STA2202H1: Time Series Analysis Assignment 2

Student Name: Yulin Wang ID Number: 1003942326

2022-10-13

Q1 For the Box and Cox power transformations, use calculus to show that, for any fixed $x > 0$, as $\lambda \to 0$, $\frac{x^\lambda - 1}{\lambda} \to log(x)$.

For $lim_{\lambda \to 0} \frac{x^\lambda - 1}{\lambda}$, with an indeterminate form of both numerator and denominator having limit of 0, so we are entitled to apply L'Hospital's rule:

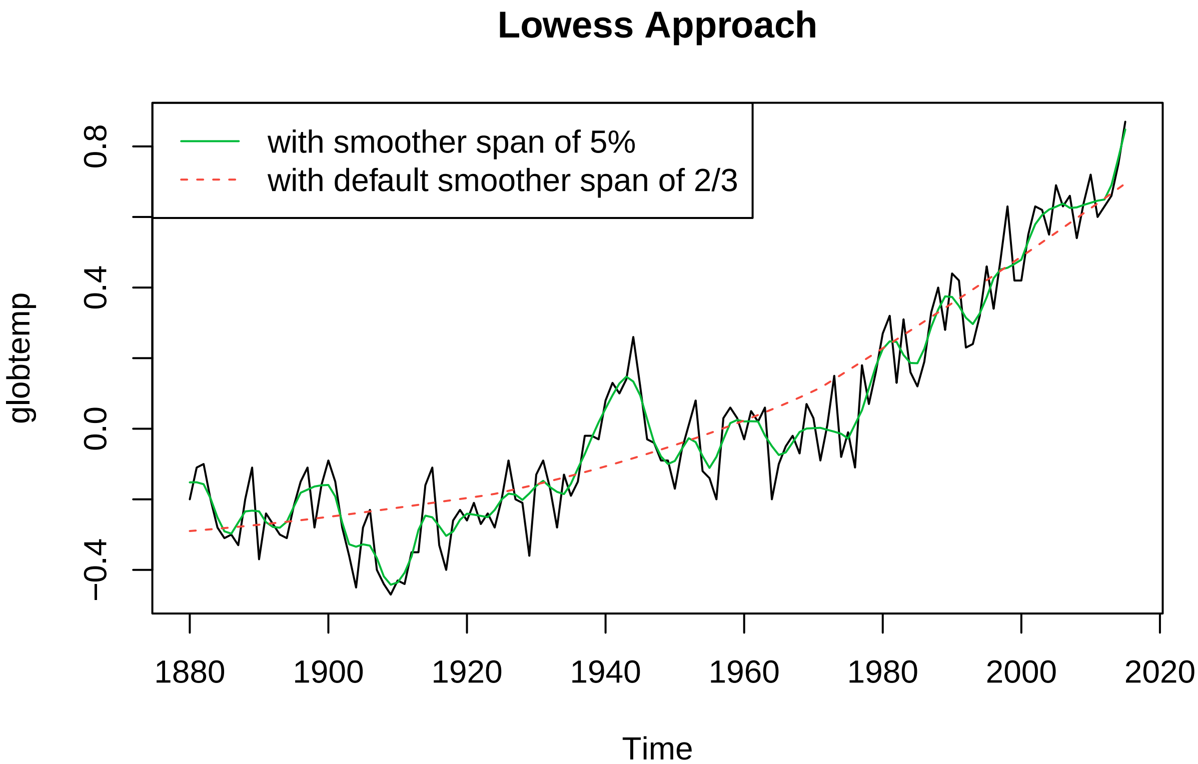
$$
\begin{aligned}
lim_{\lambda \to 0} \frac{x^\lambda - 1}{\lambda} &= lim_{\lambda \to 0} \frac{e^{log(x^\lambda)} - 1}{\lambda} \\
&= lim_{\lambda \to 0} \frac{e^{\lambda log(x)} - 1}{\lambda} \\
&= lim_{\lambda \to 0} \frac{\frac{d\{e^{\lambda log(x)} - 1\}}{d\lambda}}{\frac{d\lambda}{d\lambda}} \qquad \text{; take derivative of both numerator and denominator w.r.t. } \lambda \\
&= lim_{\lambda \to 0} \frac{e^{\lambda log(x)} \cdot \frac{d\{\lambda log(x)\}}{d\lambda}}{1} \\
&= lim_{\lambda \to 0} \{x^\lambda \cdot log(x)\} \\
&= x^0 log(x) \\
&= log(x)
\end{aligned}
$$

Thus, for any fixed $x > 0$, as $\lambda \to 0$, $\frac{x^\lambda - 1}{\lambda} \to log(x)$.

Q2

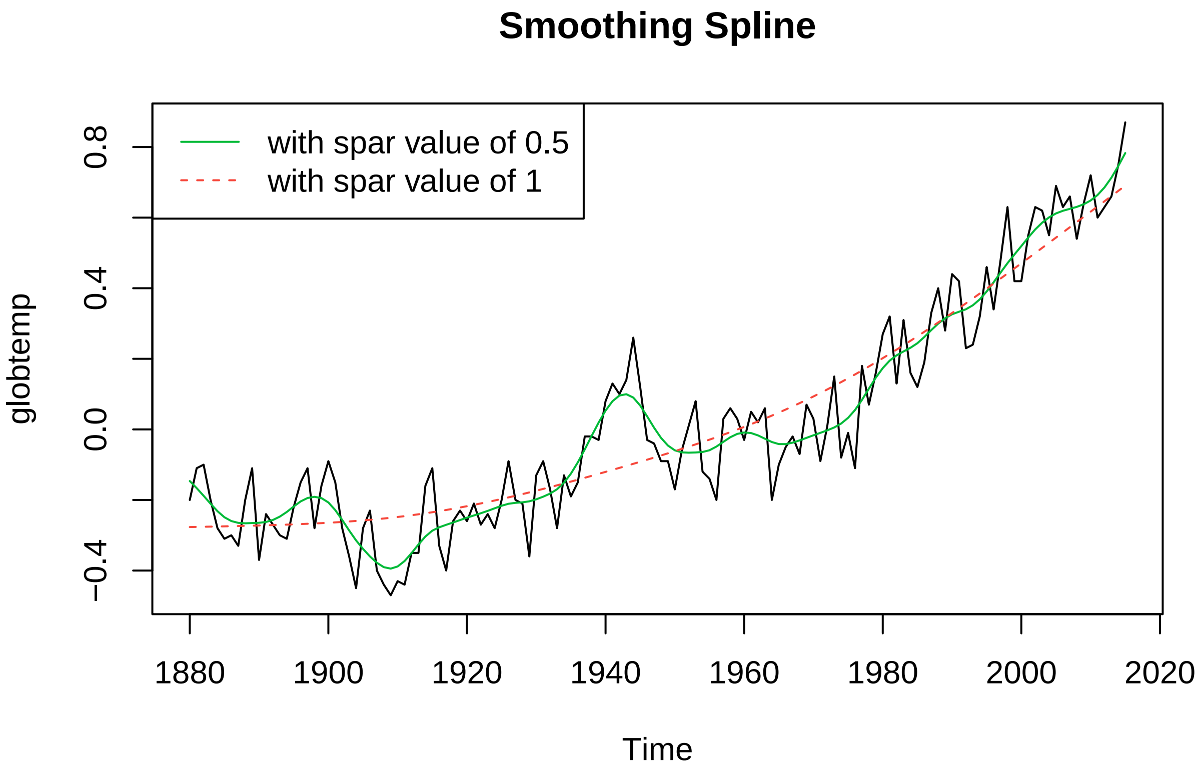
(1) Use Lowess Approach.

```r
library(astsa)
# Lowess approach
plot(globtemp, main="Lowess Approach")
lines(lowess(globtemp, f=0.05), lwd=1, col=3)  # use 5% of the data to obtain the estimate
lines(lowess(globtemp), lty=2, lwd=1, col=2) # with default span: f=2/3 -> more smoothness
legend("topleft", legend = c("with smoother span of 5%", "with default smoother span of 2/3"),
       col=c(3, 2), lty = 1:2)
```



(2) Use Smoothing Spline.

```r
# Spline approach
plot(globtemp, main="Smoothing Spline")
# use a spar value of 0.5 to emphasize the cycle
lines(smooth.spline(time(globtemp), globtemp, spar=0.5), lwd=1, col=3)
# use a spar value of 1 to emphasize the trend
lines(smooth.spline(time(globtemp), globtemp, spar=1), lty=2, lwd=1, col=2)
legend("topleft", legend = c("with spar value of 0.5", "with spar value of 1"),
       col=c(3, 2), lty = 1:2)
```



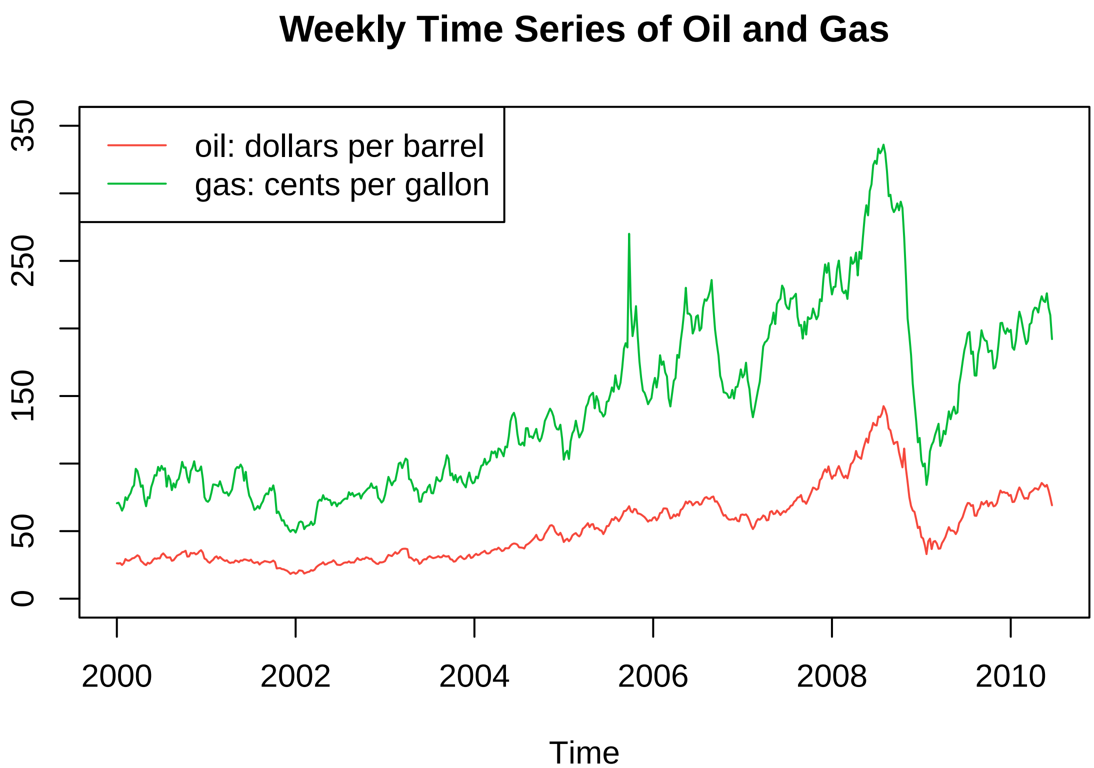
(3) Comment

As the above two plots show, the estimated trends (red dashed lines) after applying each smoothing technique are almost the same. Moreover, it seems that the smoothing spline technique is a little smoother than the Lowess technique.

Q3

1.

```
# Combine two time series data as data frame with ts object.
combined_data = ts.intersect(oil, gas, dframe=TRUE)
plot(combined_data$oil, col=2, main="Weekly Time Series of Oil and Gas", ylab ="", ylim=c(0, 350))
lines(combined_data$gas, col=3)
legend("topleft", legend = c("oil: dollars per barrel", "gas: cents per gallon"),
       col=c(2, 3), lty = 1:1)
```



Both of these two series are not stationary.

Recall the two conditions of a time series being stationary:
1. The mean is constant and doesn't depend on time t.
2. The autocovariance function depends on s and t only through their difference $|s - t|$.

The above plot shows that both of the mean values of oil and gas prices tend to increase relatively constantly with time until year 2008. However, it then decreases dramatically and then increases again between year 2008 and 2010. Thus, it is clear that the mean functions of the oil and gas prices are not constant over time. Therefore, these two series are not stationary.

2.

Let x_t be the oil or gas price series and assume y_t is the percentage change in price.
That is $y_t = \frac{x_t - x_{t-1}}{x_{t-1}}$ and $y_t \in [-1, 1]$. Then we have:

$$\begin{aligned} x_t &= x_{t-1} y_t + x_{t-1} \\ x_t &= x_{t-1}(y_t + 1) \\ y_t + 1 &= \frac{x_t}{x_{t-1}} \\ log(y_t + 1) &= log(\frac{x_t}{x_{t-1}}) \\ log(y_t + 1) &= log(x_t) - log(x_{t-1}) \end{aligned}$$

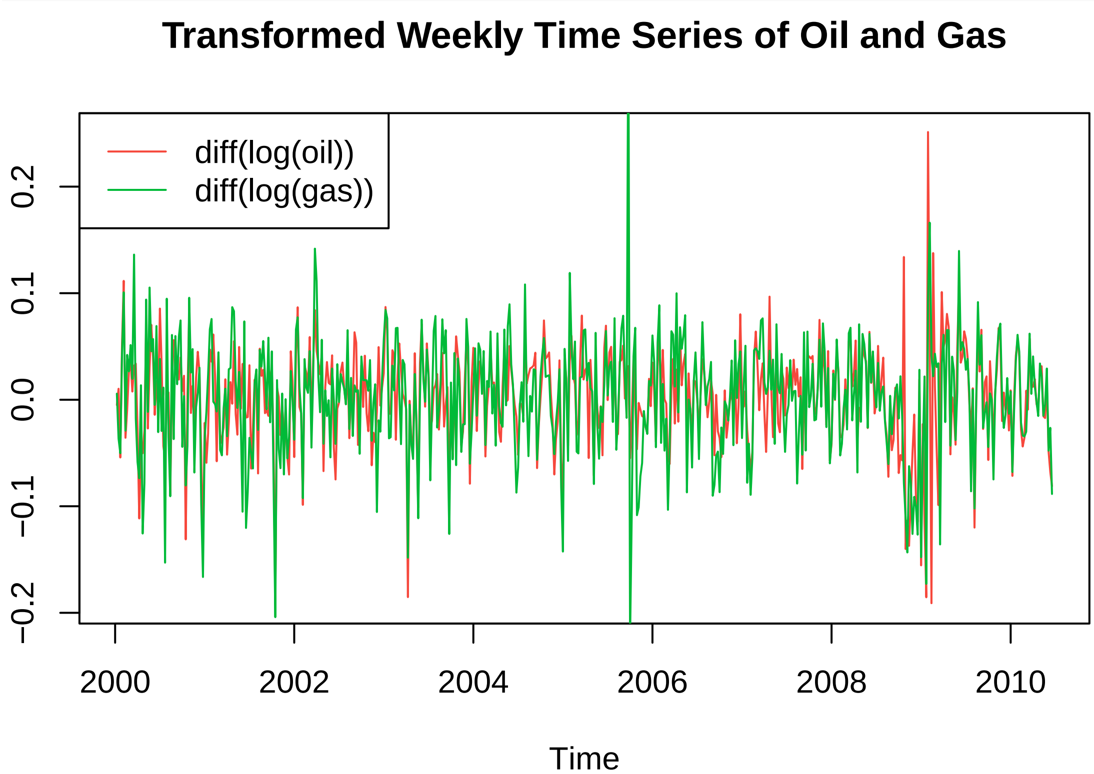
Since $log(y_t + 1) = y_t - \frac{y_t^2}{2} + \frac{y_t^3}{3} - \frac{y_t^4}{4} + ...$, and the higher-order terms are negligible for $y_t \in [-1, 1]$, so $log(y_t + 1) \approx y_t$. Then we have:

$$y_t = log(x_t) - log(x_{t-1}) = \nabla log x_t$$

3.

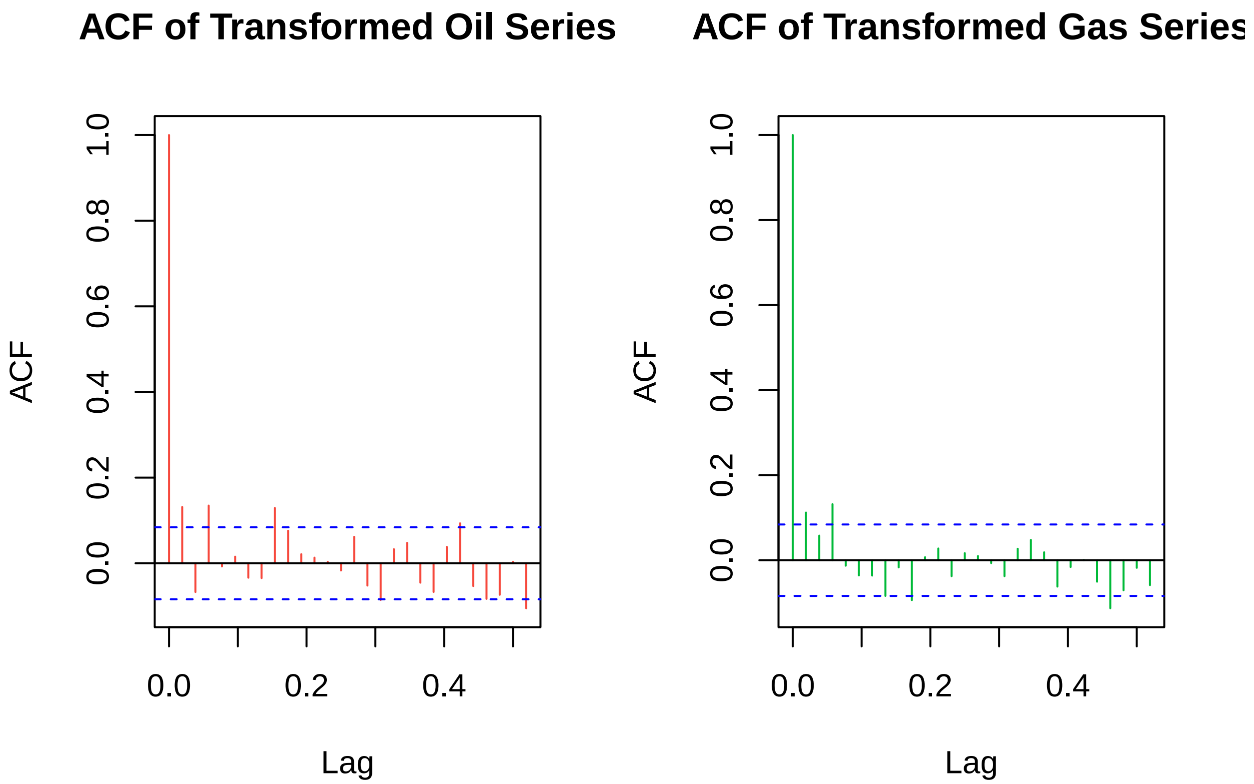
```r
# Transform the data as described in part (2).
diff_oil <- diff(log(oil))
diff_gas <- diff(log(gas))

plot(diff_oil, main="Transformed Weekly Time Series of Oil and Gas", col=2, ylab = "")
lines(diff_gas, col=3)
legend("topleft", legend = c("diff(log(oil))", "diff(log(gas))"),
       col=c(2, 3), lty = 1:1)
```



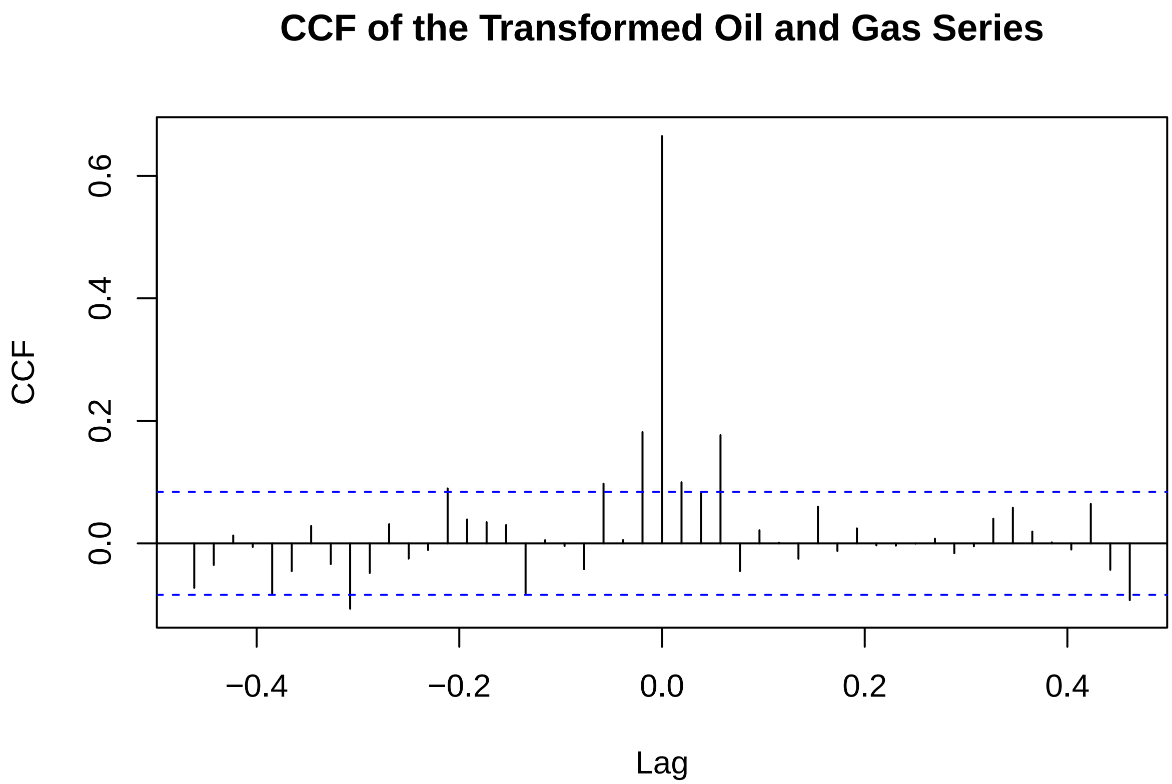
From the above superimposed data plot, the transformed data seems to be stationary with the exception of gas prices near 2006 and oil prices in around 2009.

```r
par(mfrow=c(1,2))
acf(diff_oil, main="ACF of Transformed Oil Series", col=2)
acf(diff_gas, main="ACF of Transformed Gas Series", col=3)
```



The above two sample ACF plots tend to be damped exponential dying down with oscillation, indicating the two series are relatively stationary.

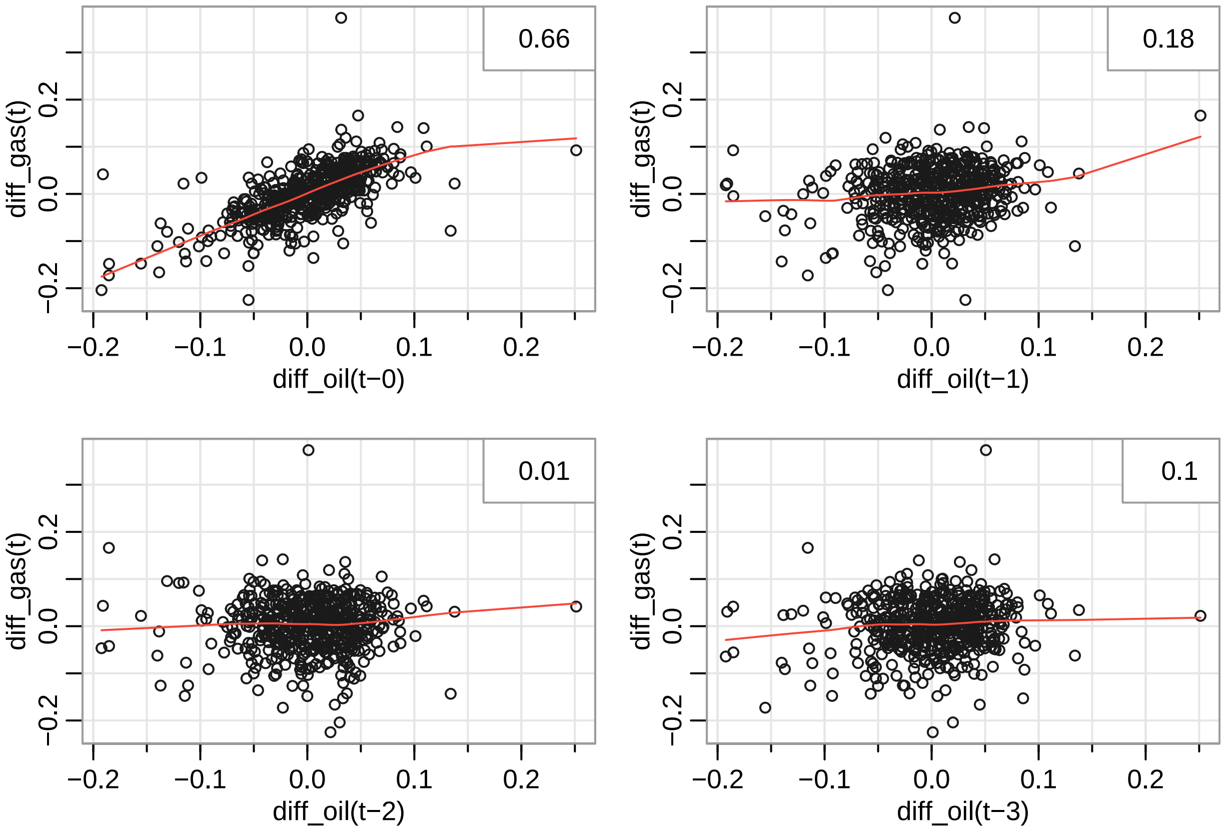
```r
ccf(diff_oil, diff_gas, ylab="CCF", main="CCF of the Transformed Oil and Gas Series")
```



The CCF plot shows that the two series are highly correlated. The most dominant cross correlation peaks at greater than 0.6 when the lag is zero. This means that the correlations in this region are positive where an above average oil price is likely to be correlated to an above average gas price at the same time.

4.

```
lag2.plot(diff_oil, diff_gas, 3)  # up to three weeks lag
```



The above scatterplots show the gas growth rate series against the lagged oil growth rate series at lag up to three weeks, with sample cross-correlations in the upper right corner and the Lowess fits in red. They indicate a strong linear relationship between gas growth rate and oil growth rate at lag zero with a sample cross-correlation of 0.66, as well as positive but weaker linear relationships with sample cross-correlations of 0.18, 0.01, and 0.1 at lags $h = 1$, 2, and 3 weeks, respectively.

5.

5 (i)

```
dummy <- ifelse(diff_oil < 0, 0, 1) # indicator of increase in the oil price, I_t
# Combine time series data as a data frame
merged_data <- ts.intersect(diff_gas, diff_oil, oilL1 = lag(diff_oil,-1), dummy, dframe=TRUE)
model1 <- lm(diff_gas ~ dummy + diff_oil + oilL1, data = merged_data)
summary(model1)
```

```
##
## Call:
## lm(formula = diff_gas ~ dummy + diff_oil + oilL1, data = merged_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18451 -0.02161 -0.00038  0.02176  0.34342
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.006445   0.003464  -1.860  0.06338 .
## dummy        0.012368   0.005516   2.242  0.02534 *
## diff_oil     0.683127   0.058369  11.704  < 2e-16 ***
## oilL1        0.111927   0.038554   2.903  0.00385 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04169 on 539 degrees of freedom
## Multiple R-squared:  0.4563, Adjusted R-squared:  0.4532
## F-statistic: 150.8 on 3 and 539 DF,  p-value: < 2.2e-16
```

Based on the regression summary result, we can see that all three predictor variables are significant with p-values less than 0.05, indicating this model is significant.

5 (ii)

The fitted model when there is negative growth in oil price at time t ($I_t = 0$):

$$G_t = -0.006445 + 0.683127O_t + 0.111927O_{t-1} + w_t$$

The fitted model when there is no or positive growth in oil price at time t ($I_t = 1$):

$$\begin{aligned} G_t &= -0.006445 + 0.012368 + 0.683127O_t + 0.111927O_{t-1} + w_t \\ &= 0.005923 + 0.683127O_t + 0.111927O_{t-1} + w_t \end{aligned}$$

From the above two fitted models, we can find that they are having same slope coefficients for O_t and O_{t-1}, but different intercepts. The fitted model when there is no or positive growth in oil price has a positive intercept 0.005923, while the fitted model when there is negative growth in oil price has a negative intercept -0.006445, indicating that the percentage change in gas price tends to have greater change when the oil price increases, holding other conditions constant. That is, the gas prices respond more quickly when oil prices are rising than when oil prices are falling. Thus, the results support the asymmetry hypothesis.

5 (iii)

```r
model2 <- lm(diff_gas ~ diff_oil, data = merged_data)
summary(model2)
```

```
##
## Call:
## lm(formula = diff_gas ~ diff_oil, data = merged_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18529 -0.02026  0.00049  0.02125  0.34777
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0004007  0.0018105   0.221    0.825
## diff_oil    0.7969128  0.0384920  20.703   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04216 on 541 degrees of freedom
```

```
## Multiple R-squared:  0.4421, Adjusted R-squared:  0.441
## F-statistic: 428.6 on 1 and 541 DF,  p-value: < 2.2e-16
```

```r
n <- length(diff_gas)  # sample size
# compute AIC
AIC_model1 <- AIC(model1)/n - log(2*pi)
AIC_model2 <- AIC(model2)/n - log(2*pi)
# compute BIC
BIC_model1 <- BIC(model1)/n - log(2*pi)
BIC_model2 <- BIC(model2)/n - log(2*pi)
# generate a table for comparison
model <- c("full", "reduced")
AIC <- c(round(AIC_model1, 4), round(AIC_model2, 4))
BIC <- c(round(BIC_model1, 4), round(BIC_model2, 4))
knitr::kable(cbind(model, AIC, BIC), caption="Model Comparison")
```

Table 1: Model Comparison

model	AIC	BIC
full	-5.3373	-5.2978
reduced	-5.3189	-5.2952

As the above table shows, both the AIC and BIC for this reduced model are bigger than those of the full model in part (i). Thus, this model is NOT better than the previous one.

5 (iv)

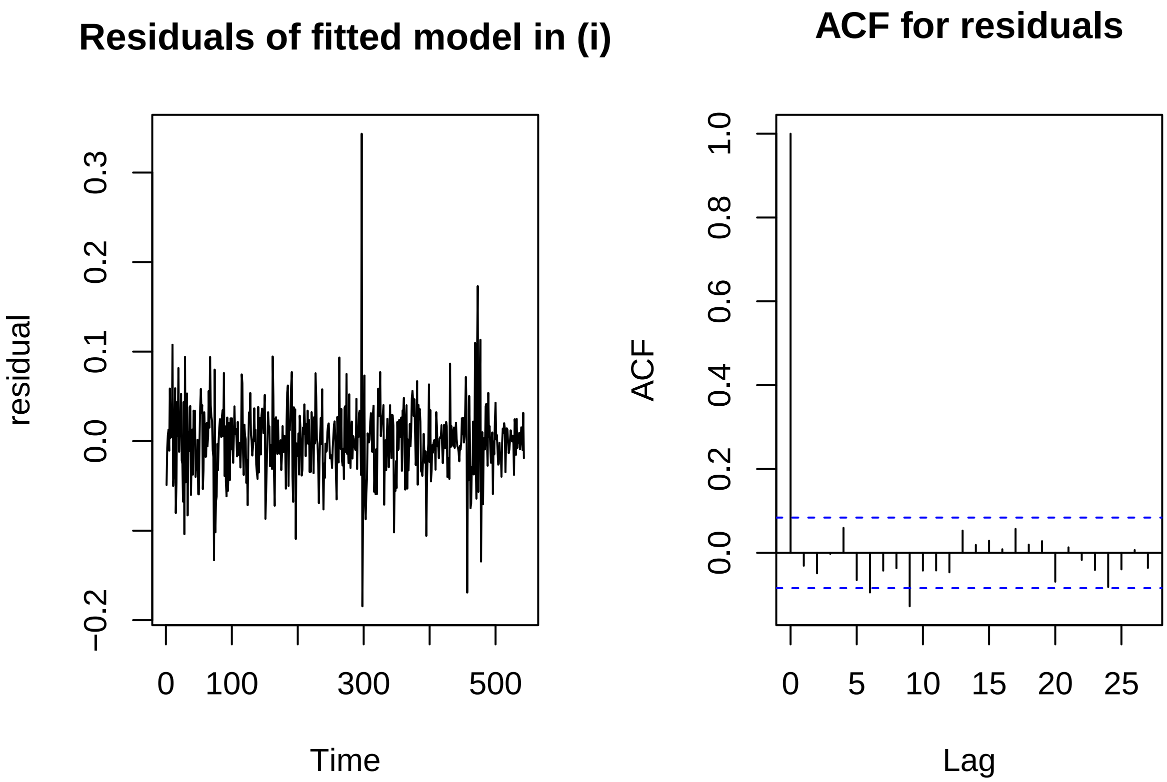
```r
# Conduct partial F test.
anova(model2, model1)
```

```
## Analysis of Variance Table
##
## Model 1: diff_gas ~ diff_oil
## Model 2: diff_gas ~ dummy + diff_oil + oilL1
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1    541 0.96149
## 2    539 0.93700  2  0.024488 7.0433 0.0009559 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the output we can see that the F test-statistic from the ANOVA is 7.0433 and the corresponding p-value is 0.0009559. Since this p-value is less than 0.05, we will reject the null hypothesis. This means that we have sufficient evidence to say that at least one of the predictor variables I_t and O_{t-1} are statistically significant. In other words, adding dummy variable of increasing in the oil price and one-lagged oil price change to the regression model can significantly improve the fit of the model.

5 (v)

```
par(mfrow=c(1,2))
plot(ts(resid(model1)), main="Residuals of fitted model in (i)", ylab="residual")
acf(resid(model1), main="ACF for residuals")
```



The above residual plot on the left shows an approximate white noise series with an exception at time of around 300. To further confirm this, we plot the ACF of the residuals on the right. It shows that the ACF of the residual is indeed approximately following a white noise, which indicates a good fitness of the model.

Q4

1.

```r
trend <- time(jj) - 1970  # center the time (the middle of 1960 and 1980)
Q = factor(cycle(jj))  # quarter effect -> 4 levels Q1, Q2, Q3, Q4
model_jj <- lm(log(jj) ~ 0 + trend + Q, na.action=NULL)  # without intercept
summary(model_jj)
```

```
##
## Call:
## lm(formula = log(jj) ~ 0 + trend + Q, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29318 -0.09062 -0.01180  0.08460  0.27644
##
## Coefficients:
##       Estimate Std. Error t value Pr(>|t|)
## trend 0.167172   0.002259   74.00   <2e-16 ***
## Q1    1.052793   0.027359   38.48   <2e-16 ***
## Q2    1.080916   0.027365   39.50   <2e-16 ***
## Q3    1.151024   0.027383   42.03   <2e-16 ***
## Q4    0.882266   0.027412   32.19   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1254 on 79 degrees of freedom
## Multiple R-squared:  0.9935, Adjusted R-squared:  0.9931
## F-statistic:  2407 on 5 and 79 DF,  p-value: < 2.2e-16
```

With the above result, we can write the fitted model as

$$x_t = 0.167172 + 1.052793Q1(t) + 1.080916Q2(t) + 1.151024Q3(t) + 0.882266Q4(t) + w_t$$

, with a residual standard error of 0.1254 on 79 degrees of freedom.

2.

Assume the model is correct, the estimated average annual increase in the logged earnings per share should be the sum of estimated coefficient values of each quarter. That is,

$$1.052793 + 1.080916 + 1.151024 + 0.882266 = 4.166999$$

3.

Assume the model is correct, the average logged earnings rate percentage change from the third quarter to the fourth quarter should be:

$$\frac{\alpha_4 - \alpha_3}{\alpha_3} \times 100\% = \frac{0.882266 - 1.151024}{1.151024} \approx -23.349\%$$

Thus, the average logged earnings rate decreases by 23.349% from the third quarter to the fourth quarter.

4.

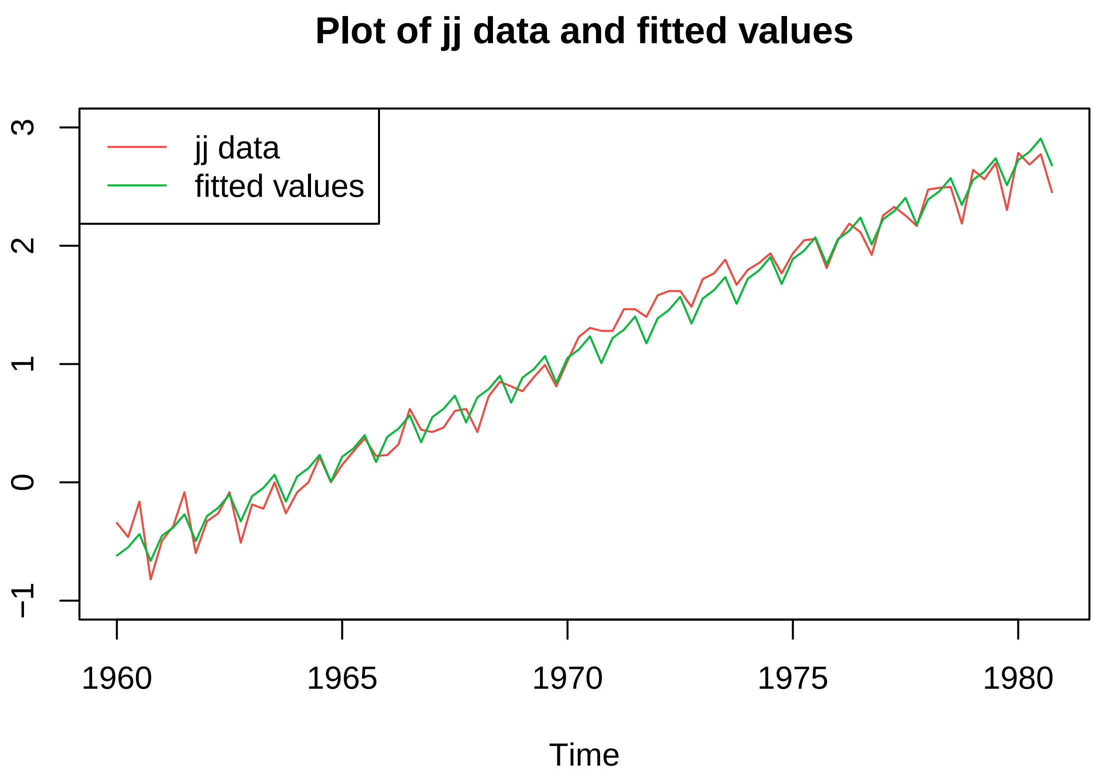
```r
# add intercept to model in part 1
model_intercept <- lm(log(jj) ~ trend + Q, na.action=NULL)  # with intercept
summary(model_intercept)
```

```
##
## Call:
## lm(formula = log(jj) ~ trend + Q, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29318 -0.09062 -0.01180  0.08460  0.27644
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.052793   0.027359  38.480  < 2e-16 ***
## trend        0.167172   0.002259  73.999  < 2e-16 ***
## Q2           0.028123   0.038696   0.727   0.4695
## Q3           0.098231   0.038708   2.538   0.0131 *
## Q4          -0.170527   0.038729  -4.403 3.31e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1254 on 79 degrees of freedom
## Multiple R-squared:  0.9859, Adjusted R-squared:  0.9852
## F-statistic:  1379 on 4 and 79 DF,  p-value: < 2.2e-16
```

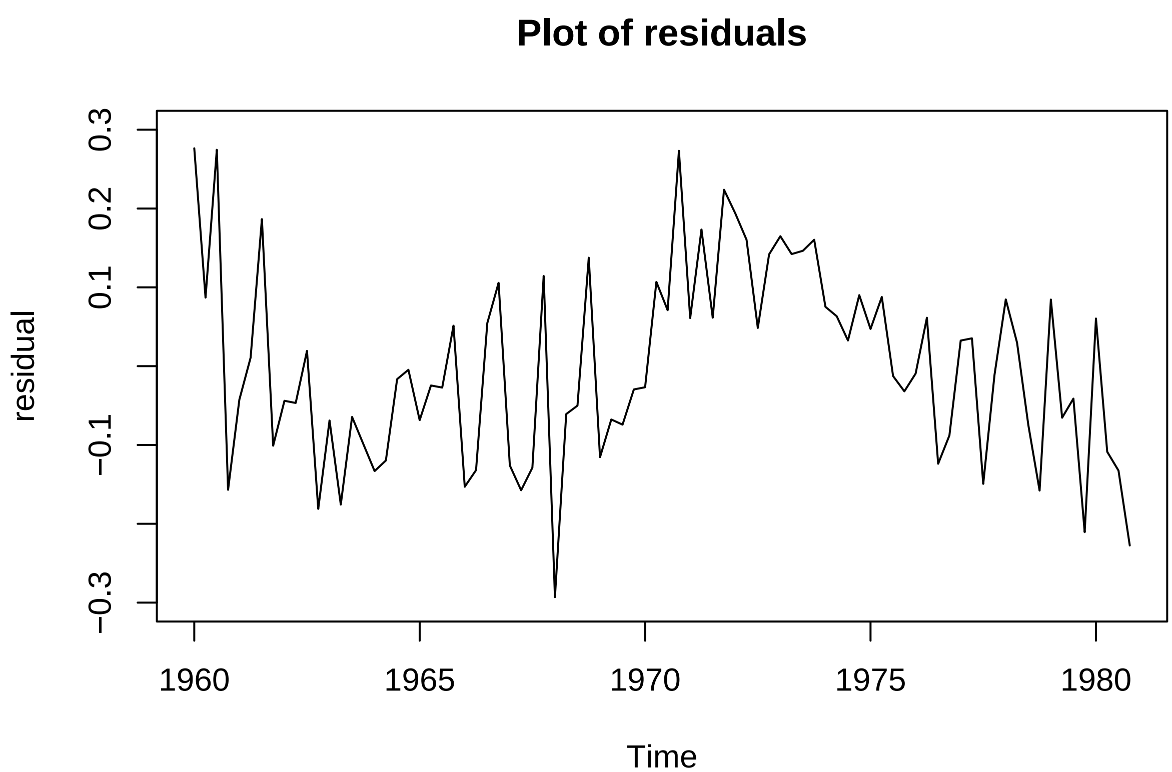
As the above result shows, the first quarter effect Q1 is missing and this affects estimated coefficients of the other three quarters, which does not make sense since our target is to include the evaluation of the seasonal effect.

5.

```r
plot(log(jj), main="Plot of jj data and fitted values", ylab="",
     col=2, ylim = c(-1, 3))  # original data xt
lines(fitted(model_jj), col=3)  # superimpose fitted values
legend("topleft", legend = c("jj data", "fitted values"),
       col=c(2, 3), lty = 1:1)
```




```
plot(log(jj) - fitted(model_jj), main="Plot of residuals", ylab="residual", ylim=c(-0.3, 0.3))
```



The above residual plot shows that the noise seems to have a constant variance without following any pattern. Thus, the residuals look white and the model fits the data well.