Assignment1_Fall2022

Alexander Tran

9/28/2022

Q1

Let $x > 0, x \in \mathbb{R}$.

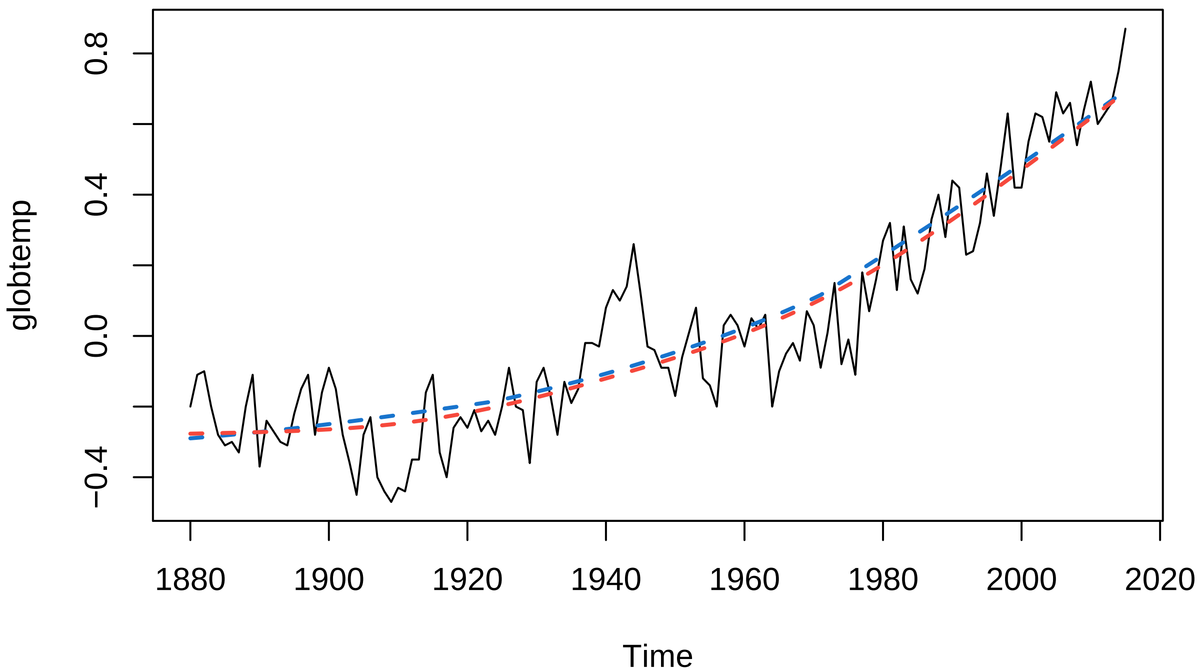
$$\lim_{\lambda \to 0} \frac{x^\lambda - 1}{\lambda} = \frac{x^0 - 1}{0} = \frac{1 - 1}{0} = \frac{0}{0}$$

Given that we have an indeterminate form $\frac{0}{0}$, we can use L'Hôpital's rule to differentiate the numerator and denominator and evaluate the limit as follows:

$$\begin{aligned}
\lim_{\lambda \to 0} \frac{x^\lambda - 1}{\lambda} &= \lim_{\lambda \to 0} \frac{x^\lambda \log(x) - 0}{1} \\
&= x^0 \log(x) \\
&= \log(x)
\end{aligned}$$

Q2

```r
plot(globtemp)
lines(lowess(time(globtemp),globtemp), lty=2,lwd=2,col=4)
lines(smooth.spline(time(globtemp),globtemp,spar=1),lty=2,lwd=2,col=2)
```

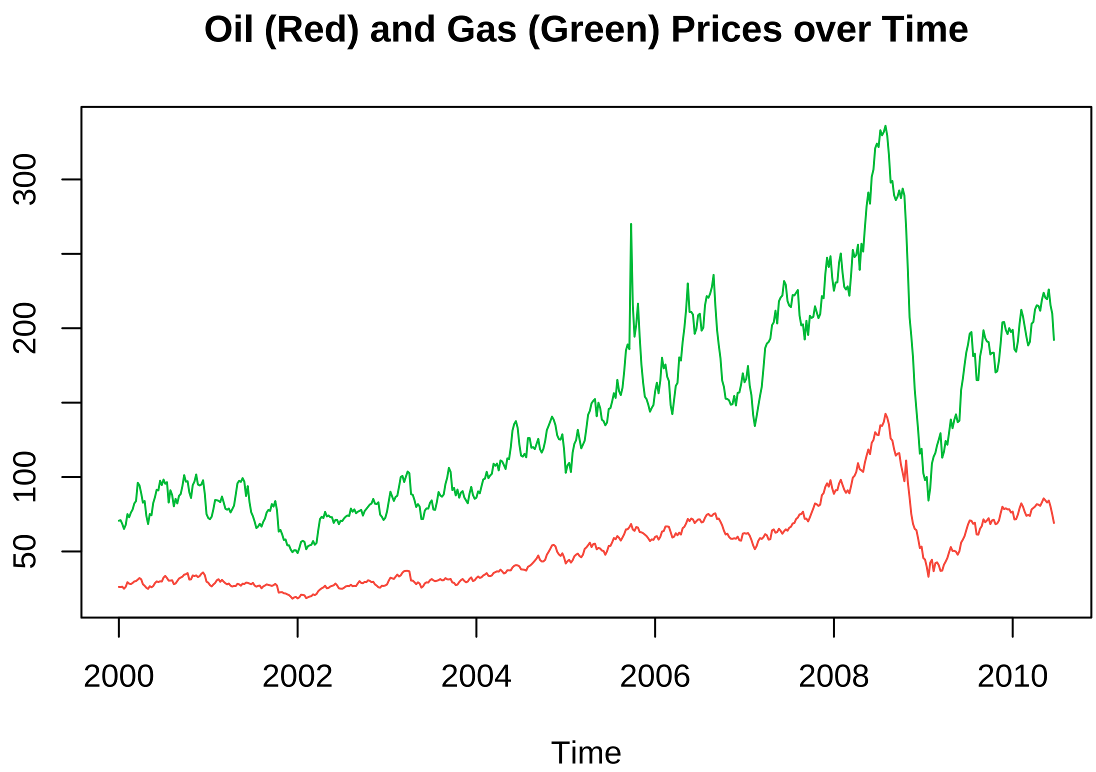


Both Lowess smoothing and smoothing splines seem to do a good job at estimating the trend. They both show that there is an increasing trend in global temperature over time, with a seemingly increasing slope.

Q3

1.

```
ts.plot(oil, gas, col=2:3, main="Oil (Red) and Gas (Green) Prices over Time")
```



I do not believe that these series are stationary as the means are not constant over time. It is clear to see that the means from 2000-2006 are very different from the means from 2006-2010.

2.

Let y_t be the percentage change in price from x_{t-1} to x_t.

$$y_t = \frac{x_t - x_{t-1}}{x_{t-1}}$$

This can be rewritten into:

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

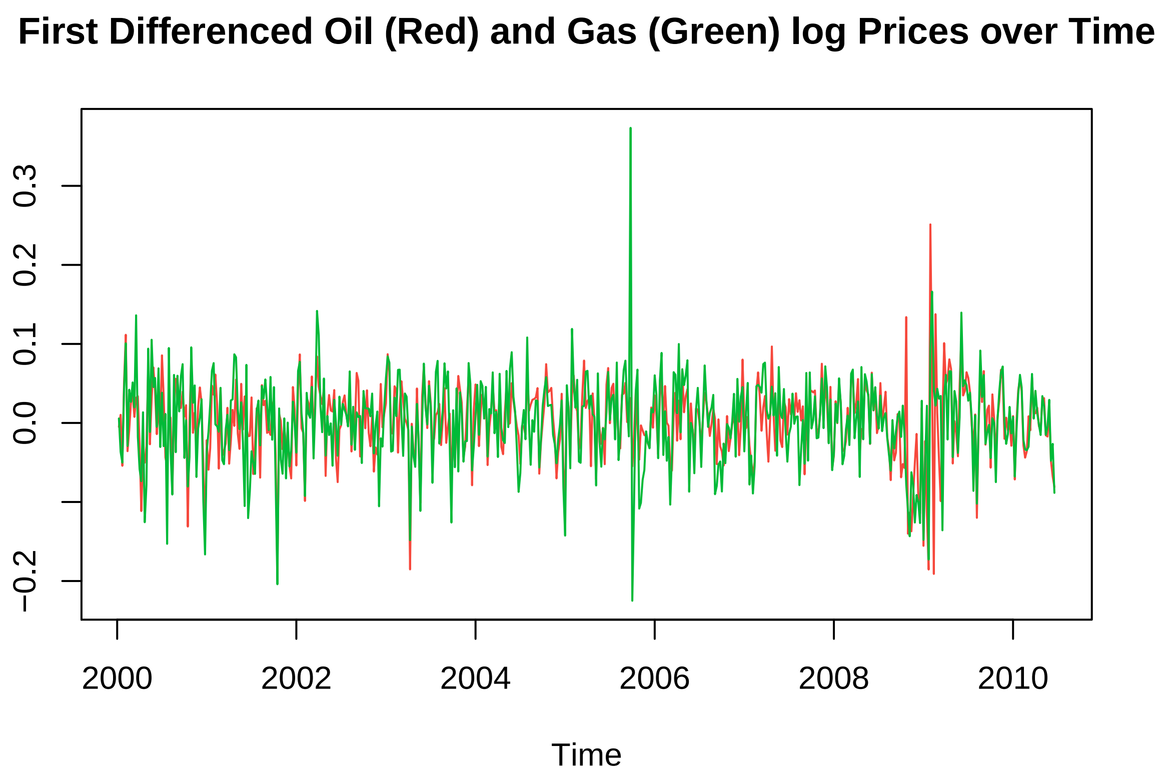
In a price series, we observe $|y_t|$ near zero. As stated in the textbook (Shumway, Stoffer), "For $|p|$ close to zero, $\log(1+p) \approx p$; let $p = (y_t - y_{t-1})/y_{t-1}$."

Thus,

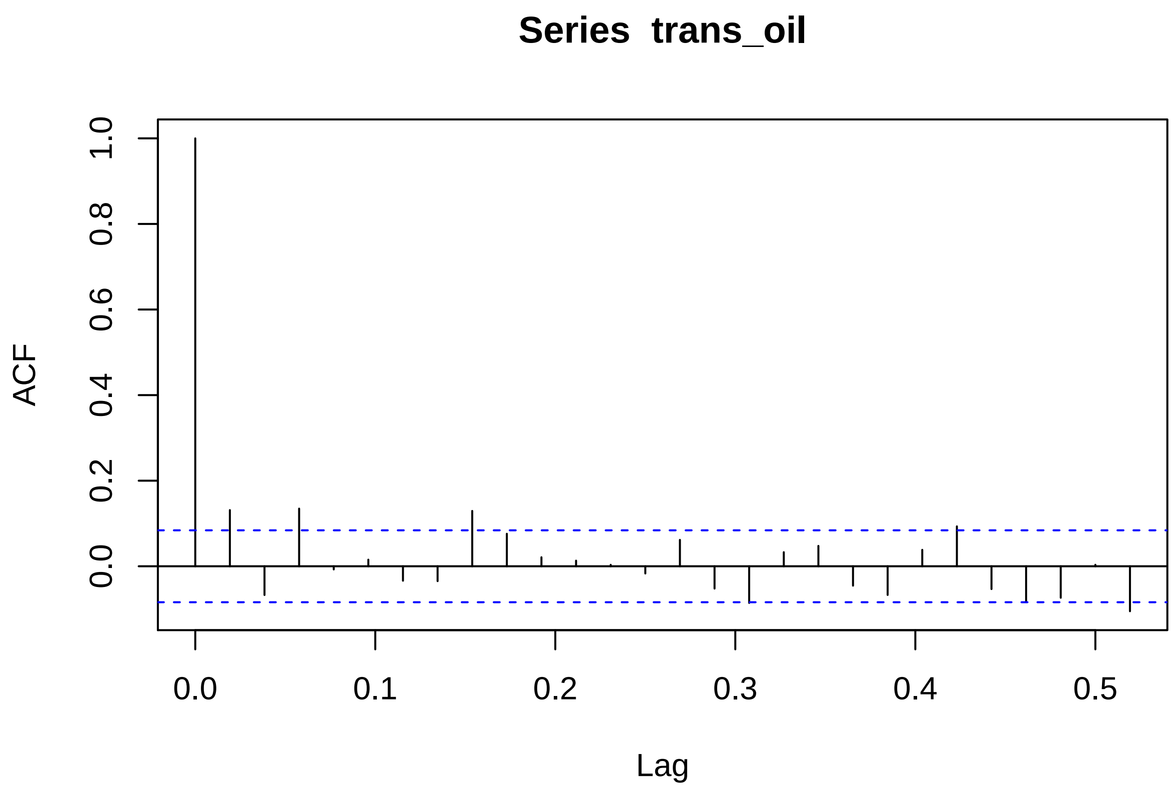
$$
\begin{aligned}
\log(y_t + 1) &= \log x_t - \log x_{t-1} \\
y_t &= \log x_t - \log x_{t-1} \\
y_t &= \nabla \log x_t
\end{aligned}
$$

3.

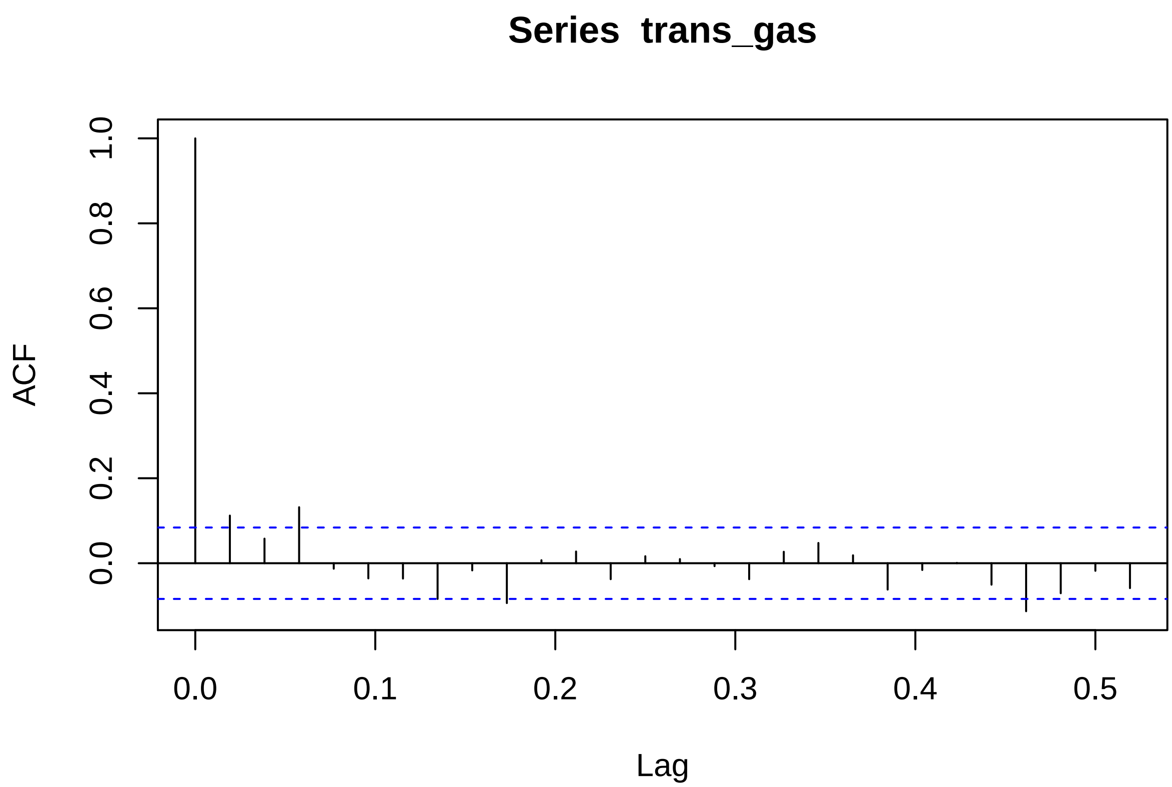
```
trans_oil <- diff(log(oil))
trans_gas <- diff(log(gas))
ts.plot(trans_oil, trans_gas, col=2:3, main="First Differenced Oil (Red) and Gas (Green) log Prices ove
```



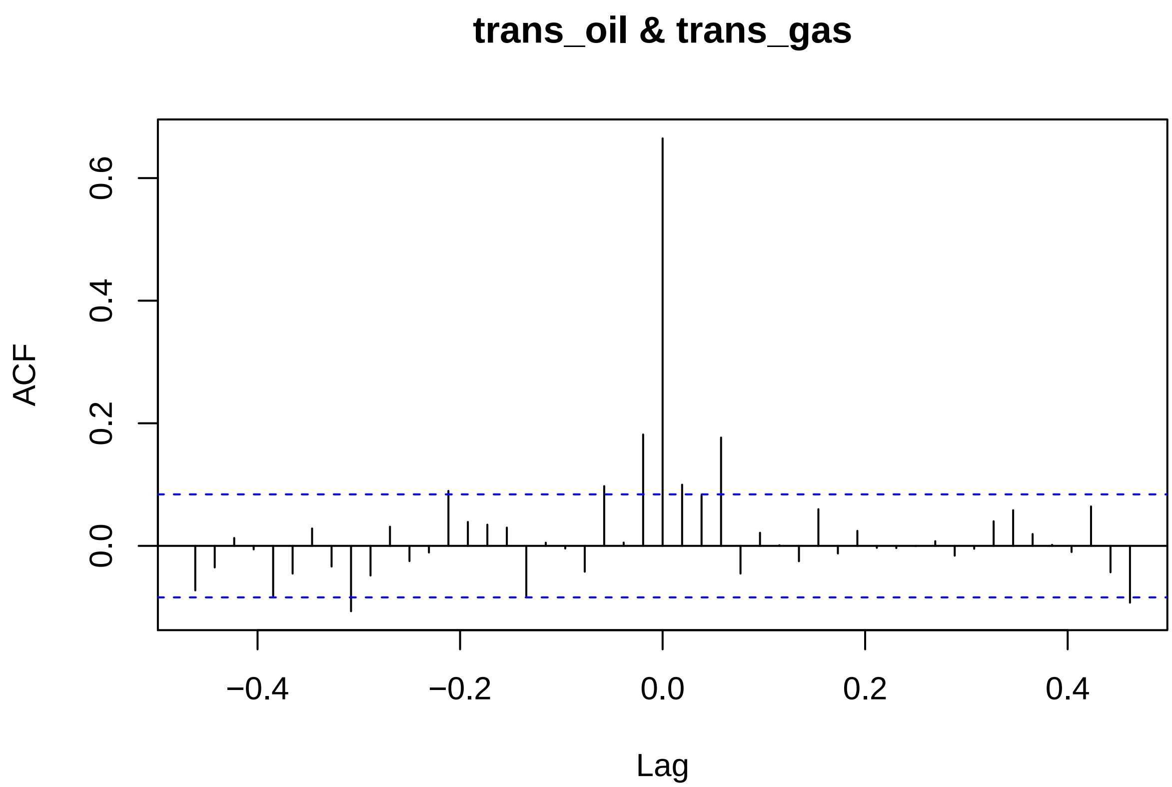

```
acf(trans_oil)
```




```
acf(trans_gas)
```




```r
ccf(trans_oil, trans_gas)
```



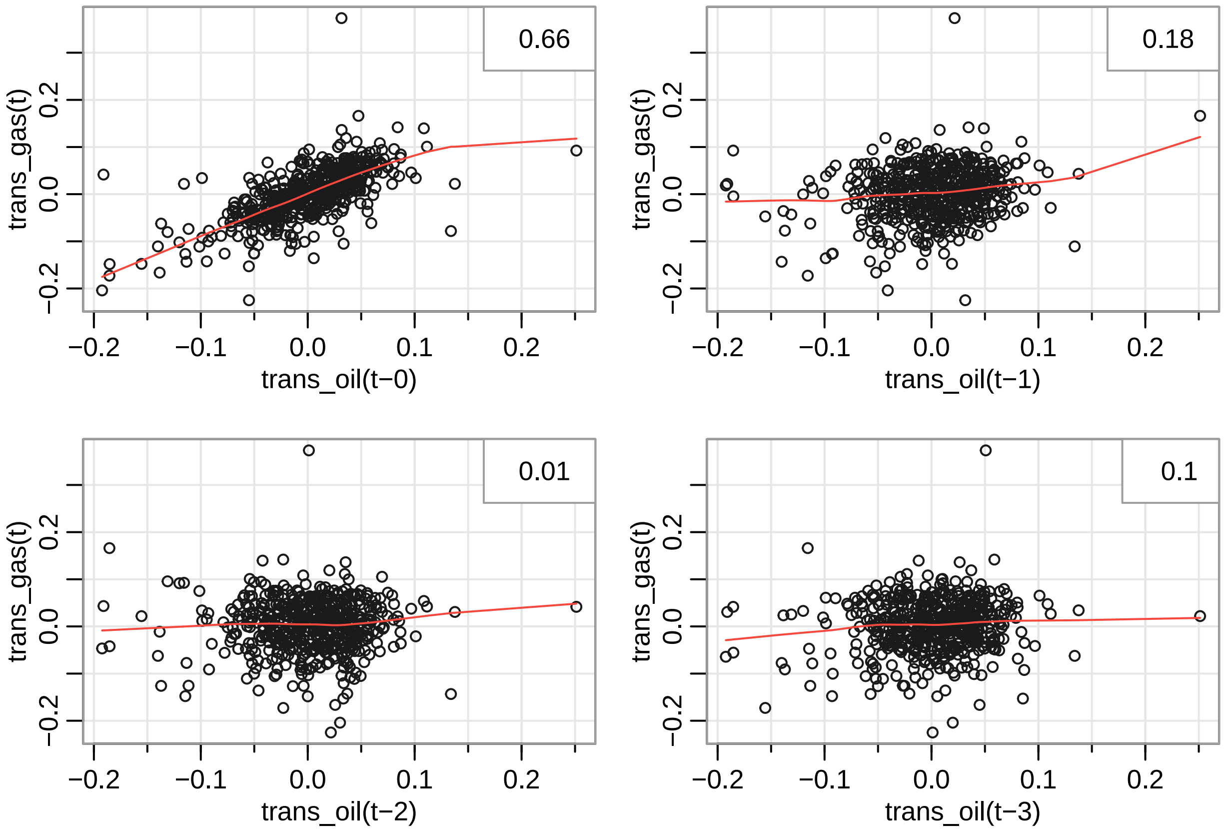
The transformed data appears to be mostly stationary despite the fact that there are unusual spikes in 2009 for oil and 2006 for gas.

The ACF plots both cut off quickly and behave like white noise in their correlation. Knowing this, we can say with confidence that the transformed series are stationary.

The CCF plot shows that the two series are highly correlated. At lag zero we observe the largest correlation of over 0.6, as well as significant correlations at other lags such as -3, -1, 1, and 3. Furthermore, the correlations are positive, meaning the prices of oil and gas generally move together. High oil prices mean high gas prices, low oil prices mean low gas prices, and vice versa.

4.

```
lag2.plot(trans_oil, trans_gas, 3)
```



The plots show that there is a strong linear relationship between gas and oil at lag zero with a correlation of 0.66. At lag 1 we observe a weakly positive linear relationship with a correlation of 0.18, and at lags 2 and 3 the relationship is either very weak or non-existent. There also appear to be a handful of outliers in the plots, some of which can be identified as the spikes seen in 2006 and 2009.

5.

i)

```
dummy = ifelse (trans_oil < 0, 0, 1)
trans_oil_lag1 <- stats::lag(trans_oil, -1)
prices = ts.intersect(trans_gas, dummy, trans_oil, trans_oil_lag1, dframe=TRUE)
reg1 <- lm(trans_gas ~ dummy + trans_oil + trans_oil_lag1, data=prices, na.action=NULL)
summary(reg1)
```

```
##
## Call:
## lm(formula = trans_gas ~ dummy + trans_oil + trans_oil_lag1,
##     data = prices, na.action = NULL)
##
## 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 *
```

```
## trans_oil       0.683127   0.058369  11.704  < 2e-16 ***
## trans_oil_lag1  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
```

The results show that the coefficients of all predictor variables are significant at at least the 0.05 level, except for the intercept. The coefficients for β_1, β_2, and β_3 are all positive, which suggest that the gas growth rate moves in the same direction with oil growth rates at both lags 0 and 1, even more so when the gas growth rate is non-negative.

ii) The fitted model when there is negative growth in oil price at time t is given by:

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

The fitted model when there is no or positive growth in oil price is given by:

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

These results support the asymmetry hypothesis, as all coefficients are positive in the case where the oil growth rate is non-negative, whereas there is a negative coefficient when the oil growth rate is negative. In other words, gasoline prices respond more quickly when oil prices are rising than when oil prices are falling.

iii)

```r
reg2 <- lm(trans_gas ~ trans_oil, data=prices, na.action=NULL)

anova(reg1)
```

```
## Analysis of Variance Table
##
## Response: trans_gas
##                  Df  Sum Sq Mean Sq F value    Pr(>F)
## dummy             1 0.52668 0.52668 302.968 < 2.2e-16 ***
## trans_oil         1 0.24493 0.24493 140.896 < 2.2e-16 ***
## trans_oil_lag1    1 0.01465 0.01465   8.428  0.003846 **
## Residuals       539 0.93700 0.00174
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(reg2)
```

```
## Analysis of Variance Table
##
## Response: trans_gas
##            Df  Sum Sq Mean Sq F value    Pr(>F)
## trans_oil   1 0.76178 0.76178  428.63 < 2.2e-16 ***
## Residuals 541 0.96149 0.00178
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For the regression $G_t = \beta_0 + \beta_1 I_t + \beta_2 O_t + \beta_3 O_{t-1} + w_t$, the AIC is given by $\log(\frac{SSE}{n}) + \frac{n+2k}{n} = \log(\frac{0.93700}{543}) + \frac{543+2(5)}{543} = -5.343765$, and the BIC is given by $\log(\frac{SSE}{n}) + \frac{k \log n}{n} = \log(\frac{0.93700}{543}) + \frac{5 \log(543)}{543} = -6.304197$

For the regression $G_t = \beta_0 + \beta_1 O_t + w_t$, the AIC is given by $\log(\frac{SSE}{n}) + \frac{n+2k}{n} = \log(\frac{0.96149}{543}) + \frac{543+2(3)}{543} = -5.325331$, and the BIC is given by $\log(\frac{SSE}{n}) + \frac{k \log n}{n} = \log(\frac{0.96149}{543}) + \frac{3 \log(543)}{543} = -6.30159$

The values of the AIC and BIC of the first regression model with more parameters are slightly lesser than those of the second model, suggesting that its predictive power is slightly better. However, because the difference is so small even with the addition of two more parameters, I believe that the second model may be preferred. This is further explored in part (iv).

iv)

```r
anova(reg1, reg2)
```

```
## Analysis of Variance Table
##
## Model 1: trans_gas ~ dummy + trans_oil + trans_oil_lag1
## Model 2: trans_gas ~ trans_oil
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1    539 0.93700
## 2    541 0.96149 -2 -0.024488 7.0433 0.0009559 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

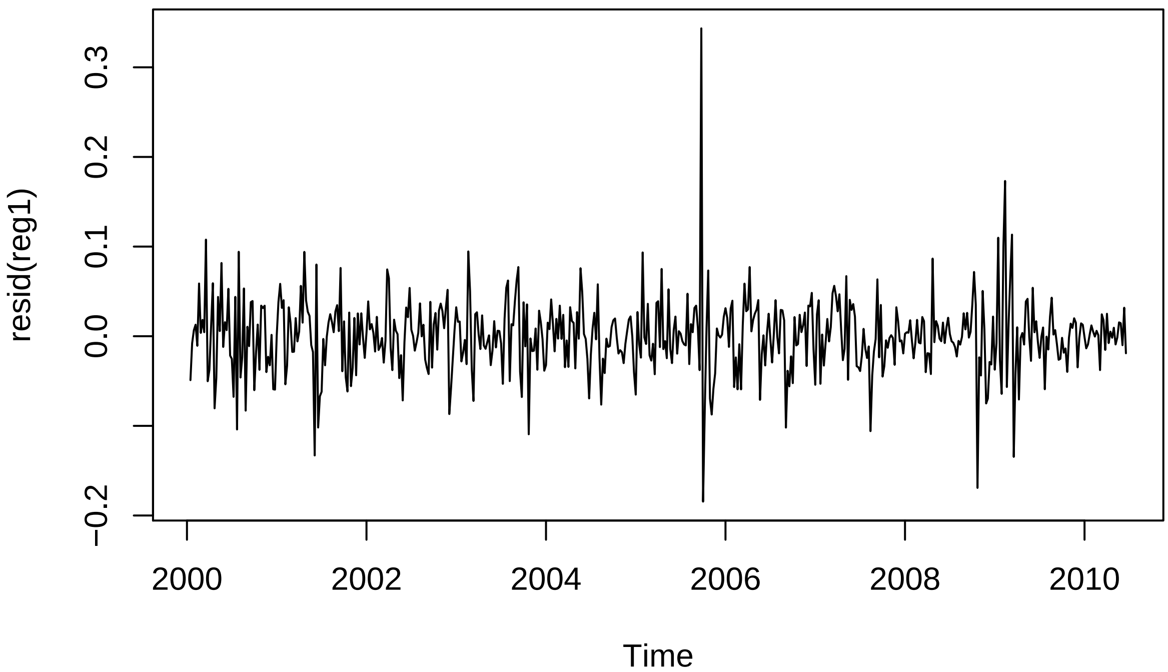
The p-value given by the ANOVA table shows that adding the two provides a significantly better fit. Taking this into account on top of what was found in part (iii), the full model is preferred over the reduced model.

v)

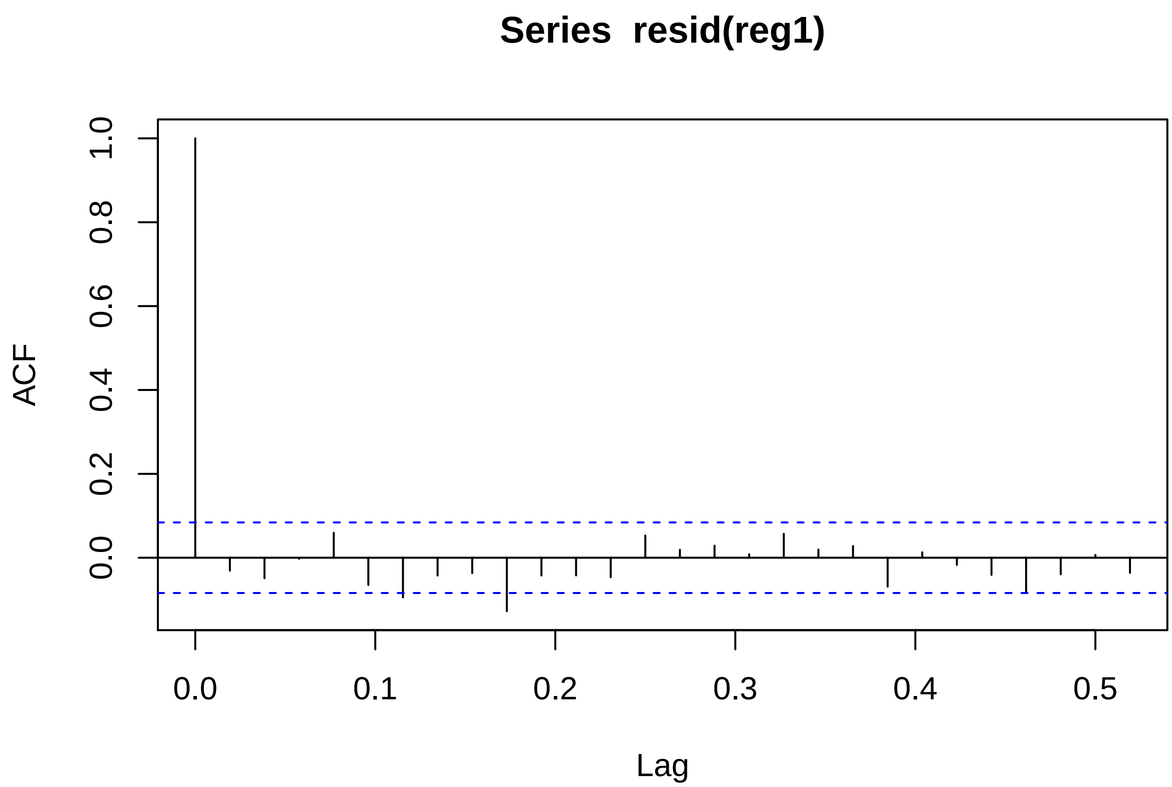
```r
summary(reg1)$sigma
```

```
## [1] 0.0416942
```

```r
plot(resid(reg1))
```




```
acf(resid(reg1))
```



The residual standard error is 0.04169 on 539 degrees of freedom, which is not very large considering the units of the variables that we are working with. The residuals time series plot behaves mostly like white noise, which is further evidenced by the ACF plot, which cuts off very quickly. Knowing this, we can say with confidence that any unexplained variation in the model is simply due to random white noise.