Econ 104 Project 2 Group 12

Seungyeon Yoo, AJ Jeng, Youxue Xia, Kevin Carranza

2023-11-17

Contents

```r
library(AER)
data("PepperPrice")
```

1. Exploratory Data Analysis

(a) Briefly discuss the question you are trying to answer.

We are attempting to model the prices of black and white pepper based on the prices of the black and white variants of pepper from 1973 to 1996.
The two variables are the spot prices for black pepper and white pepper respectively.

(b) Cite the dataset and give a summary of what the dataset is about

```r
citation(package = "AER")
```

```
## To cite AER, please use:
##
##   Christian Kleiber and Achim Zeileis (2008). Applied Econometrics with
##   R. New York: Springer-Verlag. ISBN 978-0-387-77316-2. URL
##   https://CRAN.R-project.org/package=AER
##
## A BibTeX entry for LaTeX users is
##
##   @Book{,
##     title = {Applied Econometrics with {R}},
##     author = {Christian Kleiber and Achim Zeileis},
##     year = {2008},
##     publisher = {Springer-Verlag},
##     address = {New York},
##     note = {{ISBN} 978-0-387-77316-2},
##     url = {https://CRAN.R-project.org/package=AER},
##   }
```

As introduced earlier, the data set PepperPrice is a time series of the average monthly European spot prices for black and white pepper in US dollars per ton from 1973 to 1996. These are the two previously mentioned variables:

`black` - spot price for black pepper

`white` - spot price for white pepper

As a clarification, spot price is the price of the asset at that moment in time.

(c) First check for completeness and consistency of the data (if there are NAs or missing observations, replace with the value of the previous observation; make a note of this)

```r
# Checking for missing values
paste("There is", sum(is.na(PepperPrice)), "missing value(s)") # No NAs present
```

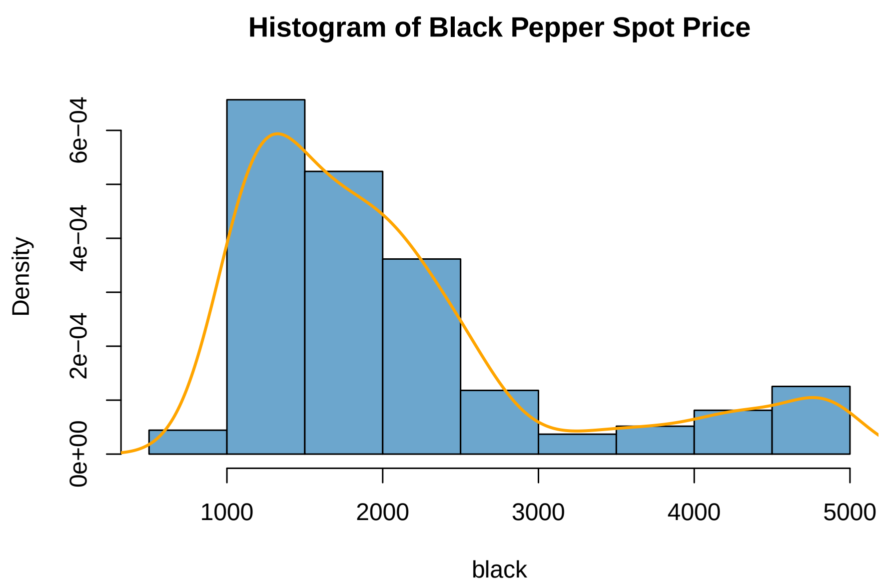
```
## [1] "There is 0 missing value(s)"
```

As seen in the output, there are no missing values.

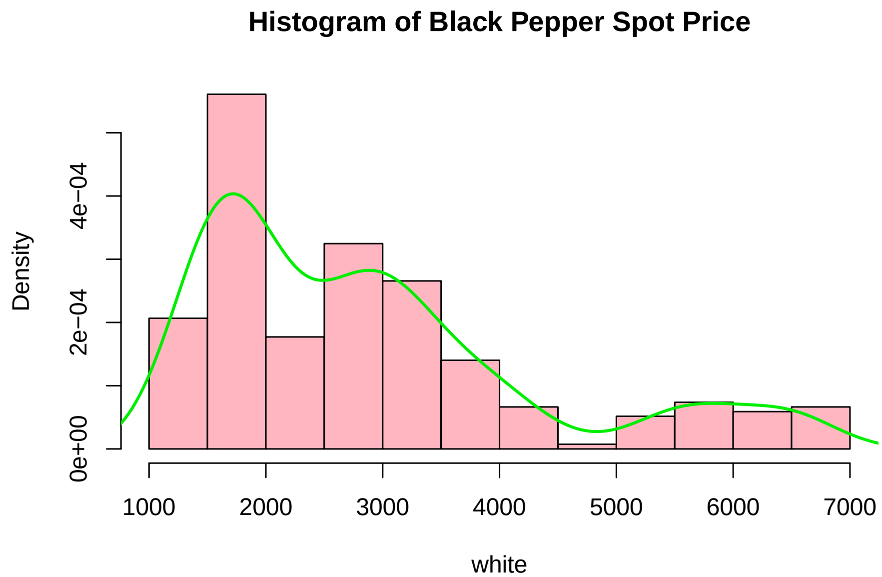
(d) Provide descriptive analyses of your variables. This should include the histogram with overlying density, boxplots, cross correlation. All figures/statistics must include comments.

```r
# Seperating two variables
black <- PepperPrice[,1]
white <- PepperPrice[,2]
```

```r
hist(black, prob = TRUE, col = "skyblue3",
     main = "Histogram of Black Pepper Spot Price")
lines(density(black), col = "orange", lwd = 2)
```




```r
hist(white, prob = TRUE, col = "lightpink",
     main = "Histogram of Black Pepper Spot Price")
lines(density(white), col = "green2", lwd = 2)
```



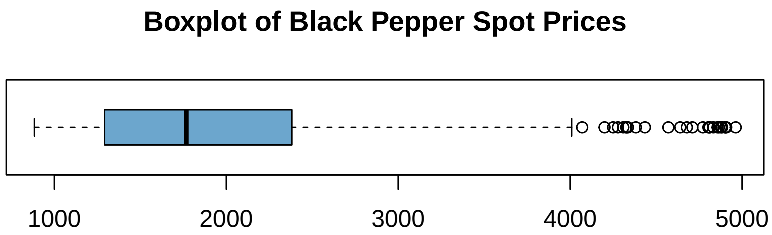
As seen in the histograms, both black and white pepper follow a similar distribution, albeit at different values. Both have high density at cheaper values in the 1000 to 2000 range then fall before slowly rising toward the more expensive values.

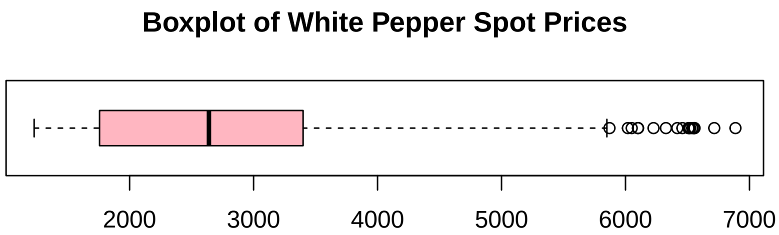
```r
#summary stats
summary(PepperPrice)
```

```
##      black          white     
##  Min.   : 884   Min.   :1230  
##  1st Qu.:1292   1st Qu.:1758  
##  Median :1768   Median :2640  
##  Mean   :2077   Mean   :2875  
##  3rd Qu.:2382   3rd Qu.:3400  
##  Max.   :4963   Max.   :6887
```

White pepper seems to have higher values for all metrics, suggesting that white pepper is usually more expensive than black pepper.

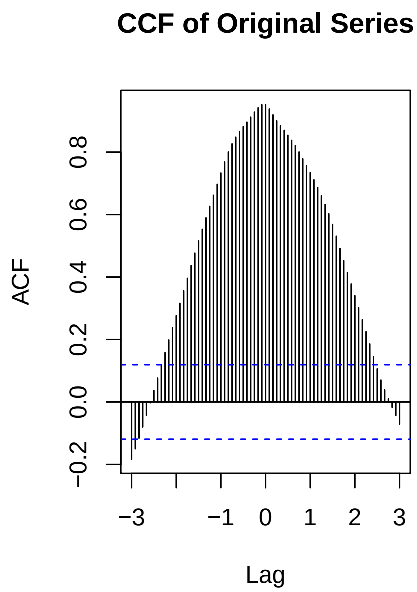
```r
# Boxplots
par(mfrow = c(2,1))
boxplot(black, horizontal = TRUE, col = "skyblue3",
        main = "Boxplot of Black Pepper Spot Prices")
boxplot(white, horizontal = TRUE, col = "lightpink",
        main = "Boxplot of White Pepper Spot Prices")
```

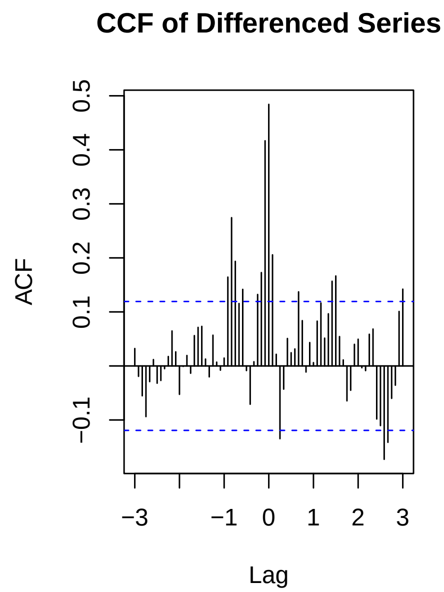




The box plots reflect what was discussed for the summary statistics and histograms. White pepper is more expensive than black, but both have relatively similar distributions.

```r
par(mfrow = c(1,2))
# ccf for original values
ccf(black, white, lag.max = 36,
    main = "CCF of Original Series")

# ccf for differenced values
ccf(diff(black), diff(white), lag.max = 36,
    main = "CCF of Differenced Series")
```





For the original series, we can observe that the cross correlation slowly decays as the lags are further away. This reflects the characteristics of long term trend or cyclic component in the original time series of Black Pepper and White Pepper spot price. This suggests that lagged values of the spot prices for black pepper can help explain the spot prices for white pepper and vice versa.

The CCF for the differenced series exhibits relatively short memory, which means that as the lags get further away, the correlation between lagged values quickly decreases or cuts off.

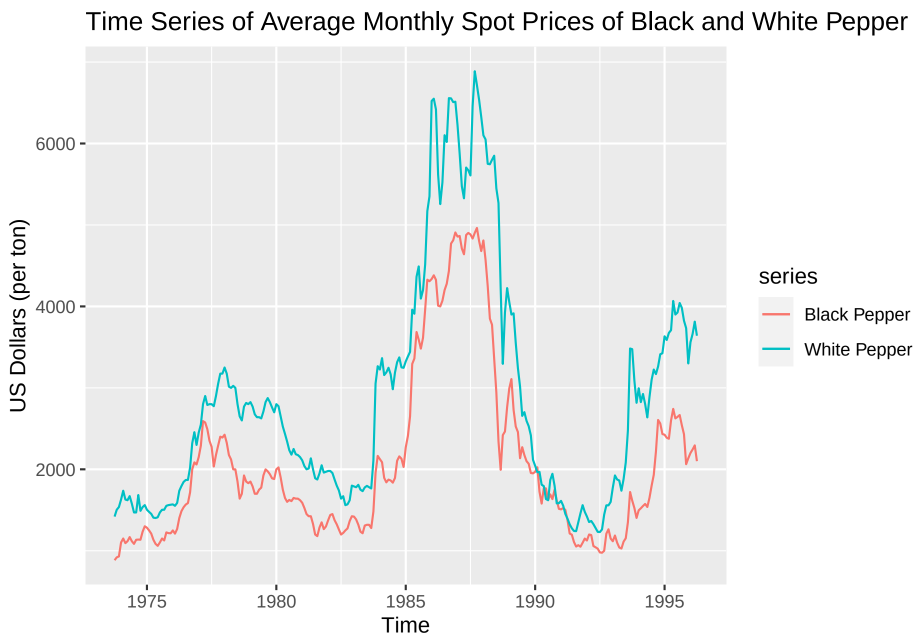
2. Data PreProcessing

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

```
# Seperating two variables
black <- PepperPrice[,1]
white <- PepperPrice[,2]
```

(a) Time Series Plots

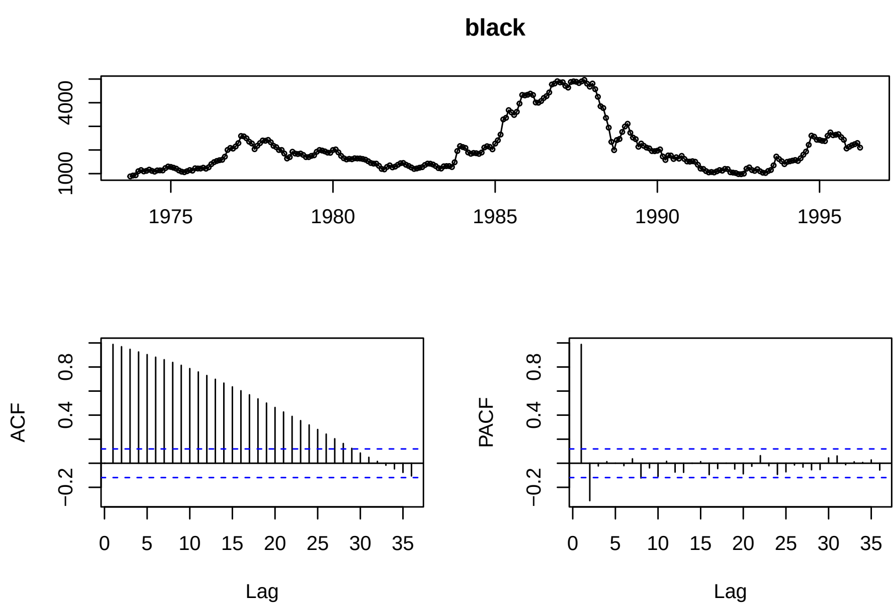
```
library(ggplot2)
autoplot(black, series = "Black Pepper") +
  autolayer(white, series = "White Pepper") +
  labs(title = "Time Series of Average Monthly Spot Prices of Black and White Pepper",
       y = "US Dollars (per ton)")
```



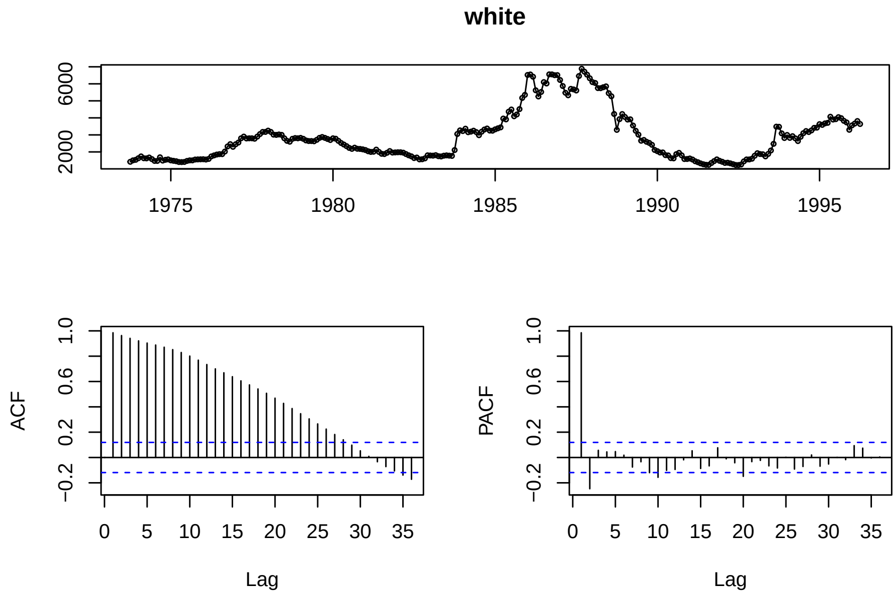
Plotting the two time series together, it is observed that two series behave very similarly. This suggests that there may be a Granger Causalty that we can examine. By further investigating the dynamics of the two time series, we may be able to identify the causal relationship between two variables, which will potentially help explain or predict the future values of the series.

```
tsdisplay(black)
```

ACF and PACF



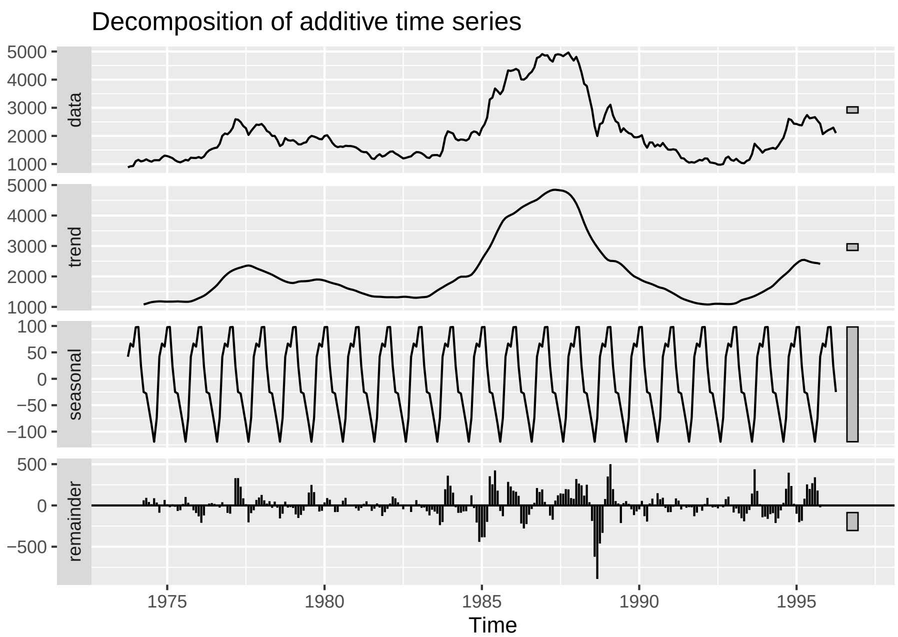
```r
tsdisplay(white)
```



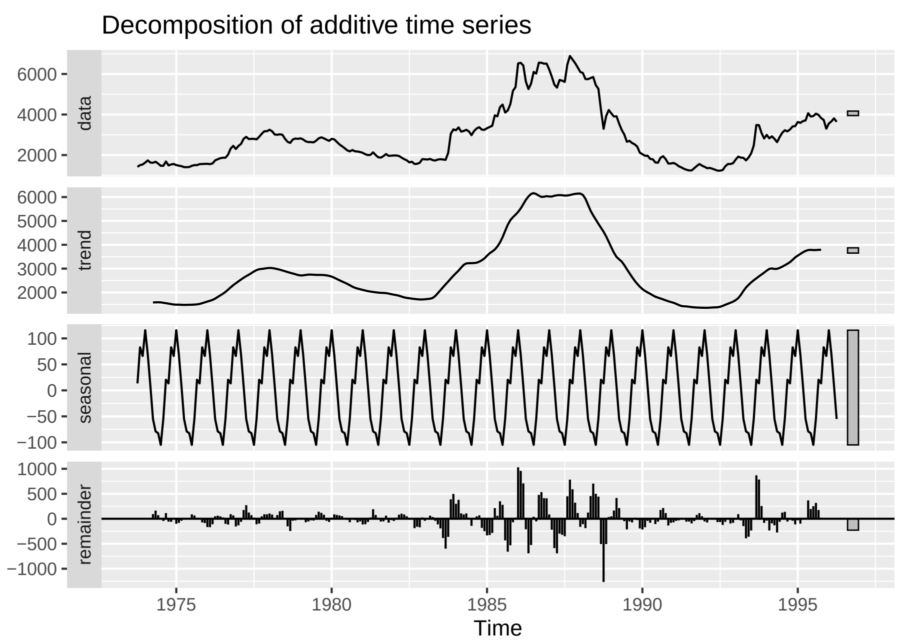
Using the ACF and PACF correlagrams, we suspect that AR(2) is the order for both variables, as the ACF exhibits a highly persistent and slowly decaying behavior and the second lag is the last one to be significant in PACF. All other lags after the second lags are below the significance lines.

```r
autoplot(decompose(black))
```

Time Series Decomposition




```r
autoplot(decompose(white))
```




```r
# There seems to be a seasonality present
# The series is not covariance stationary
# Consider taking the first order difference to remove trend component
```

```r
adf.test(black)
```

Augmented Dickey-Fuller Test (Unit Root Test)

```
##
##  Augmented Dickey-Fuller Test
##
## data:  black
## Dickey-Fuller = -1.6434, Lag order = 6, p-value = 0.7262
## alternative hypothesis: stationary
```

```r
adf.test(white)
```

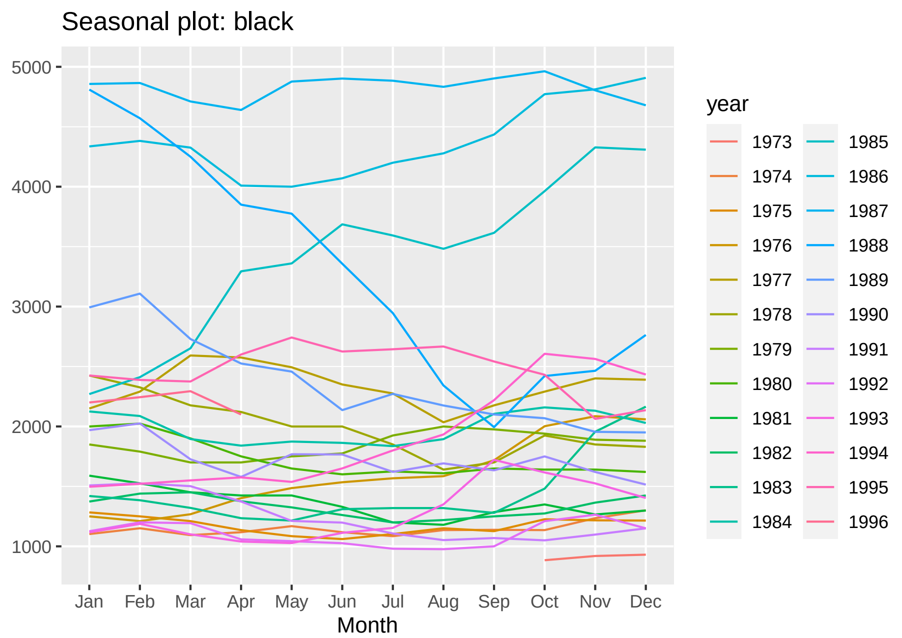
```
##
##  Augmented Dickey-Fuller Test
##
## data:  white
## Dickey-Fuller = -1.6001, Lag order = 6, p-value = 0.7444
## alternative hypothesis: stationary
```

Upon looking at the decompositon plots for both the Black and White variables, we can initially see that stationarity does not look likely because the time series plot seems to illustrate a prolonged upward trend over time, with prices peaking in the mid 1980's before declining. One can see that there is not a constant unconditional mean due to the upward trend for most of the time periods that ends in a steep decline. The ACF gradually declines as the lags increase. After running the augmented Dickey-Fuller (ADF) test on both variables (and failing to reject the null hypothesis for both), we can definitively conclude that there is not enough evidence to suggest stationarity for either variable.

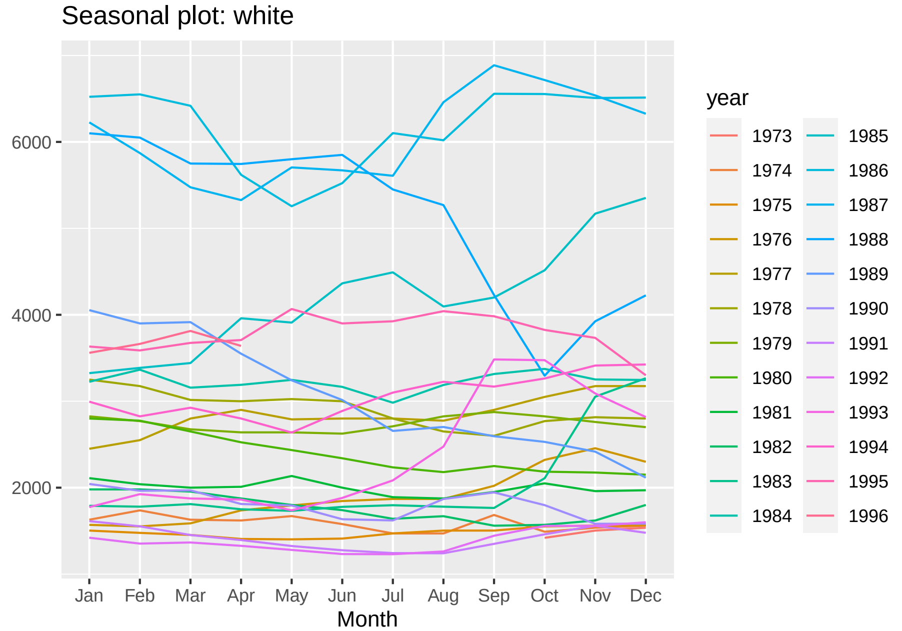
After running the ADF test on both black and white pepper prices, our assumptions that both exhibit non-stationary were confirmed by the resulting p-values, which were 0.7262 and 0.7444 for black and white pepper, respectively, and support the null hypothesis of non-stationarity.

```r


ggseasonplot(black)
```

Inspecting Seasonality




```
ggseasonplot(white)
```




```
# There is no regular seasonal patterns
```

As it was hard to visually identify the seasonal component in the original plots, we now can explicitly see that seasonality for the both plots are not regular and changes over time.

(b) Differencing the Time Series

```r
# Determine the level of differencing to make out series stationary
ndiffs(black)
```

Determining the level of differencing

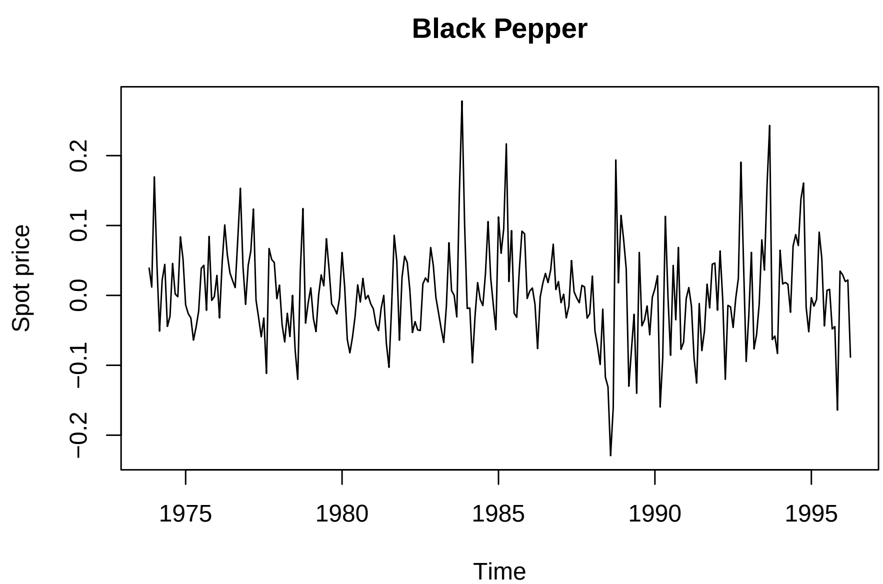
```
## [1] 1
```

```r
ndiffs(white)
```

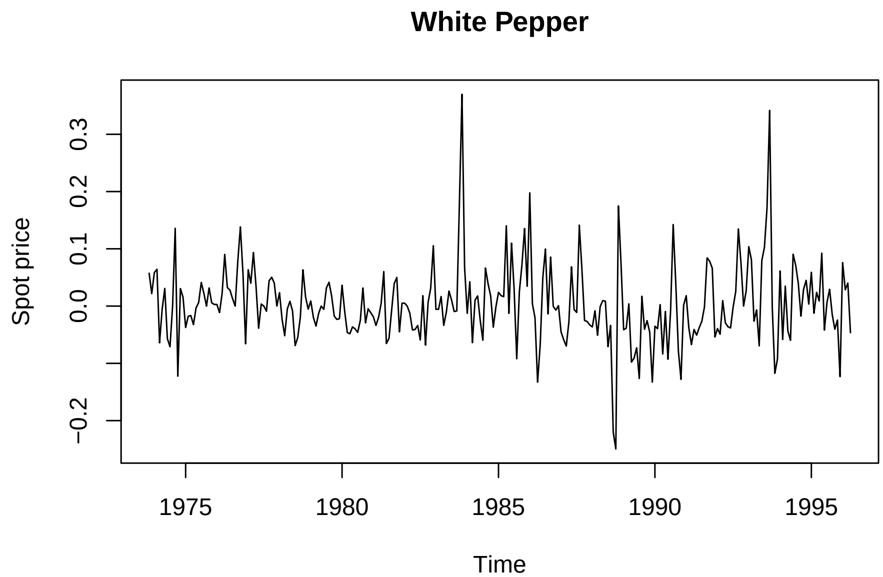
```
## [1] 1
```

By using the `ndiffs` function, we identify that first order difference is needed for the original series to be stationary.

```r
# Logging the time series to stabilize the volatility
# Taking First order difference
plot(diff(log(black)), main = "Black Pepper", ylab = "Spot price")
```

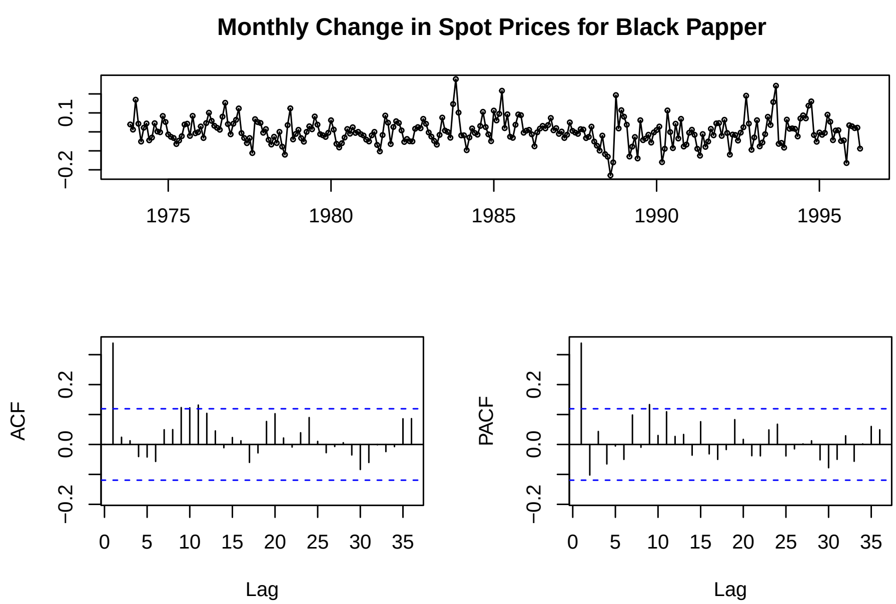



```r
plot(diff(log(white)), main = "White Pepper", ylab = "Spot price")
```

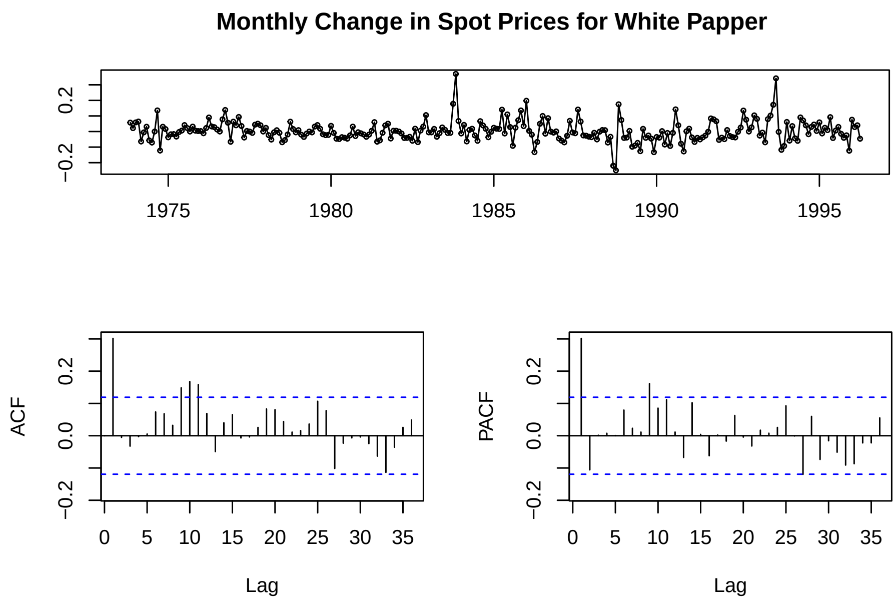


To stabilize the variances over time, the logging series is needed and by taking the first order difference, covariance stationarity is now obtained.

```r
# Redoing Part (a)
tsdisplay(diff(log(black)), main = "Monthly Change in Spot Prices for Black Papper ")
```




```r
tsdisplay(diff(log(white)), main = "Monthly Change in Spot Prices for White Papper ")
```




```r
adf.test(diff(log(black)))
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff(log(black))
## Dickey-Fuller = -5.6611, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
```

```r
adf.test(diff(log(white)))
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff(log(white))
## Dickey-Fuller = -5.336, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
```

```r
# Fail to reject the null hypothesis; now both of the series are stationary!
```

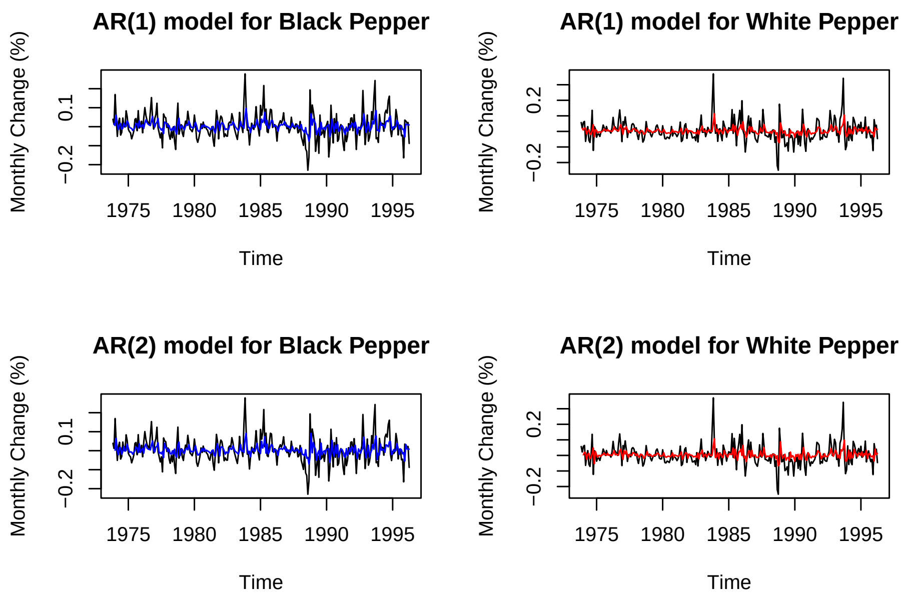
To confirm stationary behavior, an ADF test was performed on both black and white paper prices and both tests exhibited a p-value of 0.01, rejecting the null and supporting the alternative hypothesis of stationarity.

It is noteworthy that just by examining the ACF and PACF, it is not easy to tell which process or order to use to model the series. Based on visual inspection, it may be more appropriate to use ARMA model or VAR model to best estimate the fit, but for the scope of this project, we first run AR(p) process and evaluate the performance, using AIC and BIC in part 3.

3. Feature Generation, Model Testing and Forecasting.

(a) Fit an AR(p) model to the data (using part 2(a), AIC or some built in R function)

```
# Fitting an AR(1) model
stationary_black <- diff(log(black))
stationary_white <- diff(log(white))

library(dynlm)

ar1_black <- dynlm(stationary_black ~ L(stationary_black, 1))
ar1_white <- dynlm(stationary_white ~ L(stationary_white, 1))

par(mfrow = c(2,2))
# Fitting an AR(1) model to the time series for black pepper
plot(stationary_black, main = "AR(1) model for Black Pepper", ylab = "Monthly Change (%)")
lines(ar1_black$fitted.values, col = "blue")

# Fitting an AR(1) model to the time series for white pepper
plot(stationary_white, main = "AR(1) model for White Pepper", ylab = "Monthly Change (%)")
lines(ar1_white$fitted.values, col = "red")



ar2_black <- dynlm(stationary_black ~ L(stationary_black, 1) + L(stationary_black, 2))
ar2_white <- dynlm(stationary_white ~ L(stationary_white, 1) + L(stationary_white, 2))

# Fitting an AR(2) model to the time series for black pepper
plot(stationary_black, main = "AR(2) model for Black Pepper", ylab = "Monthly Change (%)")
lines(ar2_black$fitted.values, col = "blue")

# Fitting an AR(2) model to the time series for white pepper
plot(stationary_white, main = "AR(2) model for White Pepper", ylab = "Monthly Change (%)")
lines(ar2_white$fitted.values, col = "red")
```




```r
aic_bic <- data.frame(AIC = c(AIC(ar1_black), AIC(ar2_black),
                              AIC(ar1_white), AIC(ar2_white)),
                      BIC = c(BIC(ar1_black), BIC(ar2_black),
                              BIC(ar1_white), BIC(ar2_white)))

rownames(aic_bic) <- c("AR(1), Black", "AR(2), Black", "AR(1), White", "AR(2), White")
aic_bic
```

```
##                    AIC       BIC
## AR(1), Black -709.3660 -698.5819
## AR(2), Black -706.6429 -692.2790
## AR(1), White -706.3176 -695.5335
## AR(2), White -703.7495 -689.3856
```

Based on the visual inspection of the model fits and AIC/BIC, we can conclude that AR(1) model is better performing on the series, compared to AR(2) model.

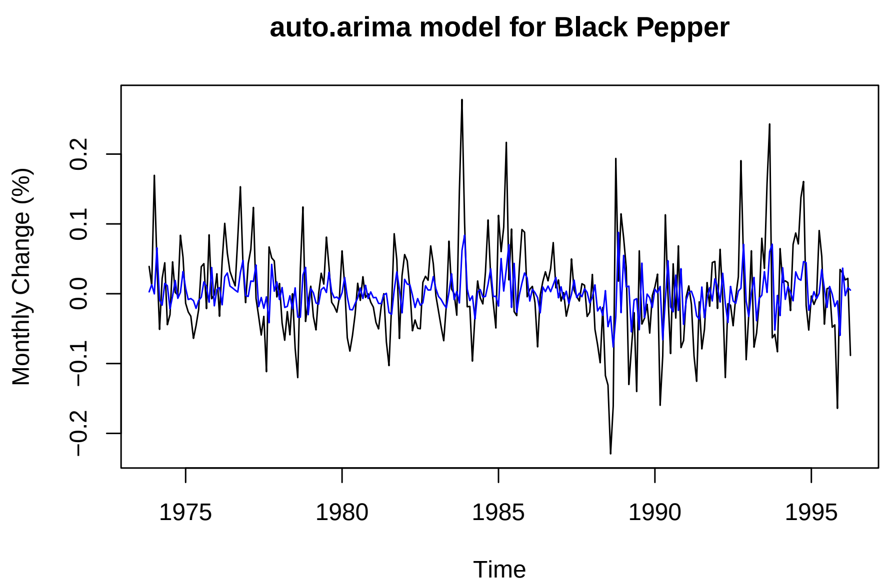
```r
auto.arima(stationary_black)
```

```
## Series: stationary_black
## ARIMA(0,0,1) with zero mean
##
## Coefficients:
##          ma1
##       0.3879
## s.e.  0.0585
##
## sigma^2 = 0.004046:  log likelihood = 361.16
## AIC=-718.33   AICc=-718.28   BIC=-711.13
```

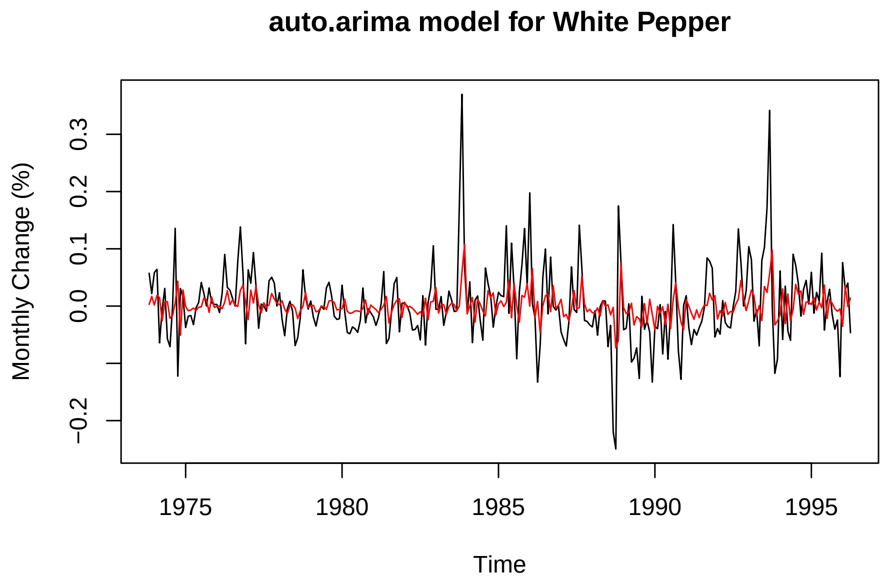
```r
auto.arima(stationary_white)
```

```
## Series: stationary_white
## ARIMA(0,0,1)(1,0,0)[12] with zero mean
##
## Coefficients:
##          ma1    sar1
##       0.3278  0.0601
## s.e.  0.0558  0.0629
##
## sigma^2 = 0.004121:  log likelihood = 359.2
## AIC=-712.39   AICc=-712.3   BIC=-701.6
```

```r
# Using auto.arima for black pepper
plot(stationary_black, main = "auto.arima model for Black Pepper",
     ylab = "Monthly Change (%)")
lines(auto.arima(stationary_black)$fitted, col = "blue")
```




```r
# Using auto.arima for white pepper
plot(stationary_white, main = "auto.arima model for White Pepper",
     ylab = "Monthly Change (%)")
lines(auto.arima(stationary_white)$fitted, col = "red")
```

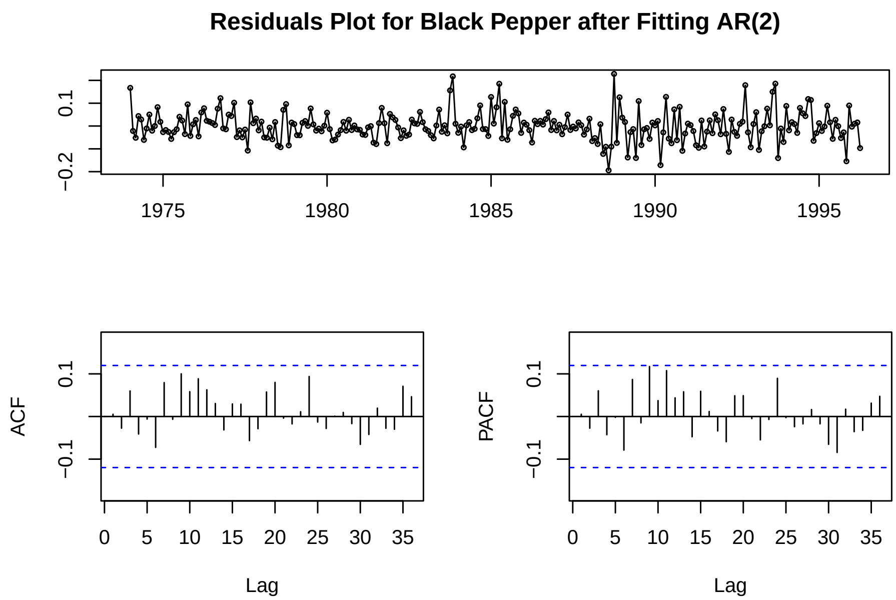


The plots shown above show the model fit create by the function 'auto.arima()', which automatically generate the "optimal" model fit.

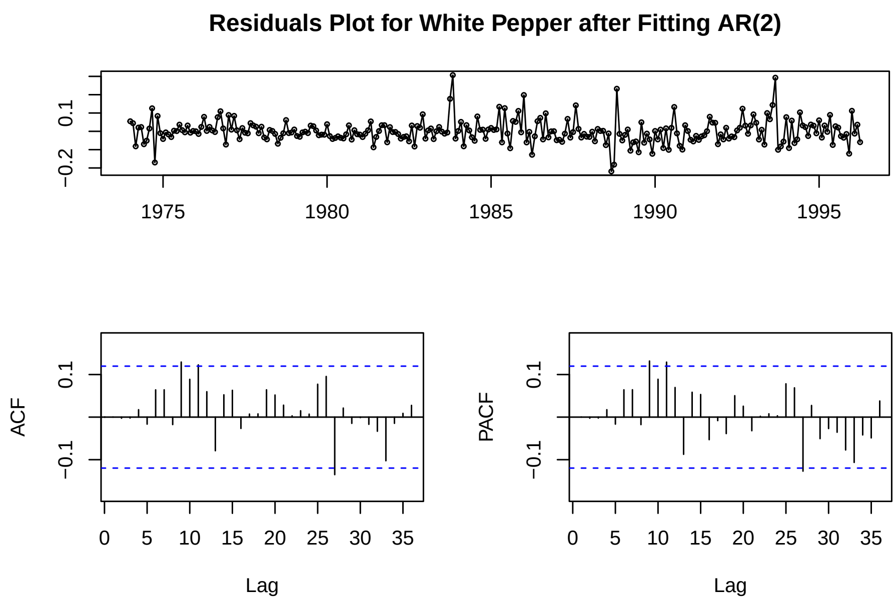
As briefly mentioned in the previous section, `auto.arima` function is suggesting MA(1) process as the ideal model fit. Although MA process will not be used to model the data in this project to focus on the scope of the project, such limitations will be further discussed in section 5.

(b) Plot and comment on the ACF of the residuals of the model chosen in 3(a). If the model is properly fit, then we should see no autocorrelations in the residuals. Carry out a formal test for autocorrelation and comment on the results.

```
# tsdisplay AR(2) model
tsdisplay(ar2_black$residuals, main =
            "Residuals Plot for Black Pepper after Fitting AR(2)")
```




```r
tsdisplay(ar2_white$residuals, main =
            "Residuals Plot for White Pepper after Fitting AR(2)")
```




```r
Box.test(ar1_black$residuals)
```

```
##
##  Box-Pierce test
##
## data:  ar1_black$residuals
## X-squared = 0.33927, df = 1, p-value = 0.5602
```

```r
Box.test(ar1_white$residuals)
```

```
##
##  Box-Pierce test
##
## data:  ar1_white$residuals
## X-squared = 0.27818, df = 1, p-value = 0.5979
```

```r
#Ljung-Box test; H0: No autocorrelation
# We fail to reject the null --> the resisuals are just like white noise!
```

After plotting the ACF of the residuals for black pepper prices, none of the lags exceed the significance bounds, which is an indication of a well-specified model. The ACF plot for the residuals of white pepper prices has one spike that marginally exceeds the significance bounds, while the remaining lags do not. This could either be an anomaly or a sign of minor autocorrelation at that specific lag, but given the overall pattern, our model appears to be well-specified and does not exhibit autocorrelation. Additionally, the PACF correlogram for both black and white pepper prices does not show significant spikes, reinforcing the absence of autocorrelation and indicating that the model has captured the essential autoregressive properties of the time series data. Finally, the Box-Pierce test confirms our observation by providing a p-value of 0.5602 for the residual of black pepper prices and 0.5979 for the residual of white pepper prices. Given that both p-values are high, we can conclude that the model could potentially be improved by adapting other methods, but there are no dynamics left that we can capture.

(c) Using the appropriate predictors, fit an ARDL(p,q) model to the data and repeat step (b) in part 3.

```r
# Based on AIC & BIC, ARDL(1,0) is the best model for black

aic_bic <- data.frame(AIC = c(AIC(ar1_black),
                   AIC(dynlm(stationary_black ~ L(stationary_black, 1)
                             + stationary_white,)),
                   AIC(dynlm(stationary_black ~ L(stationary_black, 1)
                             + stationary_white + L(stationary_white, 1))),
                   AIC(dynlm(stationary_black ~ L(stationary_black, 1)
                             + stationary_white + L(stationary_white, 1)
                             + L(stationary_white, 2)))),
           BIC = c(BIC(ar1_black),
                   BIC(dynlm(stationary_black ~ L(stationary_black, 1)
                             + stationary_white)),
                   BIC(dynlm(stationary_black ~ L(stationary_black, 1)
                             + stationary_white + L(stationary_white, 1))),
                   BIC(dynlm(stationary_black ~ L(stationary_black, 1)
                             + stationary_white + L(stationary_white, 1)
                             + L(stationary_white, 2)))))

rownames(aic_bic) <- c("AR(1), Black", "ARDL(1,0), Black",
                       "ARDL(1,1), Black", "ARDL(1,2), Black")
aic_bic
```

```
##                          AIC       BIC
```

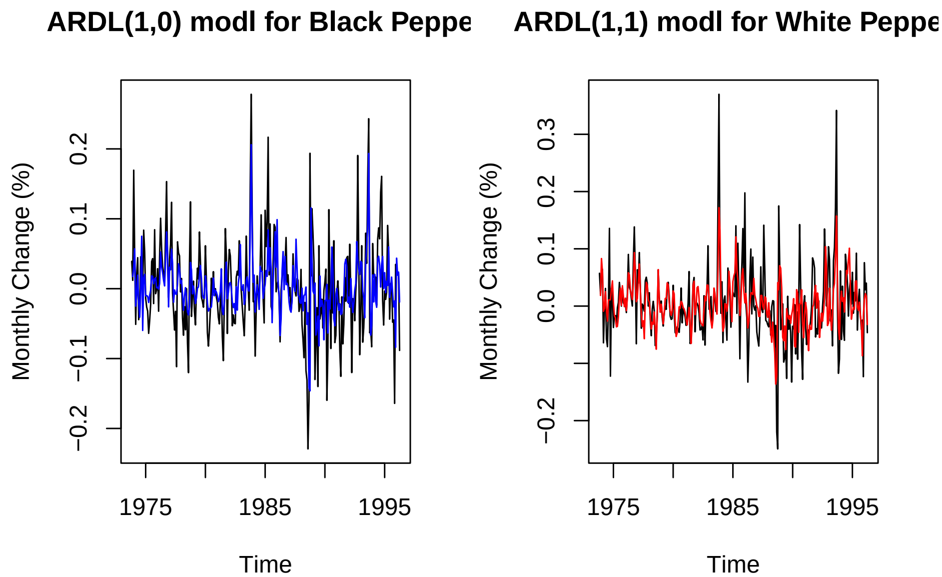
```
## AR(1), Black     -709.3660 -698.5819
## ARDL(1,0), Black -778.3891 -764.0102
## ARDL(1,1), Black -776.6069 -758.6333
## ARDL(1,2), Black -770.6981 -749.1522
```

```
aic_bic <- data.frame(AIC = c(AIC(ar1_white),
                   AIC(dynlm(stationary_white ~ L(stationary_white, 1)
                             + stationary_black,)),
                   AIC(dynlm(stationary_white ~ L(stationary_white, 1)
                             + stationary_black + L(stationary_black, 1))),
                   AIC(dynlm(stationary_white ~ L(stationary_white, 1)
                             + stationary_black + L(stationary_black, 1)
                             + L(stationary_black, 2)))),
          BIC = c(BIC(ar1_white),
                   BIC(dynlm(stationary_white ~ L(stationary_white, 1)
                             + stationary_black,)),
                   BIC(dynlm(stationary_white ~ L(stationary_white, 1)
                             + stationary_black + L(stationary_black, 1))),
                   BIC(dynlm(stationary_white ~ L(stationary_white, 1)
                             + stationary_black + L(stationary_black, 1)
                             + L(stationary_black, 2)))))

rownames(aic_bic) <- c("AR(1), White", "ARDL(1,0), White",
                       "ARDL(1,1), White", "ARDL(1,2), White")
aic_bic
```

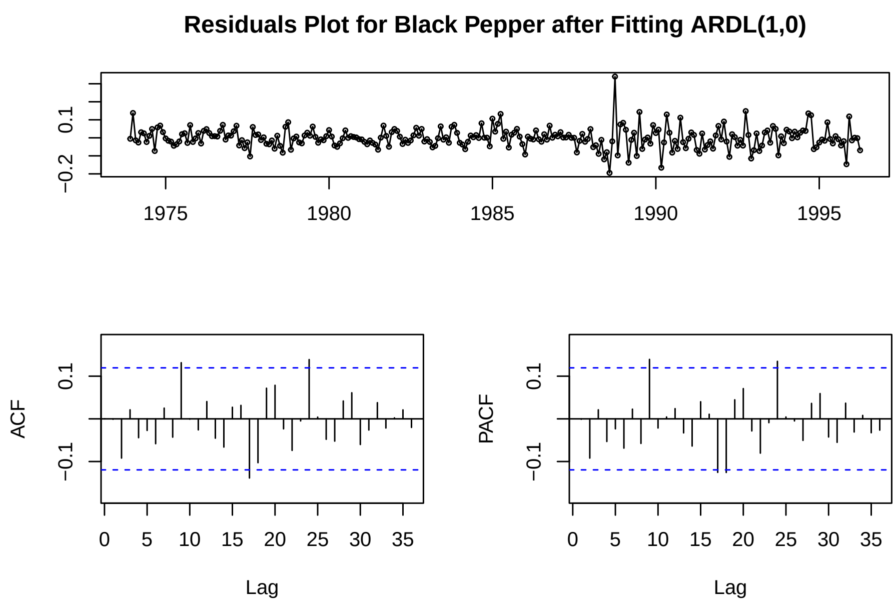
```
##                        AIC       BIC
## AR(1), White     -706.3176 -695.5335
## ARDL(1,0), White -787.6135 -773.2347
## ARDL(1,1), White -796.4300 -778.4565
## ARDL(1,2), White -790.9877 -769.4418
```

```
# Based on AIC $ BIC, ARDL(1,1) is the best model for white
```

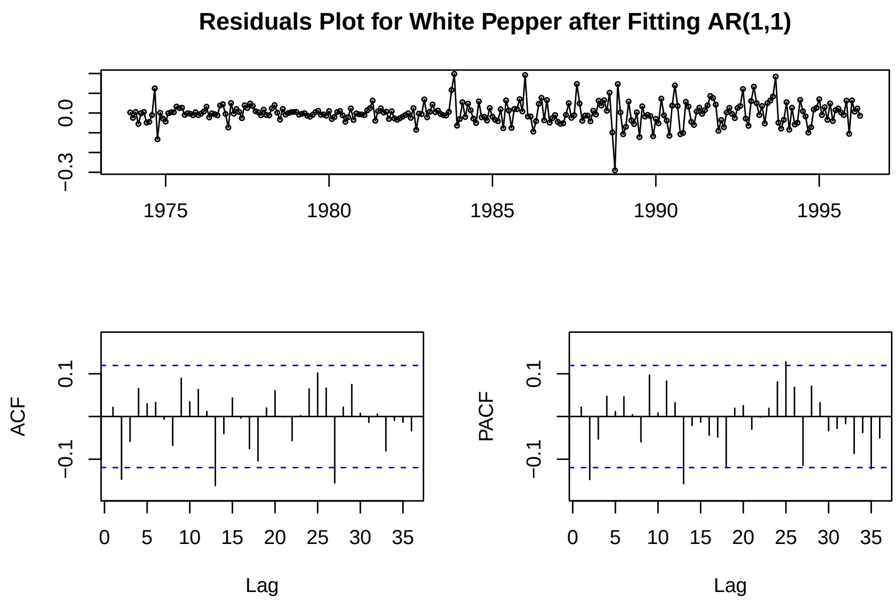
```
# Redoing step (b) in part 3
ardl10_black <- dynlm(stationary_black ~ L(stationary_black, 1) + stationary_white)

par(mfrow = c(1,2))
# Plotting ARDL(1,0) model for Black Pepper
plot(stationary_black, main = "ARDL(1,0) modl for Black Pepper",
     ylab = "Monthly Change (%)")
lines(ardl10_black$fitted, col = "blue")



ardl11_white <- dynlm(stationary_white ~ L(stationary_white, 1) + stationary_black
                      + L(stationary_black, 1))
# Plotting ARDL(1,1) model for White Pepper
plot(stationary_white, main = "ARDL(1,1) modl for White Pepper",
     ylab = "Monthly Change (%)")
lines(ardl11_white$fitted, col = "red")
```




```r
# tsdisplaying after ARDL model
tsdisplay(ardl10_black$residuals, main =
            "Residuals Plot for Black Pepper after Fitting ARDL(1,0)")
```




```r
tsdisplay(ardl11_white$residuals, main =
            "Residuals Plot for White Pepper after Fitting AR(1,1)")
```




```r
# Ljung-Box Test
Box.test(ardl10_black$residuals)
```

```
##
##  Box-Pierce test
##
## data:  ardl10_black$residuals
## X-squared = 0.00025064, df = 1, p-value = 0.9874
```

```r
Box.test(ardl11_white$residuals)
```

```
##
##  Box-Pierce test
##
## data:  ardl11_white$residuals
## X-squared = 0.12711, df = 1, p-value = 0.7215
```

```r
#Ljung-Box test; H0: No autocorrelation
# We fail to reject the null --> the residuals are just like white noise!
```

Based on each model fit, it is clear that the the models performs better than the previous AR(1) model. ARDL processes are capturing sudden spikes that were not captured in the AR models, as well as some cyclical components.

To find the best fitting model, AIC and BIC were conducted again. The best fitting model for black pepper prices is ARDL(1,0) since this model renders the lowest values. For white pepper prices, the AIC and BIC indicate that the best fitting model would be ARDL(1,1). The ACF of the residuals for both black and white pepper prices show spikes that are largely within the significance bounds. This implies that the models are properly fitted. The PACF plots for both black and white pepper prices share a similar story, with the partial autocorrelations being within the significance bounds, exhibiting white noise behavior. To confirm, we ran the Box-Pierce test on the residuals of both the black and white pepper prices. The associated p-values were 0.9874 and 0.7215, respectively, confirming that the data does not provide strong evidence for the presence of autocorrelation.

Causal relationship between two variables The best performing model fit intuitively makes sense when we look at the causal relationship between two series. Below is the VAR model and Granger Causality test:

```
library(vars)
vmod <- data.frame(stationary_black, stationary_white)
VARselect(vmod, lag.max = 10)
```

```
## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      3      1      1      3
##
## $criteria
##                    1             2             3             4             5
## AIC(n) -1.127460e+01 -1.127772e+01 -1.128383e+01 -1.125666e+01 -1.123705e+01
## HQ(n)  -1.124157e+01 -1.122266e+01 -1.120676e+01 -1.115756e+01 -1.111593e+01
## SC(n)  -1.119243e+01 -1.114077e+01 -1.109211e+01 -1.101015e+01 -1.093576e+01
## FPE(n)  1.269120e-05  1.265182e-05  1.257489e-05  1.292171e-05  1.317819e-05
##                    6             7             8             9            10
## AIC(n) -1.122131e+01 -1.120087e+01 -1.117950e+01 -1.119104e+01 -1.119628e+01
## HQ(n)  -1.107817e+01 -1.103570e+01 -1.099231e+01 -1.098182e+01 -1.096505e+01
## SC(n)  -1.086524e+01 -1.079002e+01 -1.071387e+01 -1.067063e+01 -1.062110e+01
## FPE(n)  1.338808e-05  1.366588e-05  1.396263e-05  1.380454e-05  1.373478e-05
```

```
summary(VAR(vmod, p = 3))
```

```
##
## VAR Estimation Results:
## =========================
## Endogenous variables: stationary_black, stationary_white
## Deterministic variables: const
## Sample size: 267
## Log Likelihood: 767.255
## Roots of the characteristic polynomial:
## 0.549 0.549 0.5107 0.5107 0.4293 0.4293
## Call:
## VAR(y = vmod, p = 3)
##
##
## Estimation results for equation stationary_black:
## ===================================================
## stationary_black = stationary_black.l1 + stationary_white.l1 + stationary_black.l2 + stationary_white
##
##                      Estimate Std. Error t value Pr(>|t|)
## stationary_black.l1  0.331371   0.069131   4.793 2.76e-06 ***
## stationary_white.l1  0.115217   0.071707   1.607   0.1093
## stationary_black.l2 -0.138732   0.073001  -1.900   0.0585 .
## stationary_white.l2 -0.019424   0.071942  -0.270   0.7874
## stationary_black.l3  0.120113   0.072002   1.668   0.0965 .
## stationary_white.l3 -0.127630   0.069129  -1.846   0.0660 .
## const                0.001459   0.003869   0.377   0.7063
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
##
## Residual standard error: 0.06302 on 260 degrees of freedom
## Multiple R-Squared: 0.1525,  Adjusted R-squared: 0.133
## F-statistic: 7.799 on 6 and 260 DF,  p-value: 9.927e-08
##
##
## Estimation results for equation stationary_white:
## =================================================
## stationary_white = stationary_black.l1 + stationary_white.l1 + stationary_black.l2 + stationary_whit
##
##                      Estimate Std. Error t value Pr(>|t|)
## stationary_black.l1  0.377283   0.066734   5.654 4.13e-08 ***
## stationary_white.l1  0.154806   0.069220   2.236  0.02617 *
## stationary_black.l2 -0.070607   0.070470  -1.002  0.31730
## stationary_white.l2 -0.163485   0.069447  -2.354  0.01931 *
## stationary_black.l3  0.197366   0.069505   2.840  0.00487 **
## stationary_white.l3 -0.089122   0.066732  -1.336  0.18288
## const                0.001623   0.003735   0.434  0.66431
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.06084 on 260 degrees of freedom
## Multiple R-Squared: 0.2136,  Adjusted R-squared: 0.1955
## F-statistic: 11.77 on 6 and 260 DF,  p-value: 1.131e-11
##
##
##
## Covariance matrix of residuals:
##                  stationary_black stationary_white
## stationary_black         0.003972         0.001778
## stationary_white         0.001778         0.003701
##
## Correlation matrix of residuals:
##                  stationary_black stationary_white
## stationary_black           1.0000           0.4638
## stationary_white           0.4638           1.0000
```

Based on the results above, we can see that the lagged values of black pepper's spot price have statistically significant impact on the spot prices of white pepper. In other word, the past values of black pepper's spot prices help predict the future values of white pepper's spot price, and this will be a strong tool for forecasting the future price.

```r
# Examining Granger Causalty
grangertest(stationary_black ~ stationary_white, order = 2)
```

```
## Granger causality test
##
## Model 1: stationary_black ~ Lags(stationary_black, 1:2) + Lags(stationary_white, 1:2)
## Model 2: stationary_black ~ Lags(stationary_black, 1:2)
##   Res.Df Df      F Pr(>F)
## 1    263
## 2    265 -2 1.5981 0.2042
```

```r
grangertest(stationary_white ~ stationary_black, order = 2)
```

```
## Granger causality test
##
## Model 1: stationary_white ~ Lags(stationary_white, 1:2) + Lags(stationary_black, 1:2)
## Model 2: stationary_white ~ Lags(stationary_white, 1:2)
##   Res.Df Df      F    Pr(>F)
## 1    263
## 2    265 -2 14.108 1.513e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As shown above, the spot price of black pepper "Granger causes" the spot price of white pepper. This means that changes or fluctuations in the black pepper series are statistically significant in terms of explaining or predicting the spot price of white pepper.

In short, based on the VAR model and the Granger Causualty test, the causal relationship between two variable is unidirectional, specifically from black pepper to white pepper.

This also explains why the order of ARDL model for black pepper is ARDL(1,0) because the past values of white pepper does not "Granger cause" black pepper's price. In contrast, the lagged values of black pepper do have a statistically significant meaning on white pepper's price (specifically, lag 1 of black pepper is significant at a 0.01 significance level), which explains the order of ARDL(1,1) for white pepper.

4. Summary and Findings

As discussed in 3b and 3c, both AR and ARDL process perform considerably well. However, after comparing models based on AIC and BIC results, it was observed that ARDL models generally perform better than pure AR models. A simple metric to visualize this is the lower values of AIC and BIC for the ARDL models when compared to the AR models. We also found out that there is a unidirectional causal relationship between two variables; the spot prices for black pepper "Granger cause" the spot prices for white pepper. This also explains the improved performance of ARDL models in comparison to AR models. In conclusion, among the models that were tested in this project, ARDL(1,0) and ARDL(1,1) are the best performing model for `Black` and `White` Pepper respectively.

5. Limitations and Future work

The scope of this project focuses on modeling time series with AR and ARDL process. Thus, some subtle dynamics that could be captured by other methods, such as the VAR model, ARIMA, or ETS, may not have been taken into account. Additionally, when auto.arima was used in section 3, the MA process may have been a better fit for the series, considering the quick mean reverting behavior of the differenced series. For future reference, finding a way to model the series with a lower order model or setting different weight on each parameter using ETS, could possibly not only improve the model, but also better forecast the future values while not overfitting the model.