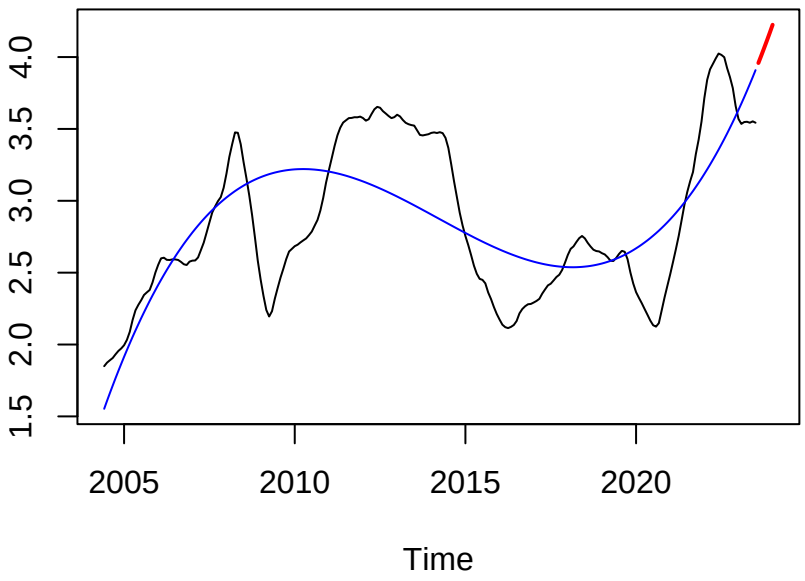
3.7 Forecasting the Real price

We want to use the mean seasonal factors that we got because we assume they do not change in the long term. For the trend component, we will use a linear prediction to predict that in our best effort.

```
##
```

```
## Call:
## lm(formula = ma5 ~ t + t2 + t3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.99022 -0.27797  0.03091  0.28902  0.67011
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.497e+00  1.043e-01   14.35   <2e-16 ***
## t            5.668e-02  3.903e-03   14.52   <2e-16 ***
## t2          -5.703e-04  3.922e-05  -14.54   <2e-16 ***
## t3           1.607e-06  1.116e-07   14.39   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3891 on 226 degrees of freedom
## Multiple R-squared:  0.5126, Adjusted R-squared:  0.5061
## F-statistic: 79.23 on 3 and 226 DF,  p-value: < 2.2e-16
```





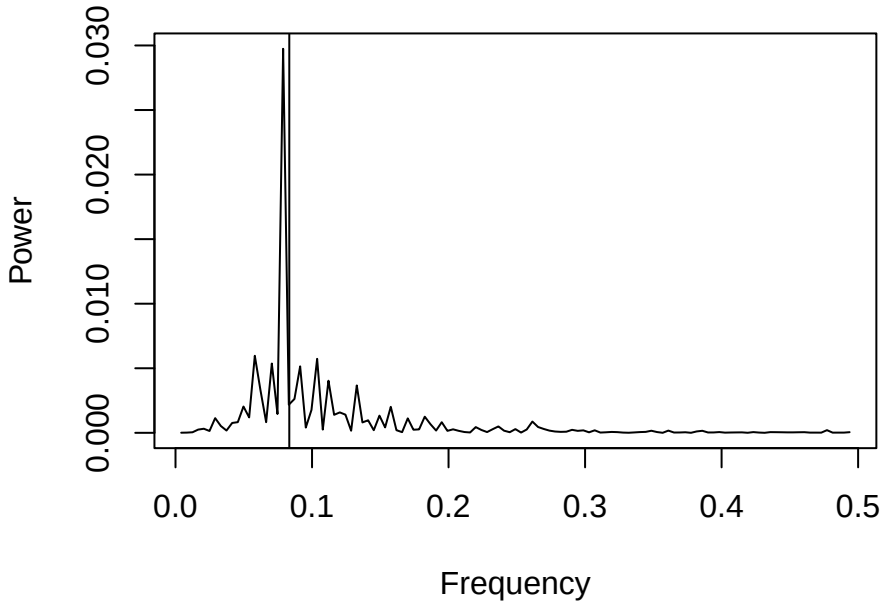

```
## [1] 4.140846 4.162582 4.111223 4.020402 3.931375
```

Though the values are not very closed to the real data points, it gives some insight about the general trend of the data.

3.8 Spectral analysis (Extra points)

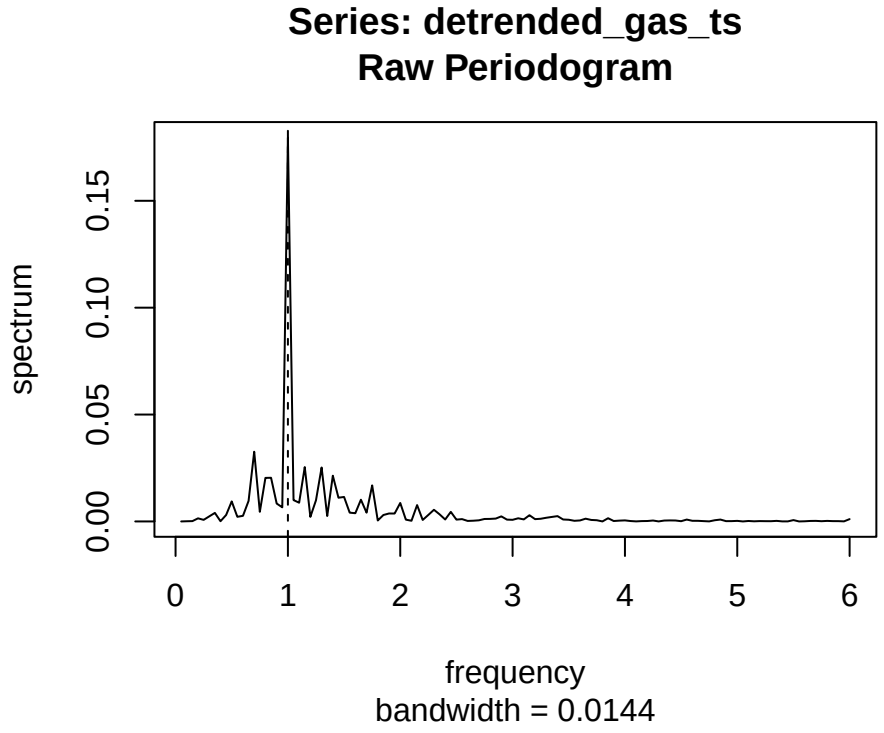
After detrend and deseasonalize the data, we can use spectral analysis to insvestigate periodic behavior in the modified residuals.

First we use the Fast Fourier Transform(FFT) similar to Example 4.1.2 to check if there is indeed a strong periodic behavior. In such way we obatin a frequency power graph of the data:



We can see that the peak of the frequency is very close to the line at 1/12, which indicate there might be a yearly(12 month period) recurring pattern.

In R we could also use the Spec.pgram fuction to fine-tune the spectral analysis(as in Example 4.3.1). In such way we obatin the following plot(the unit of x-axis is year):



There is indeed a very clear peak at one year.

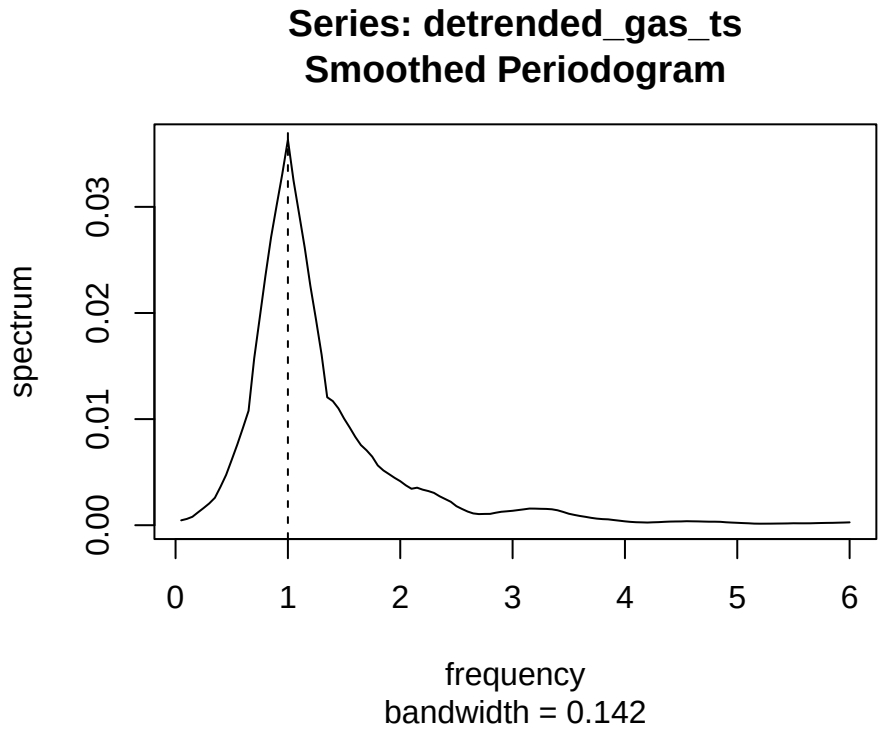
Using the numerical values of this analysis, the following confidence intervals are obtained at the level $\alpha = .1$:

```
## [1] 0.04869648
```

```
## [1] 7.09522
```

This confidence interval is very small. but if its large, we may shrink it by a smoothing approach which uses an averaging procedure over a band of neighboring frequencies, which will be unnecessary.

We can obtain a plot:



And a confidence interval:

```
## [1] 0.002428796
```

```
## [1] 0.00865321
```

4 Discussion

The advantage of our model is that we chose the ARIMA model by relatively good ACF and PACF plots, and by comparing with the auto arima function, we have a better result that can be used to predict the rough component. Also when we were choosing the ma coefficient, we chose it not to be very large so the trend are most likely got captured. And when we were using spectral analysis, we only saw one significant peak and that indicates a cycle of one year which isn't very helpful since that is quiet obvious.

The defect of our model and process is that since there are many manually controlled parameters, the model does not fix very well according to the real data. For example, when we were trying to forecast the real gas price, since we used the linear model with cubic equation, our tail of the predicted values are not really good.

5 Conclusions

The whole forecating process indicates that the gas price follows a relatively strong annually pattern. The spectual analysis shows there is indeed an annual cycle in the data which it makes sense since oil price typically rise during winter and fall during summer. However, there are many fluctuations are not describable by our model like the import and export policy and when the us government will release the gas storage, etc.

6 References

Lecture Notes.

"Time Series Analysis and Its Applications", Robert H. Shumway, Davis S. Stoffer

7 Appendix

Include all codes and additional supporting calculations here.

```
library(dplyr)
library(tseries)
library(forecast)

# load dataset
gas.dat = read.csv("../data/GASREGW.csv")
gas.dat[, 2] = as.numeric(gas.dat[, 2])
gas_ts = ts(gas.dat[,2], start = 2004, frequency = 12)
ts.plot(gas_ts)

# deseasonalization by ma
mhat = stats::filter(gas_ts, sides = 2, c(0.5, rep(1,11), 0.5)/12)
A = matrix(gas_ts, ncol=12, byrow = TRUE)
M = matrix(mhat, ncol=12, byrow = TRUE)
mu = array(0,12)
for (k in 1:6) mu[k] = sum(A[2:20,k]-M[2:20,k])/19
for (k in 7:12) mu[k] = sum(A[1:19,k]-M[1:19,k])/19
shat = mu-mean(mu)
print(shat)
shat = rep(shat, 20)
deseasonalized_gas_ts = gas_ts - ts(shat, start = start(gas_ts), frequency = 12)
ts.plot(gas_ts, deseasonalized_gas_ts, col=c("black", "blue"), lty=c(1,2))

# detrend by ma
ma5 = stats::filter(gas_ts, sides = 2, rep(1,11)/11)
ts.plot(gas_ts, ma5,col=c("black","red"))
detrended_gas_ts = gas_ts - ma5
detrended_gas_ts = na.omit(detrended_gas_ts)
```

```r
ts.plot(detrended_gas_ts, ylab="residuals", xlab="Time")


# adf test
adf.test(detrended_gas_ts, alternative = "stationary")

# acf and pacf
acf(detrended_gas_ts, main="ACF plot")
pacf(detrended_gas_ts, main="PACF plot")

# arima model
arima_model = Arima(detrended_gas_ts, order = c(2,0,1))
summary(arima_model)


# forecast the rough component
arima_forecasted_values <- forecast(arima_model, h = 12)
plot(arima_forecasted_values)

# forecast real price
ma5 = na.omit(ma5)
t = 1:length(ma5)
t2 = t^2
t3 = t^3
trend_model = lm(ma5~t+t2+t3)
summary(trend_model)

fitted_values = fitted(trend_model)
fitted_ts = ts(fitted_values, start = start(ma5), frequency = 12)
ts.plot(ma5, fitted_ts, col=c("black", "blue"))

t_max <- max(t)
n <- 6
future_t <- (t_max + 1):(t_max + n)
future_t2 <- future_t^2
future_t3 <- future_t^3
trend_predictions <- predict(trend_model, newdata = data.frame(t = future_t, t2 = future_t2, t3 = futur
trend_ts <- ts(trend_predictions, start = c(2023, 8), frequency = 12)
ts.plot(ma5, fitted_ts, trend_ts, col=c("black", "blue", "red"), lwd=c(1,1,2))


predicted_prices <- numeric(5)
for (i in 8:12) {
  predict_price <- trend_ts[i - 7] + shat[i] + arima_forecasted_values$mean[i - 7]
  predicted_prices[i - 7] <- predict_price
}


# spectual analysis
I <- abs(fft(detrended_gas_ts))^2/241
P <- (4/241)*I[1:120]
f <- 0:119/241
plot(f[-1],P[-1], type="l", xlab="Frequency", ylab="Power")
```

```
abline(v=1/12)


gas.pgram = spec.pgram(detrended_gas_ts, taper=0, log="no")
abline(v=1, lty=2)

u = qchisq(.025, 2)
l = qchisq(.975, 2)
2 * gas.pgram$spec[20]/l
2 * gas.pgram$spec[20]/u

k = kernel("daniell",c(3,3))

gas.ave = spec.pgram(detrended_gas_ts, k, taper=0, log="no")

abline(v=1, lty=2)

df = ceiling(gas.ave$df)
u=qchisq(.025,df)
l = qchisq(.975,df)
df * gas.ave$spec[40]/l
df * gas.ave$spec[40]/u
```