Analyzing UKgas data which is quarterly UK gas consumption from 1960 Q1 to 1986 Q4, in millions of therms. A quarterly time series of length 108.

```
library(fUnitRoots)
library(astsa)
library(forecast)

#storing my data and doing a basic plot to see trend and seasonality:
ukgas<-UKgas
plot(ukgas)
```

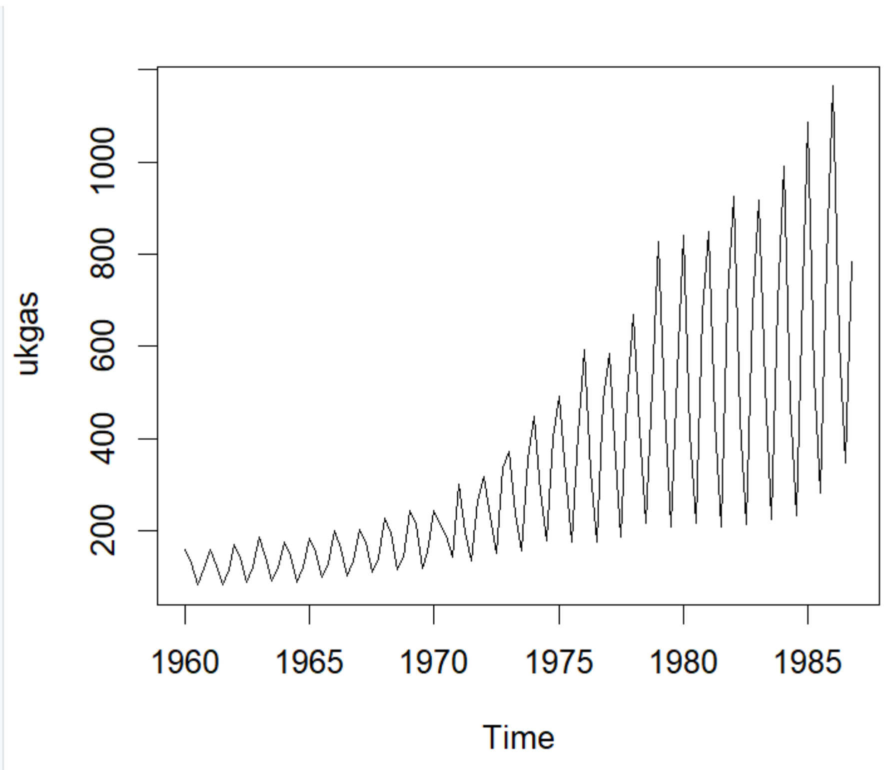


Figure 1: Basic Plot

There is a clear seasonal pattern and an upward trend.

```
#perform a lag-4 difference to remove the periodicity since it is quarterly data
X<-diff(ukgas, lag = 4)
X
plot(X)
mean(X)
abline(h = 0)
```

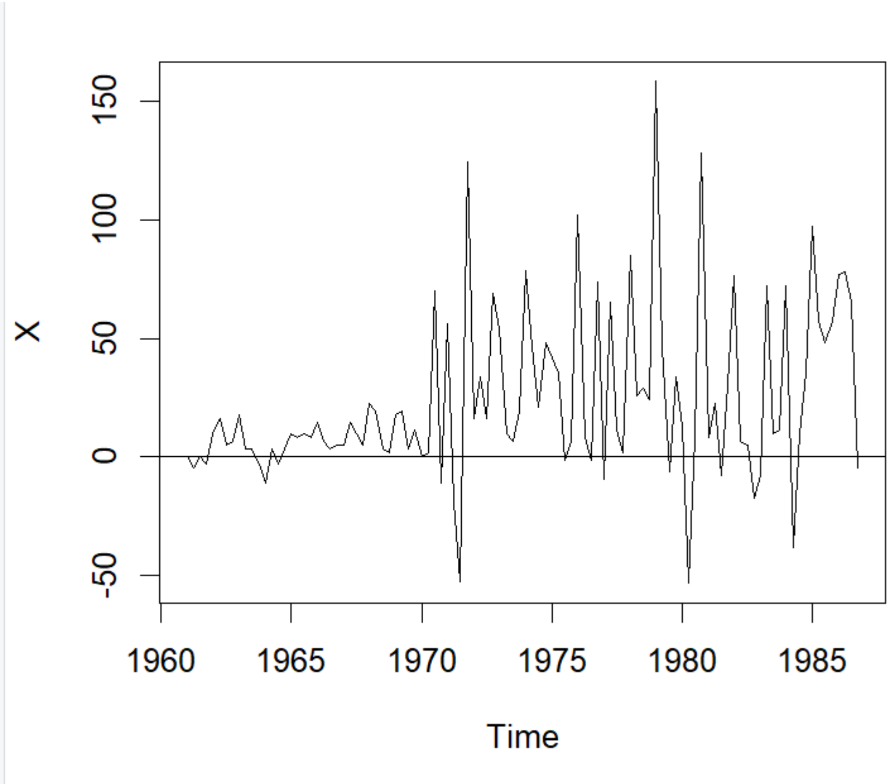


Figure 2: lag 4 diffed ukgas

It has actually removed much of the strong seasonal pattern, but we still can't say it is stationary. Our current mean is 23.19712.

```
#like in 3rd homework I am going directly to combining log transformation
#and differencing.
ukgas_log <- log(ukgas)
ukgas_log_diff <- diff(ukgas_log)
ukgas_log_diff_seasonal <- diff(ukgas_log_diff, lag=4)

#checking stationarity
library(tseries)
adf.test(ukgas_log_diff_seasonal)
```

```
> adf.test(ukgas_log_diff_seasonal)

        Augmented Dickey-Fuller Test

data:  ukgas_log_diff_seasonal
Dickey-Fuller = -7.9249, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(ukgas_log_diff_seasonal) : p-value smaller than printed p-value
```

Figure 3: stationarity results

```
Y<-ukgas_log_diff_seasonal

plot(Y)
mean(Y)
abline(h = 0)
```

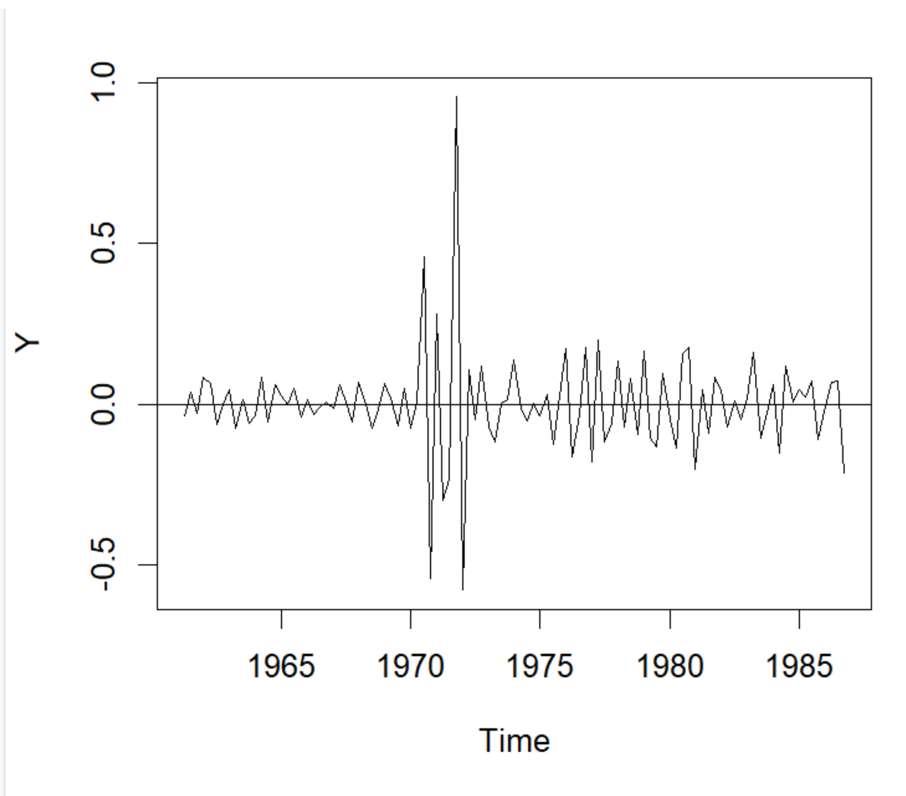
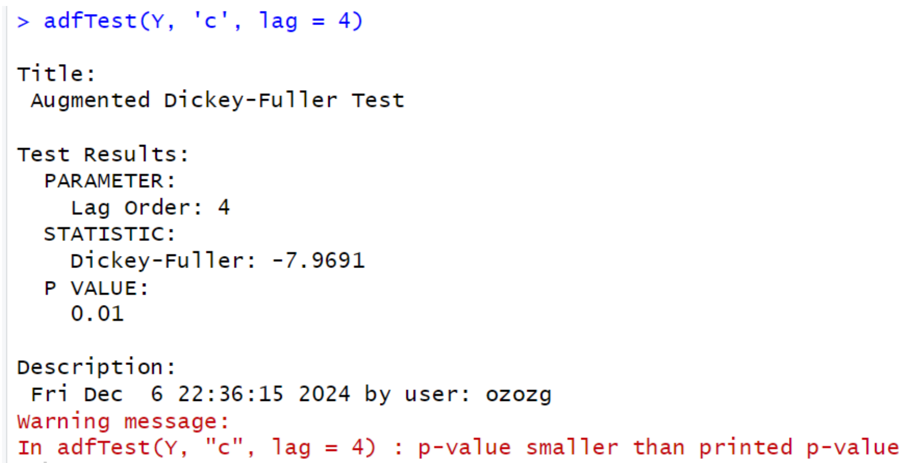


Figure 4: transformed ukgas

My current mean is -5.935059e-05. It is pretty close to zero.

Running augmented Dickey–Fuller Test

```
(length(Y)-1)^(1/3)  #4.672329 I take 4
adfTest(Y, 'c', lag = 4)
# Small p-value --> reject null hypothesis/do not difference again
```



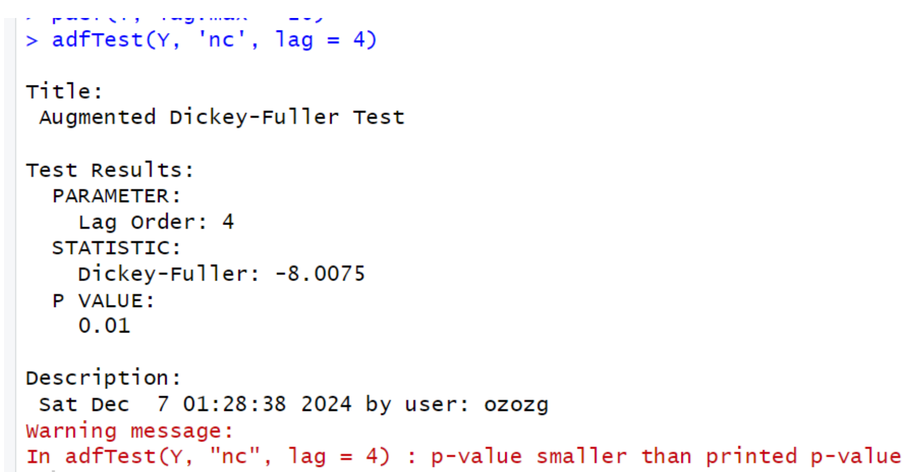
```
> adfTest(Y, 'c', lag = 4)

Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 4
  STATISTIC:
    Dickey-Fuller: -7.9691
  P VALUE:
    0.01

Description:
 Fri Dec  6 22:36:15 2024 by user: ozozg
Warning message:
In adfTest(Y, "c", lag = 4) : p-value smaller than printed p-value
```

Figure 5: ADF test result

p-value is less than 0.05, we reject the null hypothesis of a unit root.

Running it for no constant as well;

```
adfTest(Y, 'nc', lag = 4)
```



```
> adfTest(Y, 'nc', lag = 4)

Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 4
  STATISTIC:
    Dickey-Fuller: -8.0075
  P VALUE:
    0.01

Description:
 Sat Dec  7 01:28:38 2024 by user: ozozg
Warning message:
In adfTest(Y, "nc", lag = 4) : p-value smaller than printed p-value
```

Figure 6: ADF nc results

Looking at ACF and PACF of the logged differenced data

```
par(mfrow = c(1,2))
acf(Y, lag.max = 20)
pacf(Y, lag.max = 20)
```

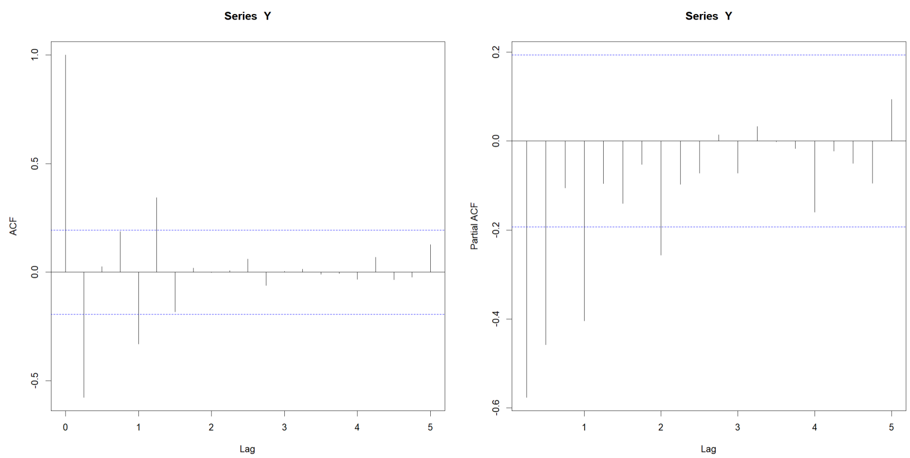


Figure 7: ACf and PACF of logged differenced data

The ACF seems to have some spikes, and then a decay or values hovering around zero for higher lags. Early spike and not much after tells me MA. Also looking at PACF, it doesn't show clear cut off. It looks like it can be MA type model.

```
d=1
q=1
D=1
Q=1 non seasonal MA(1)
```

SARIMA$(0,1,1)(0,1,1)_4$
Running autoarima on original data:

```
auto.arima(ukgas, trace = TRUE, approximation = FALSE)
```

```
> auto.arima(ukgas, trace = TRUE, approximation = FALSE)

 ARIMA(2,1,2)(1,1,1)[4]                    : Inf
 ARIMA(0,1,0)(0,1,0)[4]                    : 1099.253
 ARIMA(1,1,0)(1,1,0)[4]                    : 1079.071
 ARIMA(0,1,1)(0,1,1)[4]                    : 1032.911
 ARIMA(0,1,1)(0,1,0)[4]                    : 1030.795
 ARIMA(0,1,1)(1,1,0)[4]                    : 1032.91
 ARIMA(0,1,1)(1,1,1)[4]                    : Inf
 ARIMA(1,1,1)(0,1,0)[4]                    : 1032.078
 ARIMA(0,1,2)(0,1,0)[4]                    : 1031.488
 ARIMA(1,1,0)(0,1,0)[4]                    : 1077.142
 ARIMA(1,1,2)(0,1,0)[4]                    : 1032.042

 Best model: ARIMA(0,1,1)(0,1,0)[4]

Series: ukgas
ARIMA(0,1,1)(0,1,0)[4]

Coefficients:
          ma1
      -0.9297
s.e.   0.0336

sigma^2 = 1237:  log likelihood = -513.34
AIC=1030.67   AICc=1030.79   BIC=1035.94
```

Figure 8: Autoarima on original data

The smallest value comes from ARIMA(0,1,1)(0,1,0)[4]

```
> auto.arima(ukgas, trace = TRUE)

 ARIMA(2,1,2)(1,1,1)[4]                    : Inf
 ARIMA(0,1,0)(0,1,0)[4]                    : 1099.253
 ARIMA(1,1,0)(1,1,0)[4]                    : 1079.071
 ARIMA(0,1,1)(0,1,1)[4]                    : 1032.911
 ARIMA(0,1,1)(0,1,0)[4]                    : 1030.795
 ARIMA(0,1,1)(1,1,0)[4]                    : 1032.91
 ARIMA(0,1,1)(1,1,1)[4]                    : Inf
 ARIMA(1,1,1)(0,1,0)[4]                    : 1032.078
 ARIMA(0,1,2)(0,1,0)[4]                    : 1031.488
 ARIMA(1,1,0)(0,1,0)[4]                    : 1077.142
 ARIMA(1,1,2)(0,1,0)[4]                    : 1032.042

 Best model: ARIMA(0,1,1)(0,1,0)[4]

Series: ukgas
ARIMA(0,1,1)(0,1,0)[4]

Coefficients:
          ma1
      -0.9297
s.e.   0.0336

sigma^2 = 1237:  log likelihood = -513.34
AIC=1030.67   AICc=1030.79   BIC=1035.94
```

Figure 9: Without approximation

So there is no change in the model.
Comparing with what auto.arima does for our logged and differenced data:

```
auto.arima(Y, allowdrift = FALSE, trace = TRUE, stationary = TRUE,
approximation = FALSE)
```

```
ARIMA(2,0,2)(1,0,1)[4] with non-zero mean : -161.184
 ARIMA(0,0,0)           with non-zero mean : -77.463
 ARIMA(1,0,0)(1,0,0)[4] with non-zero mean : -119.846
 ARIMA(0,0,1)(0,0,1)[4] with non-zero mean : -161.9614
 ARIMA(0,0,0)           with zero mean     : -79.54339
 ARIMA(0,0,1)           with non-zero mean : Inf
 ARIMA(0,0,1)(1,0,1)[4] with non-zero mean : -159.8243
 ARIMA(0,0,1)(0,0,2)[4] with non-zero mean : -159.8167
 ARIMA(0,0,1)(1,0,0)[4] with non-zero mean : -161.9775
 ARIMA(0,0,1)(2,0,0)[4] with non-zero mean : -159.8186
 ARIMA(0,0,1)(2,0,1)[4] with non-zero mean : -157.8671
```

```
ARIMA(0,0,0)(1,0,0)[4] with non-zero mean : -87.09543
ARIMA(1,0,1)(1,0,0)[4] with non-zero mean : -163.4715
ARIMA(1,0,1)           with non-zero mean : -161.9497
ARIMA(1,0,1)(2,0,0)[4] with non-zero mean : -161.2158
ARIMA(1,0,1)(1,0,1)[4] with non-zero mean : -161.2161
ARIMA(1,0,1)(0,0,1)[4] with non-zero mean : -163.3795
ARIMA(1,0,1)(2,0,1)[4] with non-zero mean : Inf
ARIMA(2,0,1)(1,0,0)[4] with non-zero mean : -163.7062
ARIMA(2,0,1)           with non-zero mean : -160.8068
ARIMA(2,0,1)(2,0,0)[4] with non-zero mean : -161.5295
ARIMA(2,0,1)(1,0,1)[4] with non-zero mean : -161.6276
ARIMA(2,0,1)(0,0,1)[4] with non-zero mean : -163.9303
ARIMA(2,0,1)(0,0,2)[4] with non-zero mean : -161.6271
ARIMA(2,0,1)(1,0,2)[4] with non-zero mean : -159.3007
ARIMA(2,0,0)(0,0,1)[4] with non-zero mean : -139.6798
ARIMA(3,0,1)(0,0,1)[4] with non-zero mean : -165.3545
ARIMA(3,0,1)           with non-zero mean : -158.9575
ARIMA(3,0,1)(1,0,1)[4] with non-zero mean : -164.2443
ARIMA(3,0,1)(0,0,2)[4] with non-zero mean : -164.118
ARIMA(3,0,1)(1,0,0)[4] with non-zero mean : -163.0194
ARIMA(3,0,1)(1,0,2)[4] with non-zero mean : -161.8847
ARIMA(3,0,0)(0,0,1)[4] with non-zero mean : -157.0604
ARIMA(3,0,2)(0,0,1)[4] with non-zero mean : -164.4985
ARIMA(2,0,2)(0,0,1)[4] with non-zero mean : -163.5264
ARIMA(3,0,1)(0,0,1)[4] with zero mean     : -167.1337
ARIMA(3,0,1)           with zero mean     : -160.7212
ARIMA(3,0,1)(1,0,1)[4] with zero mean     : -166.0505
ARIMA(3,0,1)(0,0,2)[4] with zero mean     : -165.9285
ARIMA(3,0,1)(1,0,0)[4] with zero mean     : -164.8509
ARIMA(3,0,1)(1,0,2)[4] with zero mean     : -163.7379
ARIMA(2,0,1)(0,0,1)[4] with zero mean     : -165.7112
ARIMA(3,0,0)(0,0,1)[4] with zero mean     : -158.848
ARIMA(3,0,2)(0,0,1)[4] with zero mean     : -166.3703
ARIMA(2,0,0)(0,0,1)[4] with zero mean     : -141.8376
ARIMA(2,0,2)(0,0,1)[4] with zero mean     : -165.3665
```

```
 Best model: ARIMA(3,0,1)(0,0,1)[4] with zero mean

Series: Y
ARIMA(3,0,1)(0,0,1)[4] with zero mean

Coefficients:
          ar1      ar2      ar3      ma1     sma1
      -0.5620  -0.6076  -0.4454  -0.5033  -0.6159
s.e.   0.1655   0.1879   0.1901   0.1751   0.1610

sigma^2 = 0.01044:  log likelihood = 90
AIC=-168.01   AICc=-167.13   BIC=-152.2
```

Figure 10: logged diffed ukgas with no drift

Best model: ARIMA(3,0,1)(0,0,1)[4] with zero mean.
Trying it on logged diffed data:

```
model <- Arima(Y, order = c(3, 0, 1),
seasonal = list(order = c(0, 0, 1), period = 4))
#summary of the fitted model
summary(model)
```

```
> summary(model)
Series: Y
ARIMA(3,0,1)(0,0,1)[4] with non-zero mean

Coefficients:
          ar1      ar2      ar3      ma1     sma1   mean
      -0.5645  -0.6115  -0.4496  -0.5057  -0.6199  6e-04
s.e.   0.1650   0.1875   0.1902   0.1744   0.1610  8e-04

sigma^2 = 0.01049:  log likelihood = 90.27
AIC=-166.53   AICc=-165.35   BIC=-148.09

Training set error measures:
                      ME       RMSE        MAE       MPE     MAPE      MASE        ACF1
Training set 0.001791838 0.09937554 0.06784071 -642.7627 833.3422 0.4090443 -0.04060929
```

Figure 11: model results on transformed data

```
#plot diagnostics for the model
checkresiduals(model)
```

```
> checkresiduals(model)

        Ljung-Box test

data:  Residuals from ARIMA(3,0,1)(0,0,1)[4] with non-zero mean
Q* = 4.0299, df = 3, p-value = 0.2583

Model df: 5.   Total lags used: 8
```

Figure 12: diagnostics

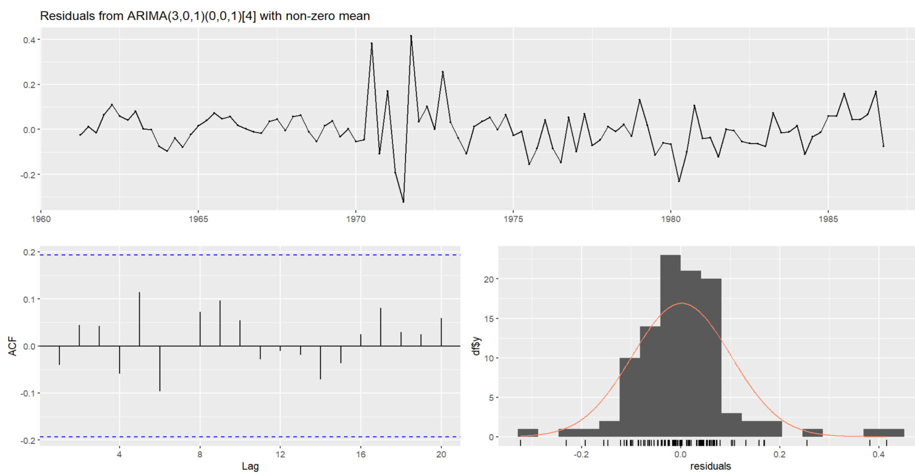


Figure 13: plots

AIC = -166.53 which is quite low, indicating a decent fit (aic being significantly lower than 1030.67 in previous one, and acf plot having no spikes are couple of the reasons why I went with this model). AICc = -165.35 is also low. From Ljung-Box Test, I got pvalue = 0.2583. pvalue ¿0.05 indicates there is no strong evidence of autocorrelation remaining in the residuals. The ACF doesn't show any spikes. The residual plot shows fluctuations around zero with no visible pattern. I think the model fits to the data, although admittedly it is a bit complex.
Running sarima as well

```
sarima(Y, 3,0,1, 0,0,1, 4)
```

```
Coefficients:
      Estimate     SE t.value p.value
ar1    -0.5645 0.1650 -3.4222  0.0009
ar2    -0.6115 0.1875 -3.2609  0.0015
ar3    -0.4496 0.1902 -2.3643  0.0201
ma1    -0.5057 0.1744 -2.8993  0.0046
sma1   -0.6199 0.1610 -3.8512  0.0002
xmean   0.0006 0.0008  0.7127  0.4777

sigma^2 estimated as 0.009875498 on 97 degrees of freedom

AIC = -1.61683  AICc = -1.608335  BIC = -1.437771
```

Figure 14: Sarima coefficients

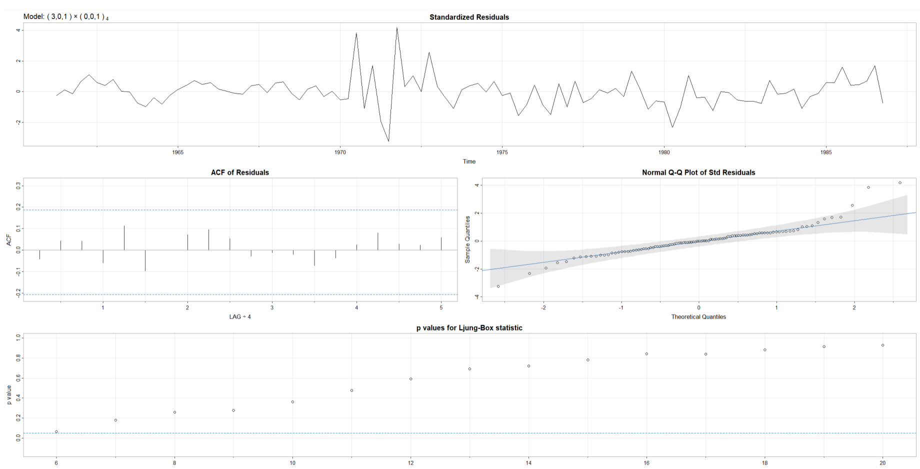


Figure 15: sarima plots

The residuals fluctuate around zero without an obvious trend. The ACF has no spikes. The QQ plot unfortunately has some deviation at the tails. all the p values are above the line. My fitted model is;

$$X_t = (1-0.5645B-0.6115B^2-0.4496B^3)(1-0.6199B^4)(X_t-\bar{X}) = (1-0.5057B)(1-0.5057B^4)W_t$$

Forecasting

```
#forecasting 12 steps ahead
fcast <- forecast(model, h=12)
print(fcast)
plot(fcast)
```

```
> print(fcast)
        Point Forecast       Lo 80     Hi 80      Lo 95     Hi 95
1987 Q1   0.0720534338 -0.05918116 0.2032880 -0.1286526 0.2727594
1987 Q2   0.0317064423 -0.16050924 0.2239221 -0.2622621 0.3256750
1987 Q3  -0.0469898713 -0.23920798 0.1452282 -0.3409621 0.2469824
1987 Q4   0.0755514143 -0.11861299 0.2697158 -0.2213974 0.3725003
1988 Q1  -0.0501815329 -0.24714732 0.1467843 -0.3514147 0.2510516
1988 Q2   0.0047271521 -0.19698272 0.2064370 -0.3037615 0.3132158
1988 Q3  -0.0044778119 -0.20687268 0.1979171 -0.3140141 0.3050584
1988 Q4   0.0236694051 -0.17873875 0.2260776 -0.2858872 0.3332260
1989 Q1  -0.0112763380 -0.21384527 0.1912926 -0.3210788 0.2985261
1989 Q2  -0.0046225657 -0.20763772 0.1983926 -0.3151075 0.3058623
1989 Q3   0.0003354392 -0.20268616 0.2033570 -0.3101593 0.3108302
1989 Q4   0.0091787361 -0.19387631 0.2122338 -0.3013672 0.3197246
> plot(fcast)
```

Figure 16: forecasting results

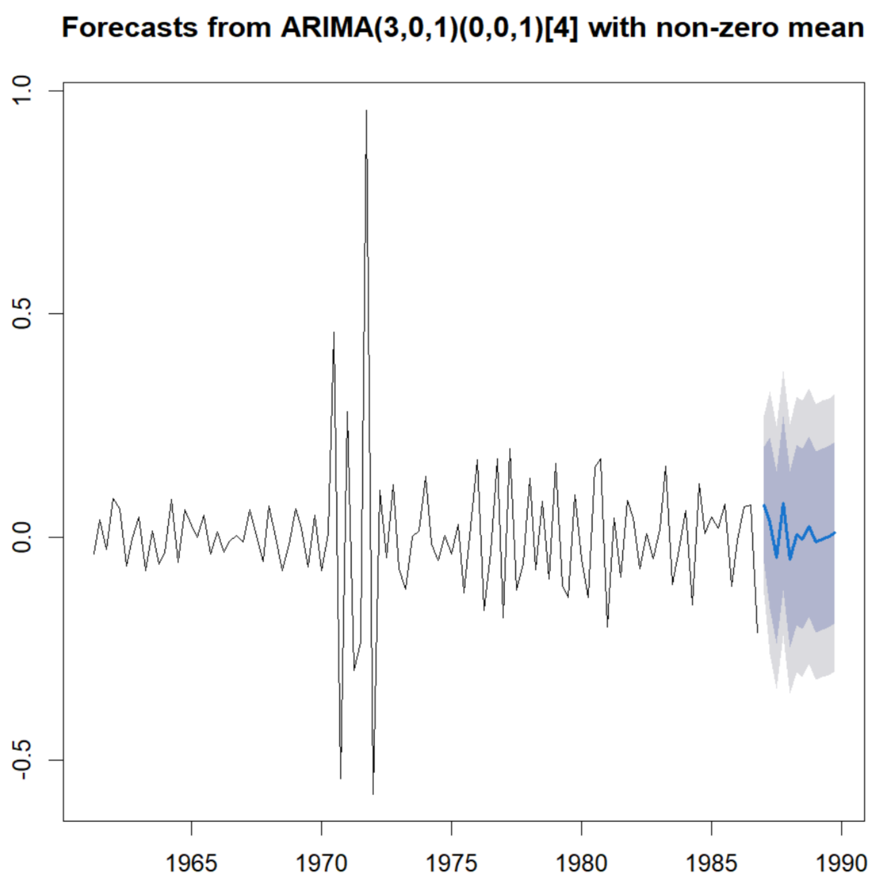


Figure 17: forecasting plot for 12 steps/4 years ahead

Exponential Smoothing Fit

```
ukgasHW <- HoltWinters(ukgas)
ukgasHW
```

```
Holt-Winters exponential smoothing with trend and additive seasonal component.

Call:
HoltWinters(x = ukgas)

Smoothing parameters:
 alpha: 0.02127689
 beta : 1
 gamma: 0.9899092

Coefficients:
         [,1]
a    621.06125
b     10.12142
s1   574.20494
s2    12.86129
s3 -263.79964
s4   162.20290
```

Figure 18: Holt Winters parameters

plotting against actual data:

```
plot(ukgasHW)
checkresiduals(ukgasHW)
```

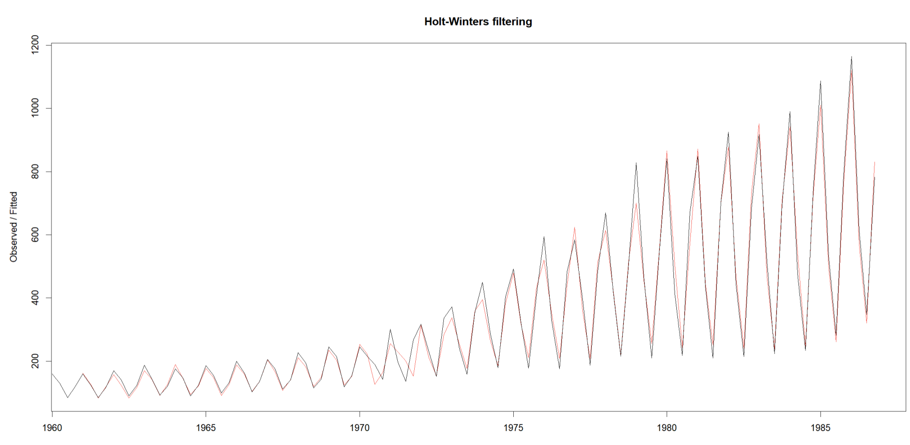


Figure 19: Holt-Winters

```
> checkresiduals(ukgasHW)

        Ljung-Box test

data:  Residuals from HoltWinters
Q* = 10.111, df = 8, p-value = 0.2573

Model df: 0.   Total lags used: 8
```

Figure 20: Holt-Winters residuals

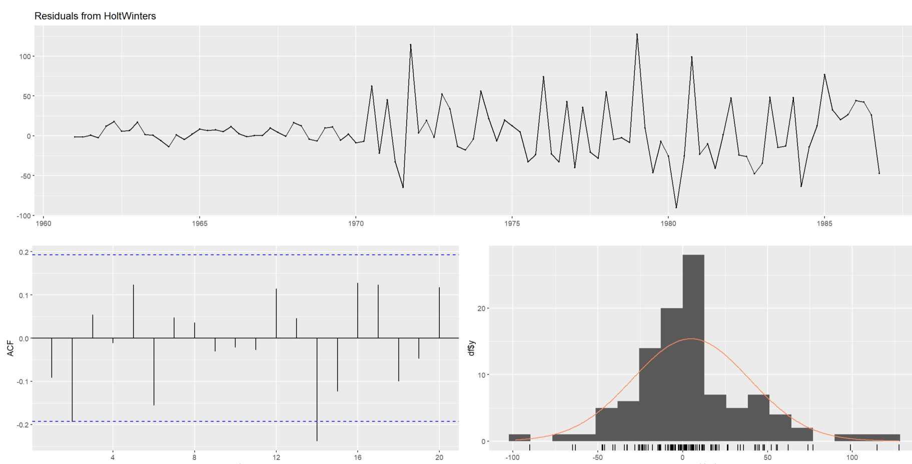


Figure 21: Holt-Winters residuals plots

While residuals fluctuate around zero, there are some periods that has deviations. I see only one spike in ACF. The histogram of residuals is a bit skewed which means the model can be improved. But the Ljung box pvalue of 0.2573 is really higher than 0.05.

```
ukgasforecastHW <- forecast(ukgasHW, h = 12)
plot(ukgasforecastHW)
```

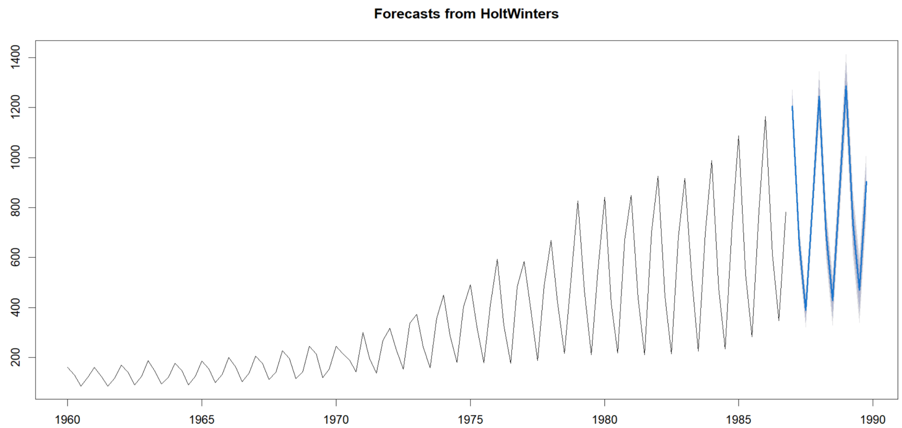


Figure 22: Holt-Winters forecast

Applying holt-winters on transformed data:

```
transformedukgasHW<-HoltWinters(Y)
transformedukgasHW
```

```
> transformedukgasHW
Holt-Winters exponential smoothing with trend and additive seasonal component.

Call:
HoltWinters(x = Y)

Smoothing parameters:
 alpha: 0.007639264
 beta : 0.9470262
 gamma: 0.06341883

Coefficients:
            [,1]
a   -0.0271486014
b    0.0001084341
s1   0.0269357672
s2   0.0086248934
s3  -0.0029523138
s4   0.0080380637
```

Figure 23: Holt-Winters parameters on transformed data

```
plot(transformedukgasHW)
checkresiduals(transformedukgasHW)
```

```
> checkresiduals(transformedukgasHW)

        Ljung-Box test

data:  Residuals from HoltWinters
Q* = 55.034, df = 8, p-value = 4.349e-09

Model df: 0.   Total lags used: 8
```

Figure 24: Ljung-box on residuals

the p value here is way too low.

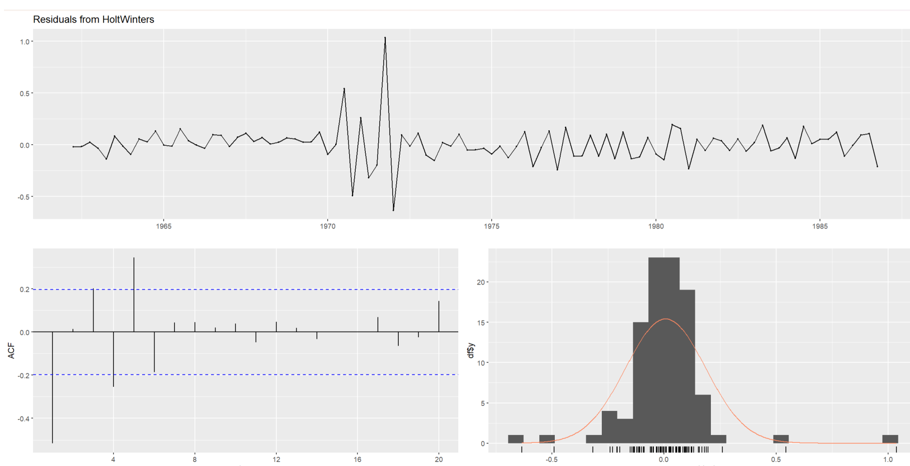


Figure 25: Residuals for HW on transformed data

I see a couple spikes on ACF, but the histogram distribution is so much closer to a normal distribution. plot(transformedukgasHW)

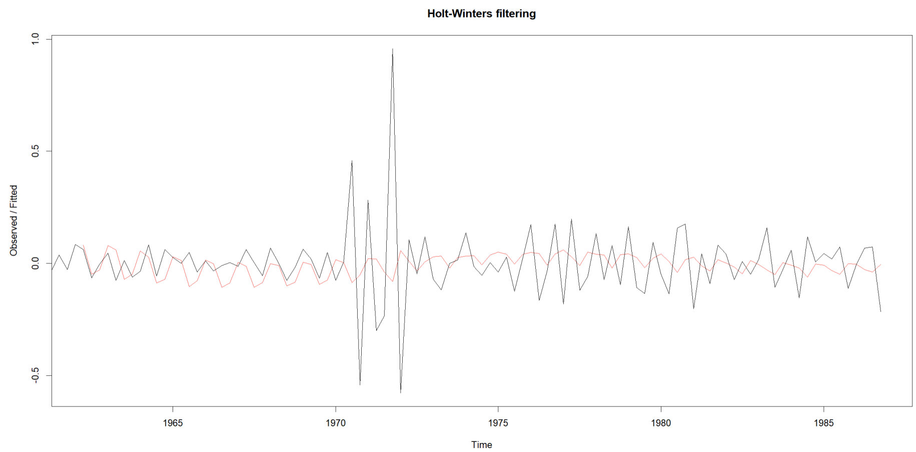


Figure 26: Holt-Winters transformed data plot

Plotting the forecast of transformed data on holt winters:

```
transformedukgasforecastHW <- forecast(transformedukgasHW, h = 12)
plot(transformedukgasforecastHW)
```

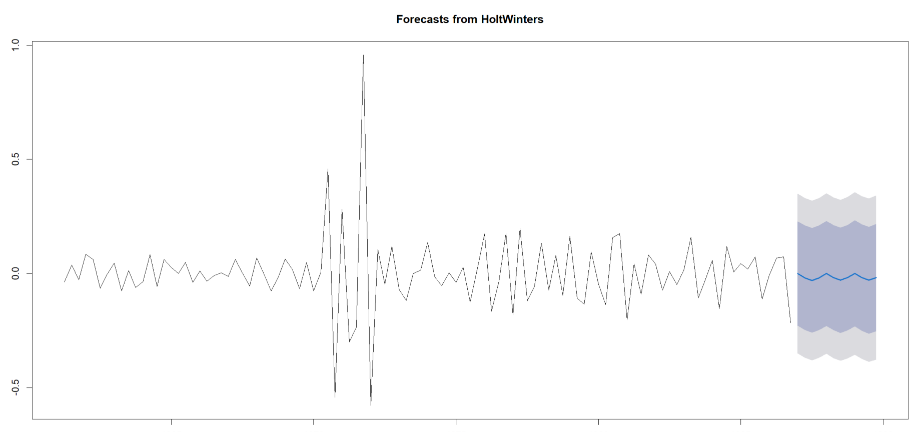


Figure 27: Holt-Winter forecast on transformed data

Comparing the ARIMA and exponential smoothing forecasts

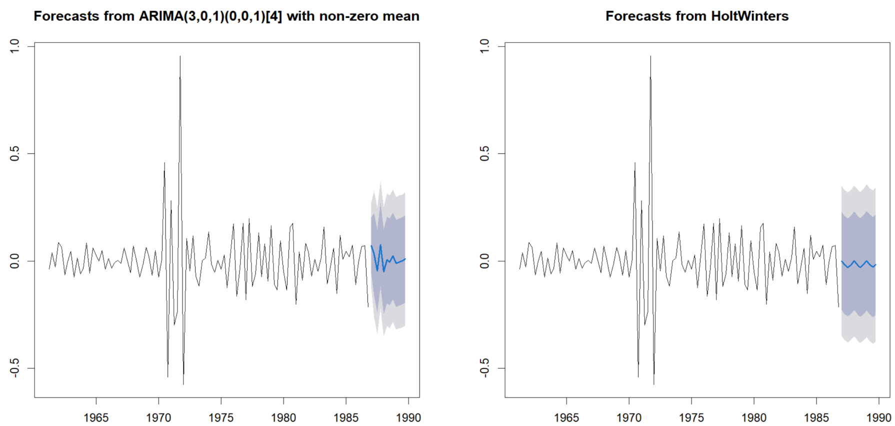


Figure 28: comparison

Honestly, my exponential smoothing forecast on transformed data doesn't seem to work. And comparing Holt Winters forecast on original data with my arima forecast on transformed data doesn't make much sense so I didn't add it. Point forecast equation manual calculation:

```
ukgasHW$coefficients
a <- as.numeric(ukgasHW$coefficients["a"])
b <- as.numeric(ukgasHW$coefficients["b"])
s1 <- as.numeric(ukgasHW$coefficients["s1"])
s2 <- as.numeric(ukgasHW$coefficients["s2"])
s3 <- as.numeric(ukgasHW$coefficients["s3"])
s4 <- as.numeric(ukgasHW$coefficients["s4"])

#one step ahead forecast Q1 1987, m is 1
forecastQ11987 <- a + 1*b + s1
forecastQ11987

#two steps ahead Q2 1987, m is 2
forecastQ21987 <- a + 2*b + s2
forecastQ21987

#comparing
hwforecast2 <- forecast(ukgasHW, h=2)
hwforecast2
```

```
> forecastQ11987 <- a + 1*b + s1
> forecastQ11987
[1] 1205.388
>
> #two steps ahead Q2 1987, m is 2
> forecastQ21987 <- a + 2*b + s2
> forecastQ21987
[1] 654.1654
> hwforecast2 <- forecast(ukgasHW, h=2)
> hwforecast2
        Point Forecast     Lo 80     Hi 80     Lo 95     Hi 95
1987 Q1      1205.3876 1161.1854 1249.5899 1137.7861 1272.9891
1987 Q2       654.1654  609.9231  698.4076  586.5027  721.8281
> 
```

Figure 29: manual calculation for point forecast