BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC KINH TẾ QUỐC

---------***--------

FORECASTING TIME SERIES DATA USING ARIMA MODEL

Group 10

No1	ID	Full name	Contribution for common section
1	11208006	Đỗ Thùy Trang	
2	11202127	Hoàng Diệu Linh	
3	11201426	Nguyễn Thị Thanh Hiền	
Sum			100%

Hà Nội, 2023

INTRODUCTION

To predict an outcome based on time series data, we can use a time series model which is called Auto Regressive Integrated Moving Average (ARIMA). It is used as the machine learning technique to analyze and predict future stock prices based on historical prices.

ARIMA (autoregressive integrated moving average) is a commonly used technique utilized to fit time series data and forecasting. It is a generalized version of ARMA (autoregressive moving average) process, where the ARMA process is applied for a different version of the data rather than the original. Three numbers p, d and q specify ARIMA model and the ARIMA model is said to be of order (p, d, q). Here p, d and q are the orders of AR part, Difference, and the MA part respectively. AR and MA- both are different techniques for stationary time series data. ARMA (and ARIMA) is a combination of these two methods for better fitting the model. In this write-up an overview of the AR and MA process will be given. The steps of building an ARIMA model will be explained. Finally, a demonstration using R will be presented.

A. Individual

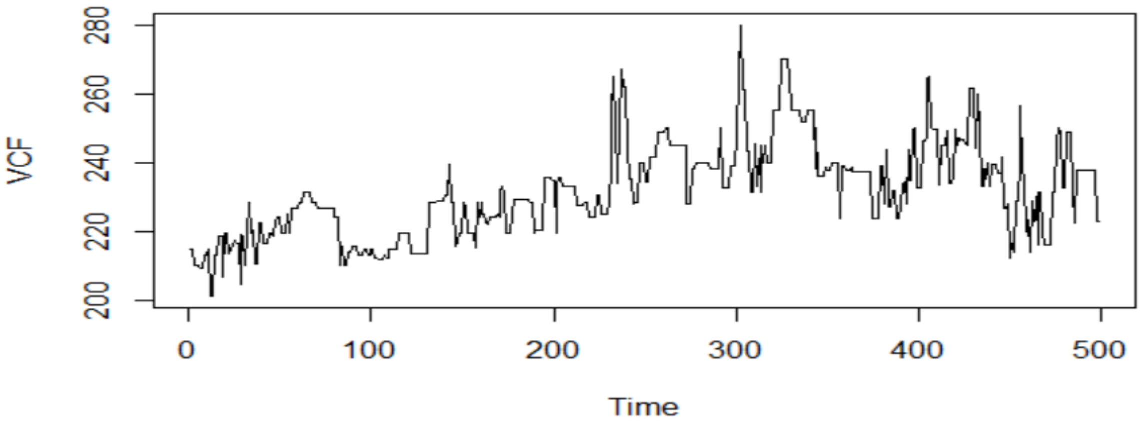
I. Đỗ Thùy Trang

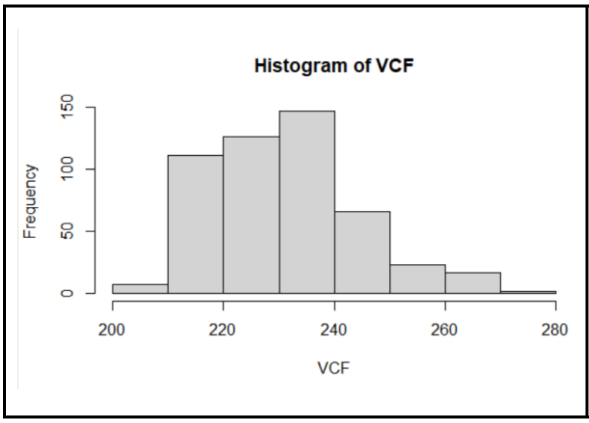
1. CTCP Vinacafé Biên Hòa (VCF: HOSE)

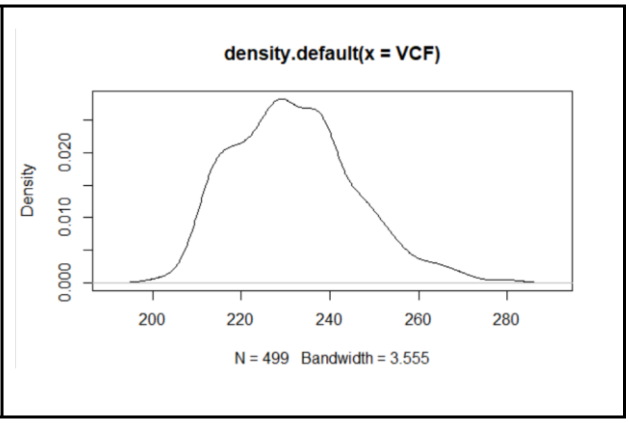
Vinacafé BH inherited the biggest achievement of Bien Hoa Coffee Factory, formerly the Vinacafé brand. Established in the 1980s and officially recognized as an intellectual property in 1993, Vinacafé brand today has become a major brand of Vietnam, selected into the National Brand Program since 2008. Vinacafé painstakingly built from a solid foundation: product quality and commitment "Taste of nature."

2. Finance series

Daily_data : <close_price> of VCF in 2021-2022

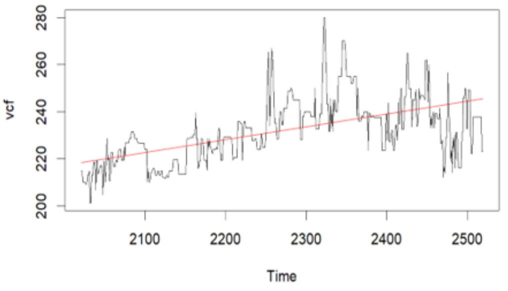
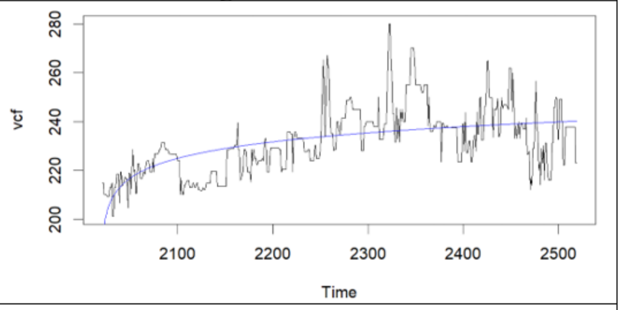
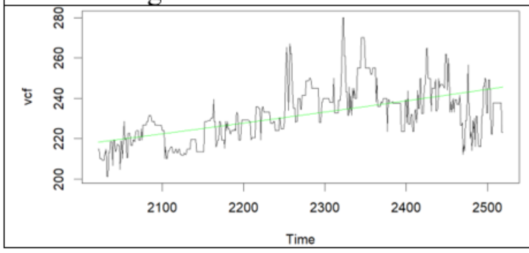
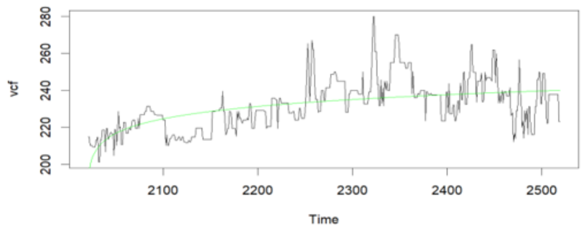






The distribution of price for VCF is right-skewed because it’s longer on the right side of its peak. This means the outliers of the distribution curve are further out towards the right and closer to the mean on the left

Compare among models:

1. Linear-linear	2. Linear- log
	
3. Log - linear	4. Log - log
	

We can see that model linear-log and log-log are fittest.

Forecast for the first next value:

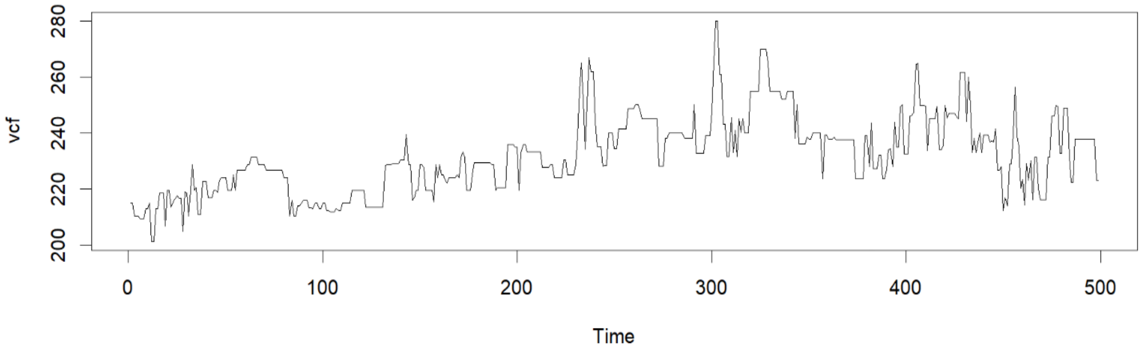
t	Linear - log	Log-log
	$vcf_t = 188.385 + 8.339.\ln(t)$	$\ln\,(vcf_{t)} = 5.255 + 0.036.\ln(t)$
500	240.209	239.541
501	240.225	239.559
502	240.242	239.576
503	240.259	239.593
504	240.275	239.610
505	240.292	239.627
506	240.308	239.644
507	240.325	239.661

Compare	Range value	Linear - log	Log-log
RMSE	Whole data	10.989	10.946
	4 last objects	11.157	10.991
MAPE	Whole data	0.0354	0.0351
	4 last objects	0.0383	0..0379

3. Stock price

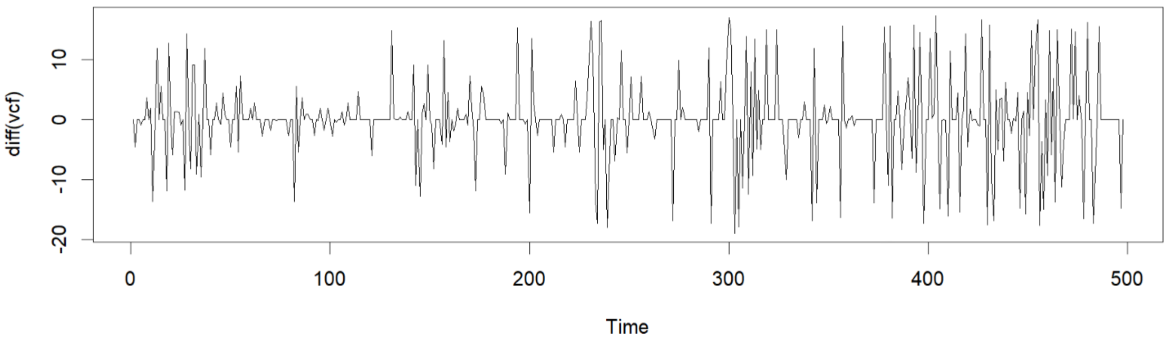
a. Plot

The following time series data that represents the stock price for VCF during 2 years' periods



Clearly the prices are trending upwards and Nonstationary over time, but there also appears to be a cyclical or seasonal trend in the data, which can be seen by the tiny "hills" that occur over time.

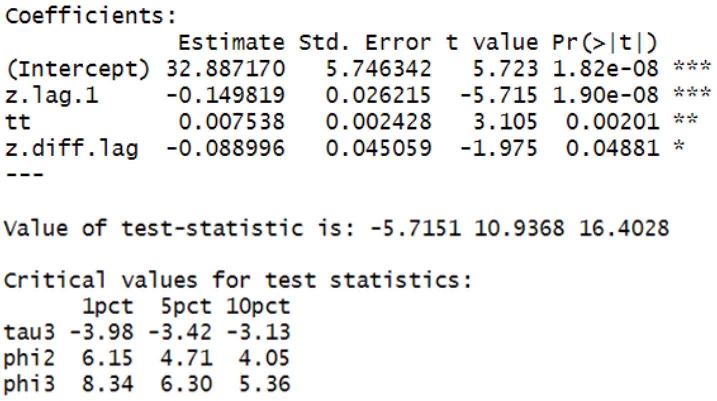
To gain a better view of this cyclical trend, we can detrend the data. In this case, this would involve removing the overall upward trend over time so that the resulting data represents just the cyclical trend. One way to detrend time series data is to simply create a new dataset where each observation is the difference between itself and the previous observation.



VCF is "different stationary": I (1). It's much easier to see the seasonal trend in the time series data in this plot because the overall upward trend has been removed.

b. Unit root test for price of VCF

Test with trend and constant

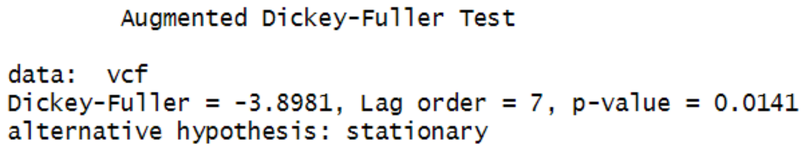

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 32.887170   5.746342   5.723 1.82e-08 ***
z.lag.1     -0.149819   0.026215  -5.715 1.90e-08 ***
tt           0.007538   0.002428   3.105  0.00201 ** 
z.diff.lag  -0.088996   0.045059  -1.975  0.04881 *  
---

Value of test-statistic is: -5.7151 10.9368 16.4028 

Critical values for test statistics: 
      1pct  5pct 10pct
tau3 -3.98 -3.42 -3.13
phi2  6.15  4.71  4.05
phi3  8.34  6.30  5.36
```

Test for significant of trend:

$$\{H_0: Trend\ is\ insignificant\ H_1: Trend\ is\ significant$$

$|\tau_{trend}| = 16.4 > |\tau_{0.05}| = 6.3$ ▯ Reject Ho ▯ $Trend\ is\ significant$

Test for unit root

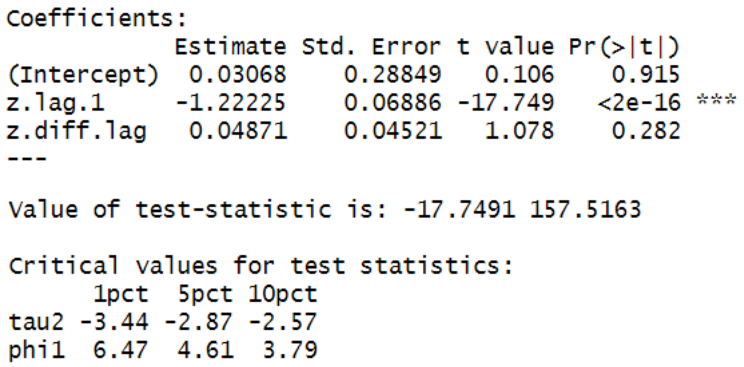


```
        Augmented Dickey-Fuller Test

data:  vcf
Dickey-Fuller = -3.8981, Lag order = 7, p-value = 0.0141
alternative hypothesis: stationary
```

$$\{H_0: \delta = 0: \ vcf\ has\ unit\ root \quad H_1: \delta < 0: \ vcf\ has\ not\ unit\ root$$

P_val = 0.014 < 0.05 but > 0.01 Not Reject Ho $VCF\ has\ \ unit\ root$

c. Unit root test for difference series



```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.03068    0.28849   0.106    0.915    
z.lag.1     -1.22225    0.06886 -17.749   <2e-16 ***
z.diff.lag   0.04871    0.04521   1.078    0.282    
---

Value of test-statistic is: -17.7491 157.5163 

Critical values for test statistics: 
      1pct  5pct 10pct
tau2 -3.44 -2.87 -2.57
phi1  6.47  4.61  3.79
```

$$\{H_0: \delta = 0: \ vcf\ has\ unit\ root \quad H_1: \delta < 0: \ vcf\ has\ not\ unit\ root$$

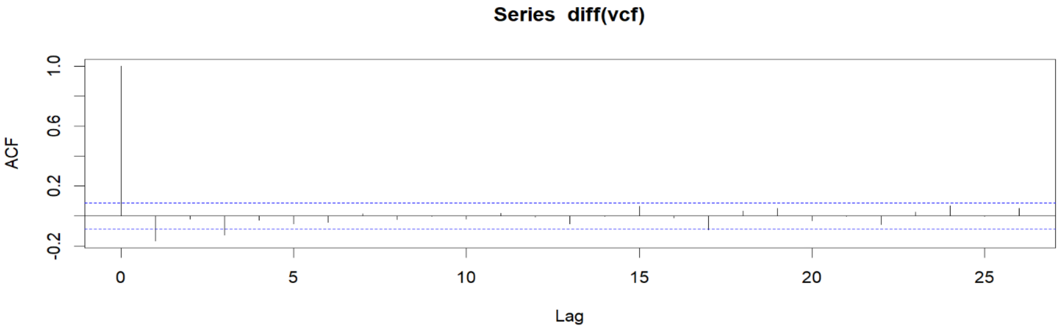
$|\tau_{ur}| = 17.7491 > \ |\tau_{0.05}| = 2.87$ Reject Ho.

Mean that time series does not have a unit root, meaning it is trend stationary.

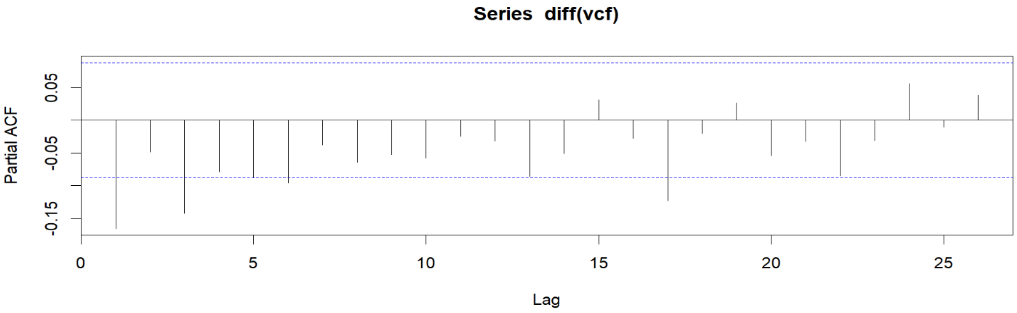
We can conclude that : $\Delta VCF_t\ has\ no\ unit\ root\ and\ stationary\ around\ a\ constant$

4. ACF & PACF to determine for ARMA of different series

Autocorrelation function (ACF). At lag *k*, this is the correlation between series values that are *k* intervals apart.



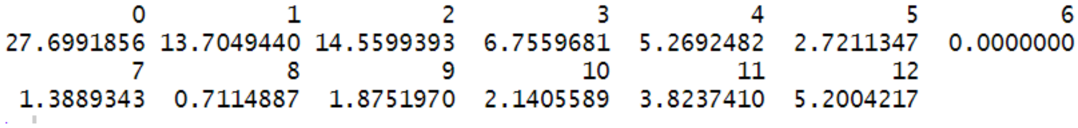
Partial autocorrelation function (PACF). At lag *k*, this is the correlation between series values that are *k* intervals apart, accounting for the values of the intervals between.



The *x* axis of the ACF plot indicates the lag at which the autocorrelation is computed; the *y* axis indicates the value of the correlation (between −1 and 1). For example, a spike at lag 1 in an ACF plot indicates a strong correlation between each series value and the preceding value, a spike at lag 2 indicates a strong correlation between each value and the value occurring two points previously, and so on.

We can observe that with ACF, the histogram has 6nd order lag and 3nd order lag PACF. The ADF test shows that the series of first difference logarithms of stock price is stationary, so we define the model ARIMA (p, d, q) suitable for prediction as ARIMA (6,1,3).

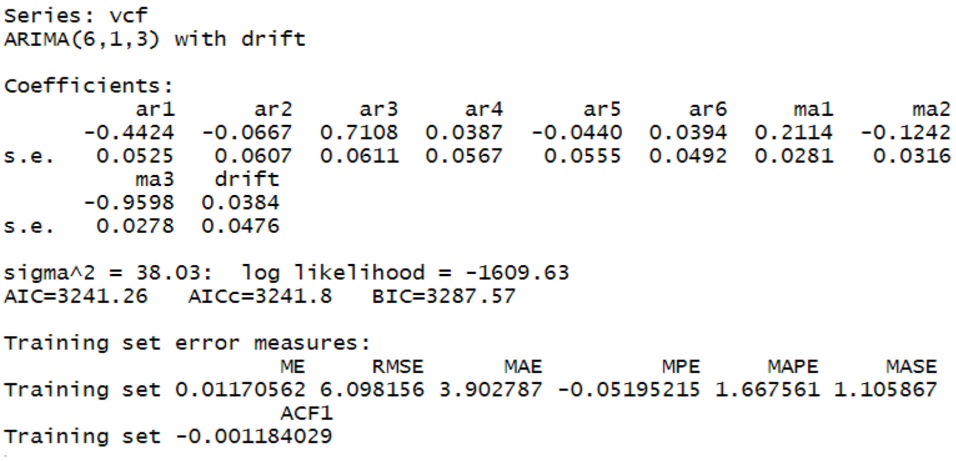
Try Criteria AIC



```
         0          1          2          3          4          5          6 
27.6991856 13.7049440 14.5599393  6.7559681  5.2692482  2.7211347  0.0000000 
         7          8          9         10         11         12 
 1.3889343  0.7114887  1.8751970  2.1405589  3.8237410  5.2004217 
```

AIC (6) = 0.000000 min Model: AR(6)

3.5. Estimate ARIMA (6,1,3) model



```
Series: vcf 
ARIMA(6,1,3) with drift 

Coefficients:
          ar1      ar2     ar3     ar4      ar5     ar6     ma1      ma2
      -0.4424  -0.0667  0.7108  0.0387  -0.0440  0.0394  0.2114  -0.1242
s.e.   0.0525   0.0607  0.0611  0.0567   0.0555  0.0492  0.0281   0.0316
          ma3   drift
      -0.9598  0.0384
s.e.   0.0278  0.0476

sigma^2 = 38.03:  log likelihood = -1609.63
AIC=3241.26   AICc=3241.8   BIC=3287.57

Training set error measures:
                     ME     RMSE      MAE         MPE     MAPE     MASE
Training set 0.01170562 6.098156 3.902787 -0.05195215 1.667561 1.105867
                     ACF1
Training set -0.001184029
```

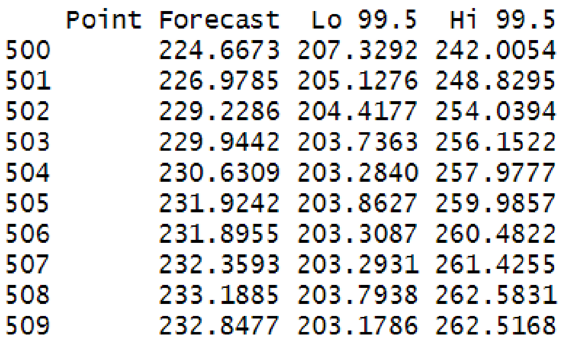
AIC=3241.26; RMSE = 6.098156; MAPE = 1.667561

Compare ARIMA (6,1,2) model if smaller AIC better model

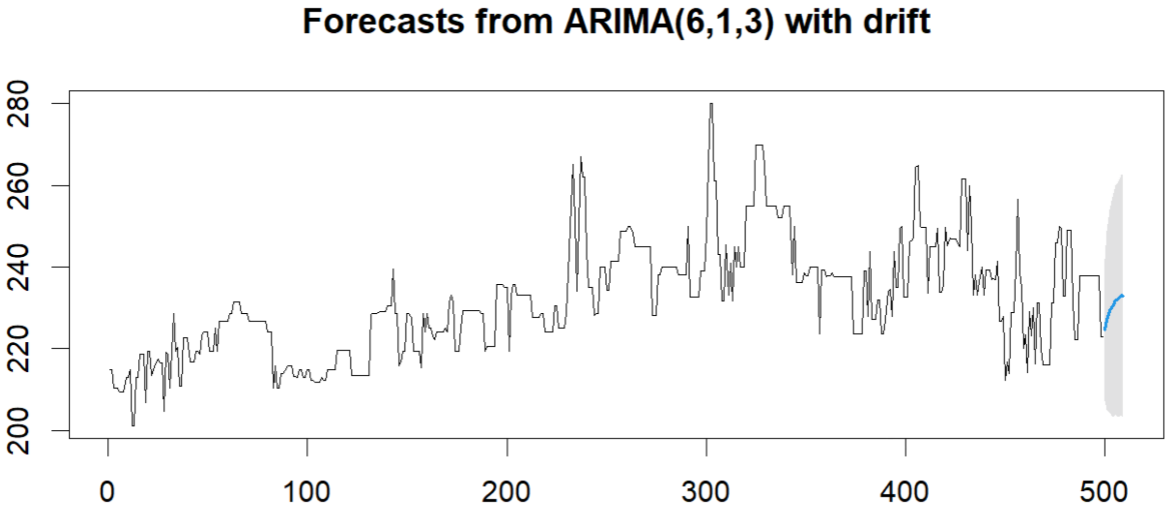
AIC=3244.71; RMSE = 6.152825; MAPE = 1.680848

From the result, it is clear that model ARIMA (6,1,3) is better than model ARIMA (6,1,2).

3.6. Forecast for the first 10 observations in 2023



```
    Point Forecast  Lo 99.5  Hi 99.5
500       224.6673 207.3292 242.0054
501       226.9785 205.1276 248.8295
502       229.2286 204.4177 254.0394
503       229.9442 203.7363 256.1522
504       230.6309 203.2840 257.9777
505       231.9242 203.8627 259.9857
506       231.8955 203.3087 260.4822
507       232.3593 203.2931 261.4255
508       233.1885 203.7938 262.5831
509       232.8477 203.1786 262.5168
```

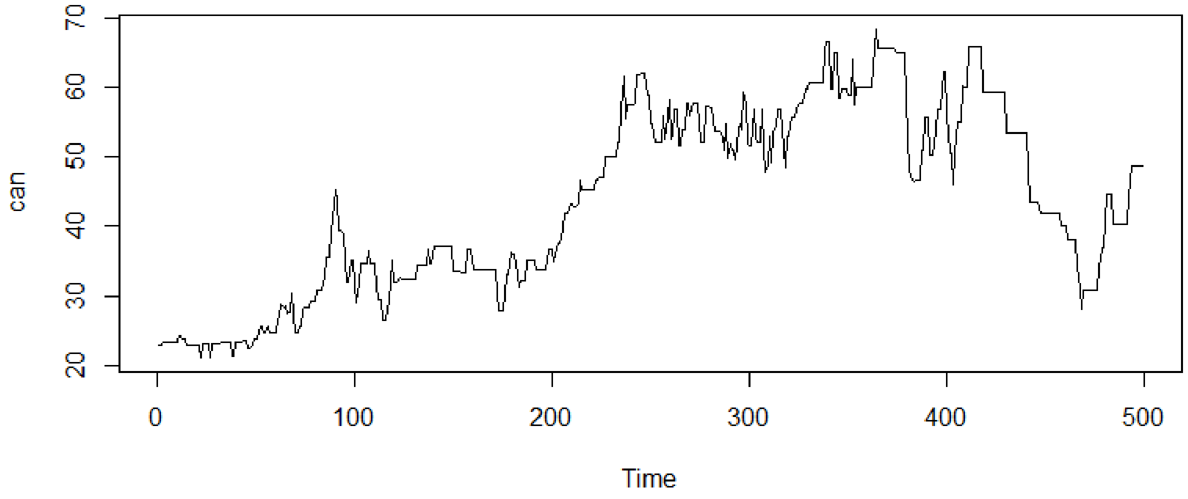


II. Hoàng Diệu Linh

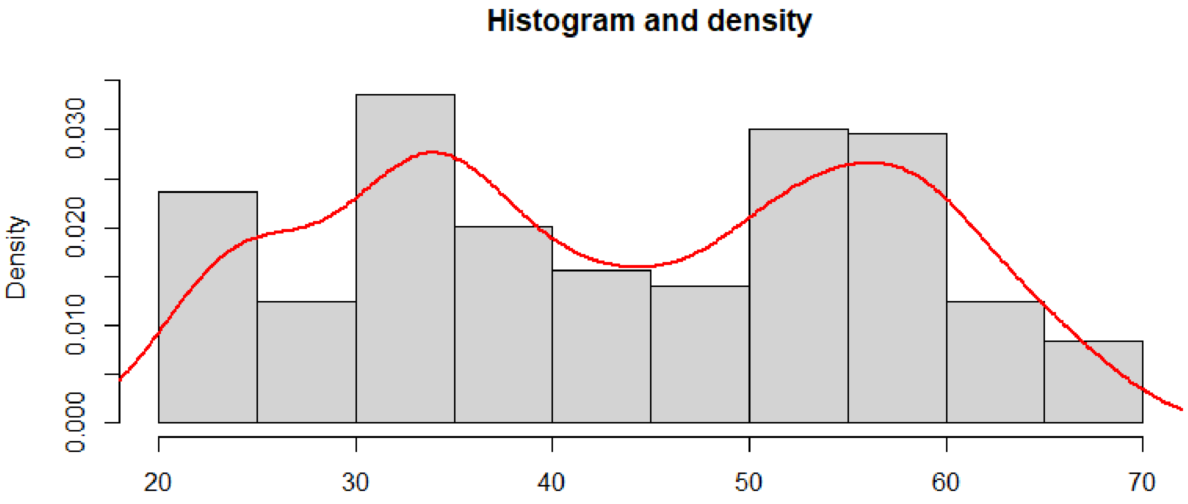
1. CTCP Đồ hộp Hạ Long (HNX: CAN)

Established in 1957, formerly known as Ha Long Canned Fish Factory, Ha Long Canned Joint Stock Company (Halong Canfoco) is considered as one of the first real canned food manufacturers in Vietnam. Today, along with the strong development of the country, Halong Canfoco is one of the first companies listed on the stock market with nearly 1,000 employees, 2 factories. The company's products are diverse from canned products such as: fish, meat, vegetables and fruits, pasteurized sausages to frozen products such as frozen sausages, spring rolls or seafood spring rolls. . The company's products are present in all provinces and cities and are exported to nearly every continent, from Europe, Asia, to the Middle East, Africa…

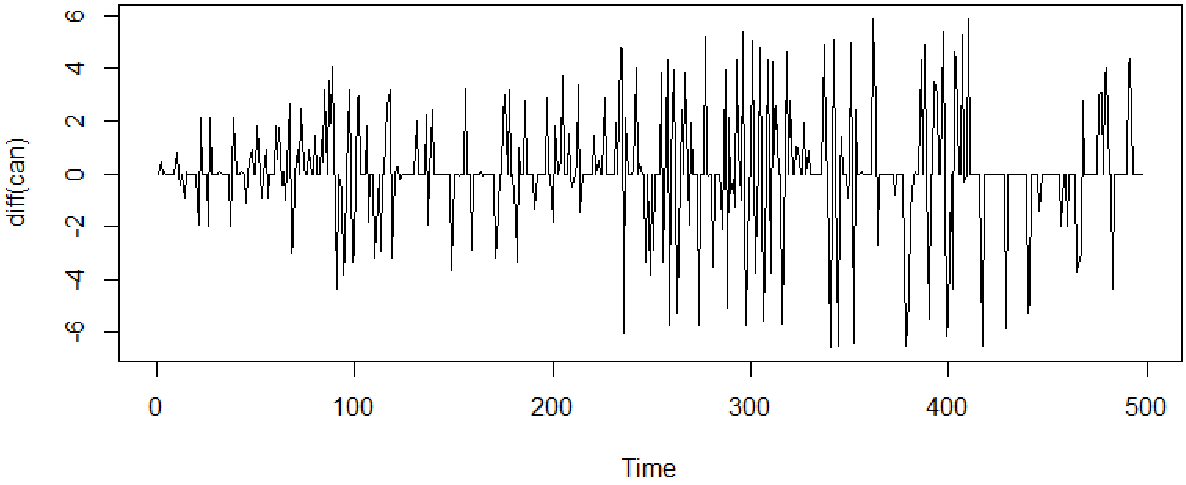
2. Finance series



The series shows a very distinct upswing from the first 400 observations; the remaining data show a very clear downturn. In general, the series tends to be volatile and unstable. Moreover, the storyline does not adequately depict the seasonal component.



After plotting the histogram of the series, we can see that the above series is a multivariate gaussian distribution



After the difference of series, we have a series that oscillates around the number 0. By subtracting the difference, we may deduce that we have a stationary series.

3. Models to forecast 8 observations in 2023

a. Linear - log

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   -7.8942     2.2368  -3.529 0.000455 ***
log(time)      9.8400     0.4212  23.364  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.172 on 497 degrees of freedom
Multiple R-squared:  0.5234,	Adjusted R-squared:  0.5225 
F-statistic: 545.9 on 1 and 497 DF,  p-value: < 2.2e-16
```

We have model: $CAN = -7.8942 + 9.84 * ln(t) + u_t$

b. Log - log

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 2.358089   0.049716   47.43   <2e-16 ***
log(time)   0.261174   0.009361   27.90   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2039 on 497 degrees of freedom
Multiple R-squared:  0.6103,	Adjusted R-squared:  0.6095 
F-statistic: 778.4 on 1 and 497 DF,  p-value: < 2.2e-16
```

We have model: $ln(CAN) = 2.358089 + 0.261174 * ln(t) + u_t$

From model we can calculate the first 8 observations:

t	CAN (Linear – log)	CAN (Log – log)
500	53.257	53.58

501	53.277	53.608
502	53.296	53.636
503	53.316	53.664
504	53.335	53.692
505	53.55	53.719
506	53.374	53.747
507	53.394	53.775

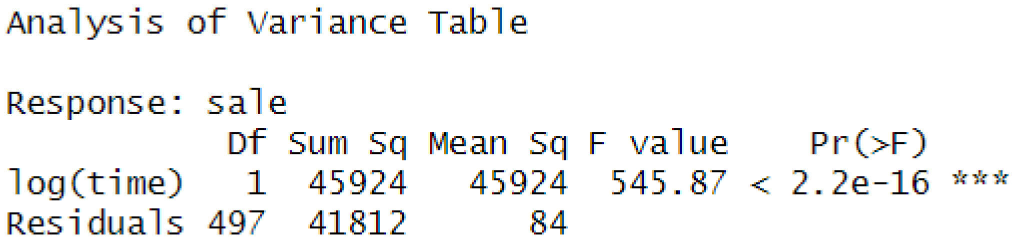
c. Compare models:

Model linear - log have: RMSE = 9.154 and MAPE = 0.196

Model log - log have: RMSE = 8.9 and MAPE = 0.175

Because both RMSE and MAPE of log - log model is smaller than linear - log model, so log - log can predict better.

Using ANOVA to test



```
Analysis of Variance Table

Response: sale
           Df Sum Sq Mean Sq F value    Pr(>F)    
log(time)   1  45924   45924  545.87 < 2.2e-16 ***
Residuals 497  41812      84
```

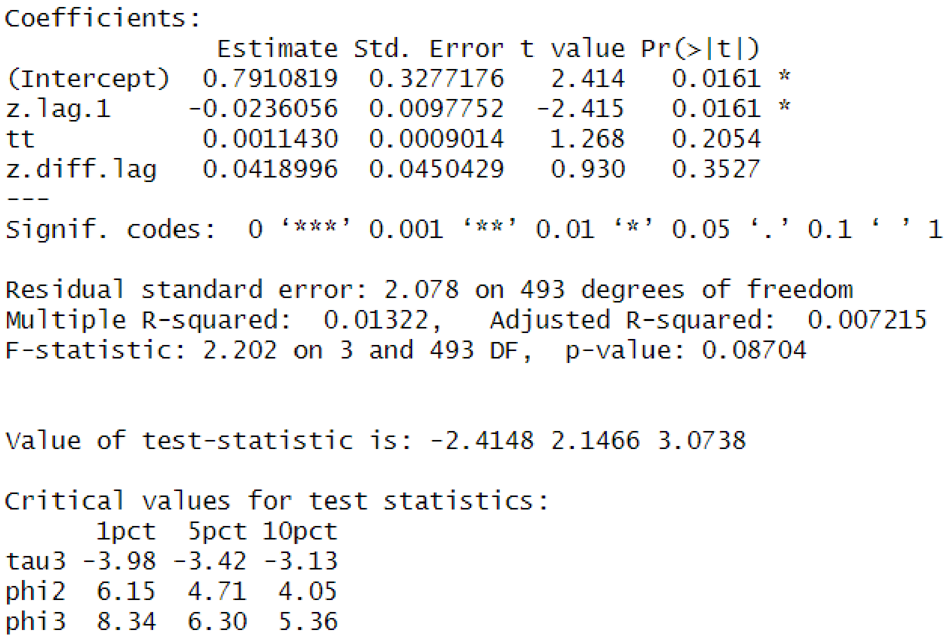
As you can see, the result shows a Df of 1 (indicating that the more complex model has one additional parameter), and a very small p-value (< .001). This means that the model log - log did lead to a significantly improved fit over the model linear - log.

4. ARIMA test for CAN

a. Unit root test

To verify that the series is stationary, we employ unit root testing. To conduct this test, we shall make use of the Dickey-Fuller test.

Test with CAN



```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.7910819  0.3277176   2.414   0.0161 *
z.lag.1     -0.0236056  0.0097752  -2.415   0.0161 *
tt           0.0011430  0.0009014   1.268   0.2054
z.diff.lag   0.0418996  0.0450429   0.930   0.3527
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.078 on 493 degrees of freedom
Multiple R-squared:  0.01322,   Adjusted R-squared:  0.007215
F-statistic: 2.202 on 3 and 493 DF,  p-value: 0.08704


Value of test-statistic is: -2.4148 2.1466 3.0738

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.98 -3.42 -3.13
phi2  6.15  4.71  4.05
phi3  8.34  6.30  5.36
```

$$\{H_0: \delta = 0 : \ CAN\ has\ unit\ root \quad H_1: \delta < 0 : \ CAN\ has\ not\ unit\ root$$

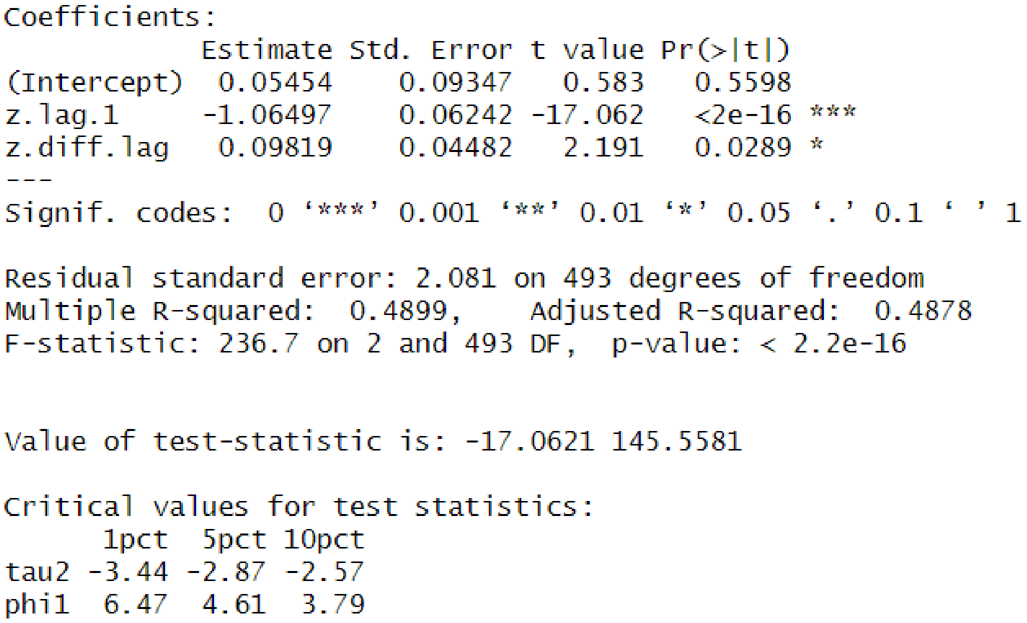
$|\tau_{stat}| = 2.4148 < \ |\tau_{0.05}| = 3.42$

→ Not reject Ho → Can has unit root → Not stationary

After testing unit root for CAN series, we have found out that CAN has unit root, which means that the series is not stationary.

We must turn the series into a stationary series in order to apply the series to the model. We obtain a stationary series—one without trend factors—by taking the series' difference once. The values of the series are centered on the series mean.

Test with difference of CAN



```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.05454    0.09347   0.583   0.5598
z.lag.1     -1.06497    0.06242 -17.062   <2e-16 ***
z.diff.lag   0.09819    0.04482   2.191   0.0289 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.081 on 493 degrees of freedom
Multiple R-squared:  0.4899,    Adjusted R-squared:  0.4878
F-statistic: 236.7 on 2 and 493 DF,  p-value: < 2.2e-16


Value of test-statistic is: -17.0621 145.5581

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.44 -2.87 -2.57
phi1  6.47  4.61  3.79
```

$$\{H_0: \delta = 0 : \ CAN\ has\ unit\ root \quad H_1: \delta < 0 : \ CAN\ has\ not\ unit\ root$$

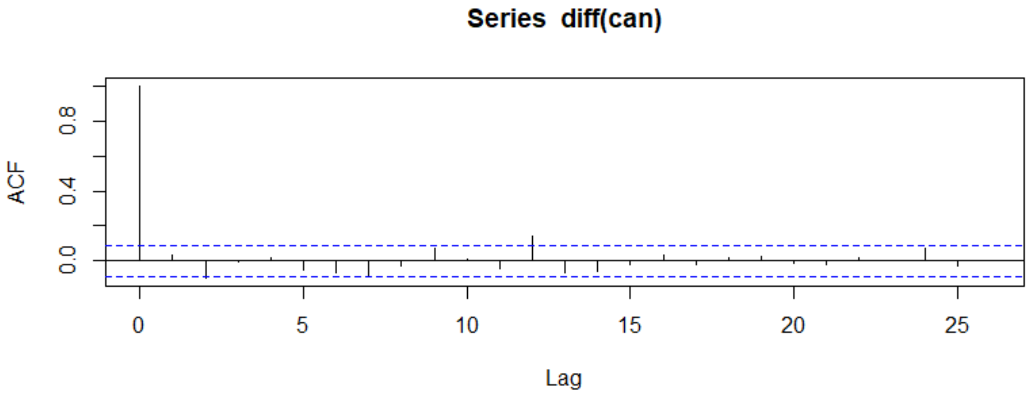
$|\tau_{stat}| = 17.0621 < \ |\tau_{0.05}| = 2.87$

→ Reject Ho → No unit root → Stationary

After testing unit root for the difference of CAN series, We have found out that the difference of CAN does not have unit root, which means that the series is stationary. So that we will use the difference of CAN for the ARIMA model.

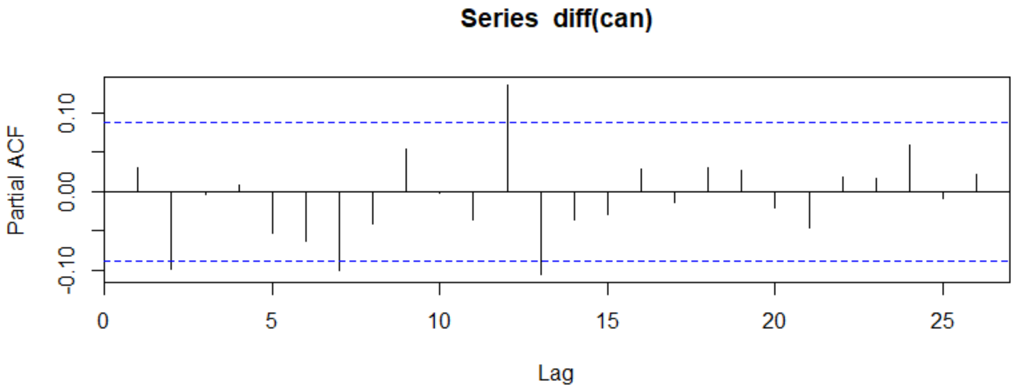
b. ACF & PACF

ACF: explains how the present value of a given time series is correlated with the past. ACF plot is a bar chart of coefficients of correlation between a time series and its lagged values. From the ACF plot we can identify the order of MA for the ARIMA model.



→ From the graph, we can conclude that MA = q = 12

PACF: is the partial autocorrelation function that explains the partial correlation between the series and lags itself. From the PACF plot we can identify the order of AR for the ARIMA model.

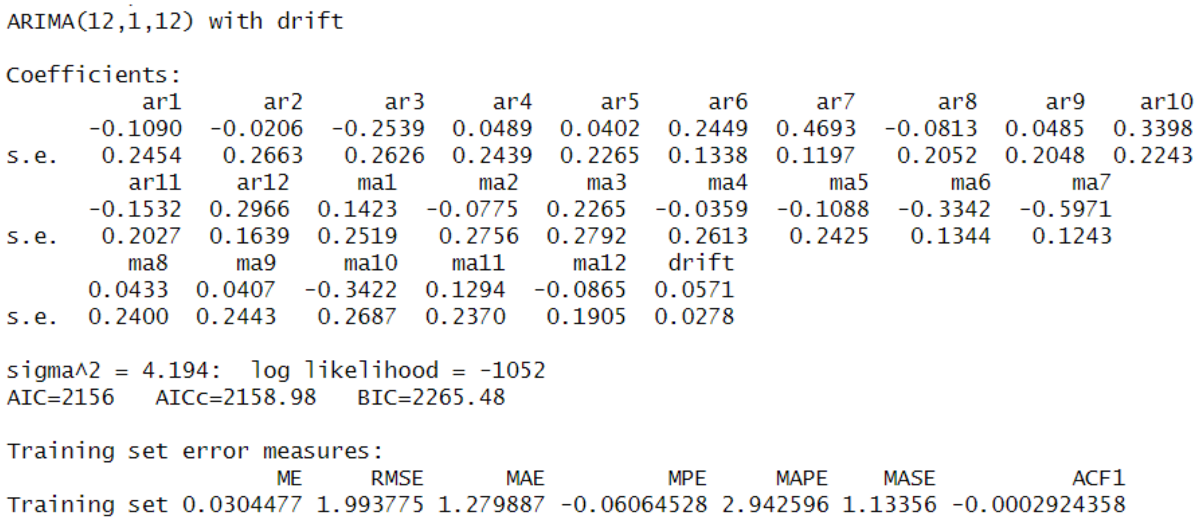


→ From the graph, we can conclude that AR = p = 12

→ Our ARIMA model is ARIMA (12,1,12)

c. Estimate ARIMA model

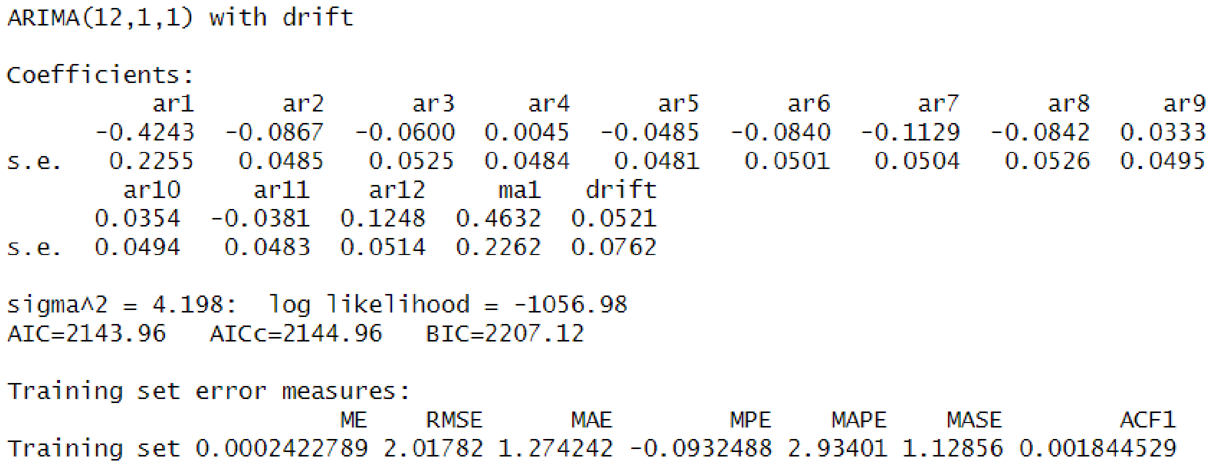
ARIMA model (12,1,12)



```
ARIMA(12,1,12) with drift

Coefficients:
          ar1      ar2      ar3     ar4     ar5     ar6     ar7      ar8     ar9    ar10
      -0.1090  -0.0206  -0.2539  0.0489  0.0402  0.2449  0.4693  -0.0813  0.0485  0.3398
s.e.   0.2454   0.2663   0.2626  0.2439  0.2265  0.1338  0.1197   0.2052  0.2048  0.2243
         ar11    ar12     ma1      ma2     ma3      ma4      ma5      ma6      ma7
      -0.1532  0.2966  0.1423  -0.0775  0.2265  -0.0359  -0.1088  -0.3342  -0.5971
s.e.   0.2027  0.1639  0.2519   0.2756  0.2792   0.2613   0.2425   0.1344   0.1243
         ma8     ma9     ma10    ma11     ma12   drift
      0.0433  0.0407  -0.3422  0.1294  -0.0865  0.0571
s.e.  0.2400  0.2443   0.2687  0.2370   0.1905  0.0278

sigma^2 = 4.194:  log likelihood = -1052
AIC=2156   AICc=2158.98   BIC=2265.48

Training set error measures:
                    ME     RMSE      MAE         MPE     MAPE    MASE          ACF1
Training set 0.0304477 1.993775 1.279887 -0.06064528 2.942596 1.13356 -0.0002924358
```

→ AIC = 2156, RMSE = 1.993775, MAPE = 2.942596

ARIMA model (12,1,1)



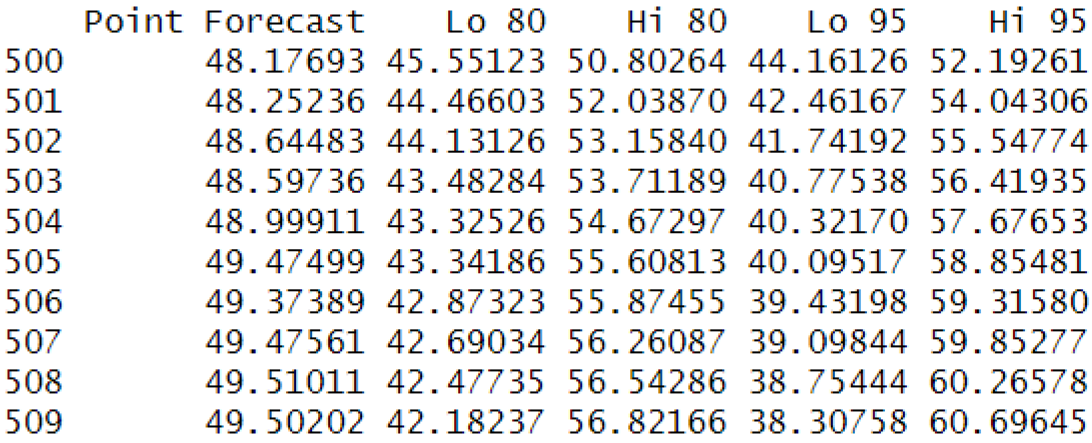
```
ARIMA(12,1,1) with drift

Coefficients:
          ar1      ar2      ar3     ar4      ar5      ar6      ar7      ar8     ar9
      -0.4243  -0.0867  -0.0600  0.0045  -0.0485  -0.0840  -0.1129  -0.0842  0.0333
s.e.   0.2255   0.0485   0.0525  0.0484   0.0481   0.0501   0.0504   0.0526  0.0495
         ar10     ar11    ar12     ma1   drift
       0.0354  -0.0381  0.1248  0.4632  0.0521
s.e.   0.0494   0.0483  0.0514  0.2262  0.0762

sigma^2 = 4.198:  log likelihood = -1056.98
AIC=2143.96   AICc=2144.96   BIC=2207.12

Training set error measures:
                       ME    RMSE      MAE        MPE    MAPE    MASE        ACF1
Training set 0.0002422789 2.01782 1.274242 -0.0932488 2.93401 1.12856 0.001844529
```

→ AIC = 2143.96, RMSE = 2.01782, MAPE = 2.93401

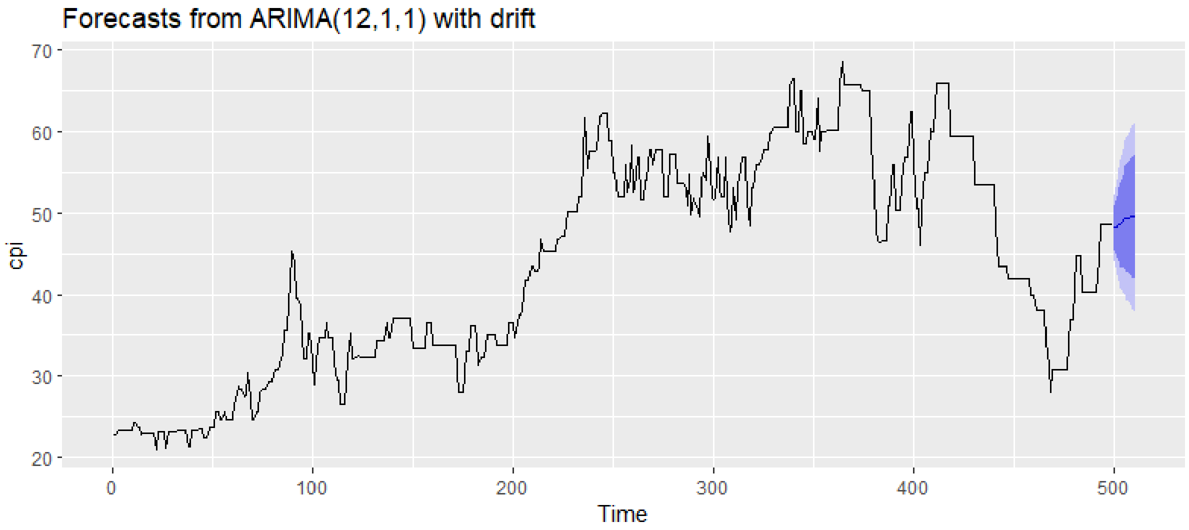
→ AIC of model ARIMA (12,1,12) higher than model ARIMA (12,1,1) → Model ARIMA (12,1,1) has better performance.

Thus, we will select the model ARIMA (12,1,1) to predict the subsequent 10 observations.

d. Forecast for the first 10 observations in 2023



```
    Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
500       48.17693 45.55123 50.80264 44.16126 52.19261
501       48.25236 44.46603 52.03870 42.46167 54.04306
502       48.64483 44.13126 53.15840 41.74192 55.54774
503       48.59736 43.48284 53.71189 40.77538 56.41935
504       48.99911 43.32526 54.67297 40.32170 57.67653
505       49.47499 43.34186 55.60813 40.09517 58.85481
506       49.37389 42.87323 55.87455 39.43198 59.31580
507       49.47561 42.69034 56.26087 39.09844 59.85277
508       49.51011 42.47735 56.54286 38.75444 60.26578
509       49.50202 42.18237 56.82166 38.30758 60.69645
```



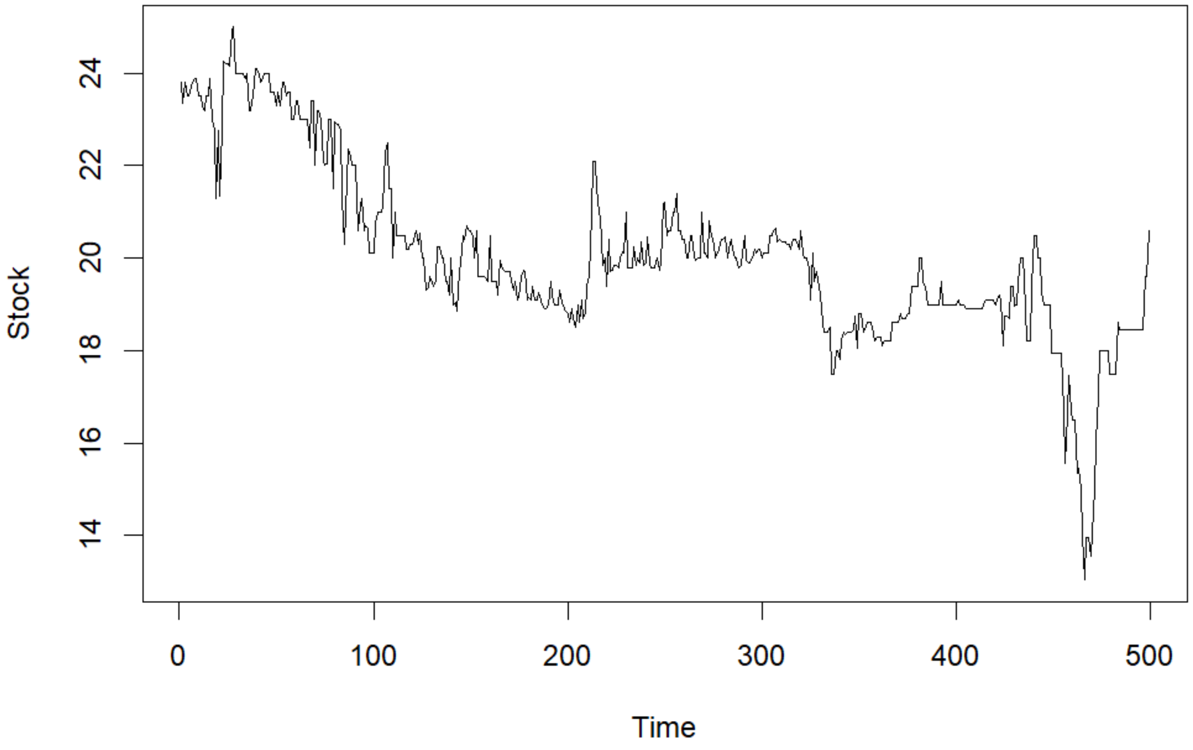
III. Nguyen Thi Thanh Hien

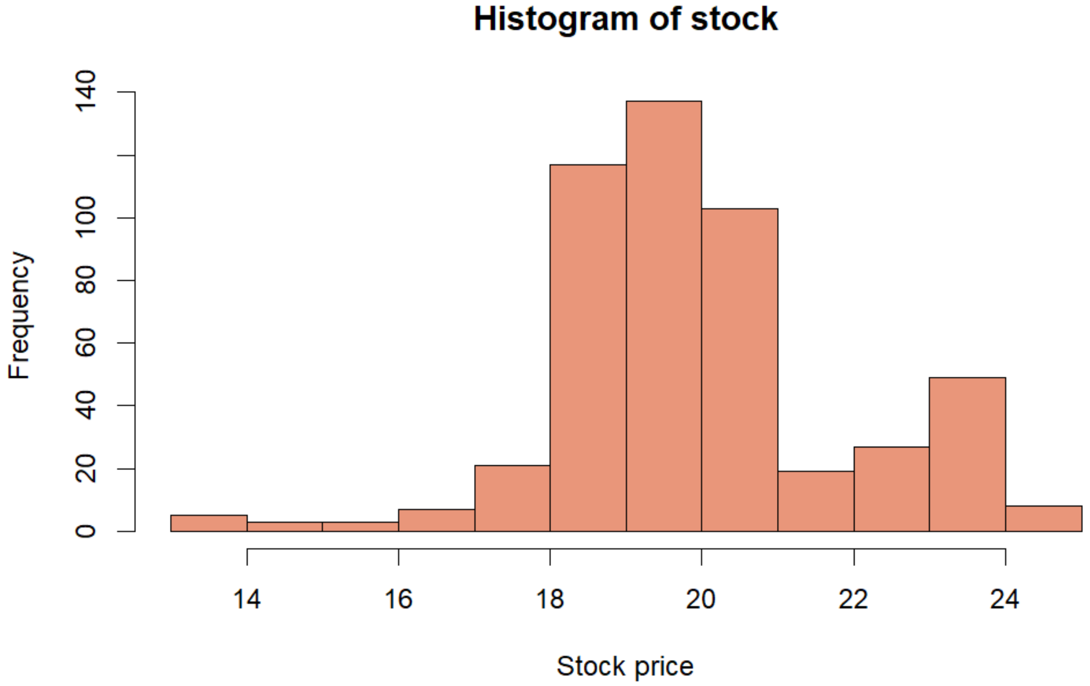
1. Chuong Duong Beverages Joint Stock Company (CDBECO)

Chuong Duong Beverages Joint Stock Company (CDBECO) is a Vietnam-based manufacturer of beverages. Its products include soft drinks, fruit drinks, mild wine and purified drinking water. The Company also offers bottling, packaging and related services to other beverage producers, as well as trades supplies and materials for the beverage industry. Moreover, it is involved in real estate trading and the offering of real estate brokerage services. As of December 31, 2012, the Company was a 51%-owned subsidiary of Saigon Beer - Alcohol - Beverage Joint Stock Corporation (SABECO). As of the same date, it had three branches located in Ho Chi Minh City, Binh Duong Province and Vinh Long Province, Vietnam.

2. Finance Series

This data shows the stock price of Chuong Duong Beverages Joint Stock Company in 2021 and 2022.





The distribution of the SCD series looks like the normal distribution except that it has a large peak at one tail. Usually this is caused by faulty construction of the histogram, with data lumped together into a group labeled "greater than."

Forecast by using Linear - Linear model:

501	17.69
502	17.68

503	17.67
504	17.66
505	17.65
506	17.64
507	17.63
508	17.62

Forecast by using Linear - Log model

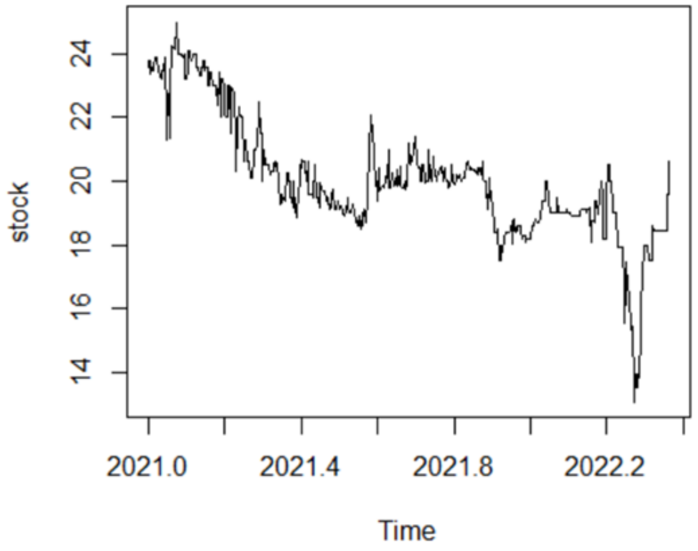
501	18.436
502	18.432
503	18.429
504	18.426
505	18.423
506	18.420
507	18.417
508	18.414

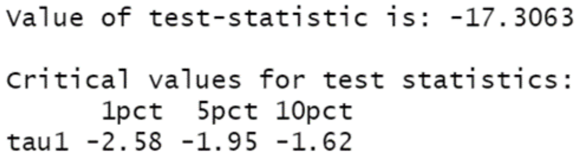
3. Stock Price

This data shows the stock price of Chuong Duong Beverages Joint Stock Company in 2021 and 2022.

a. Unit Root Test

Unit root test for SCD

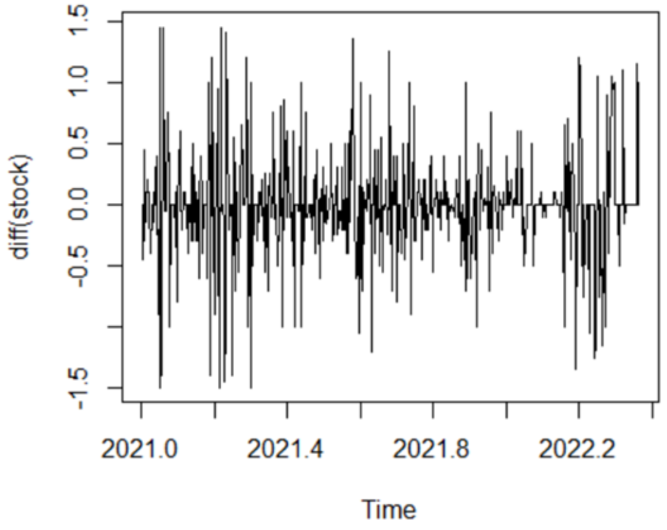


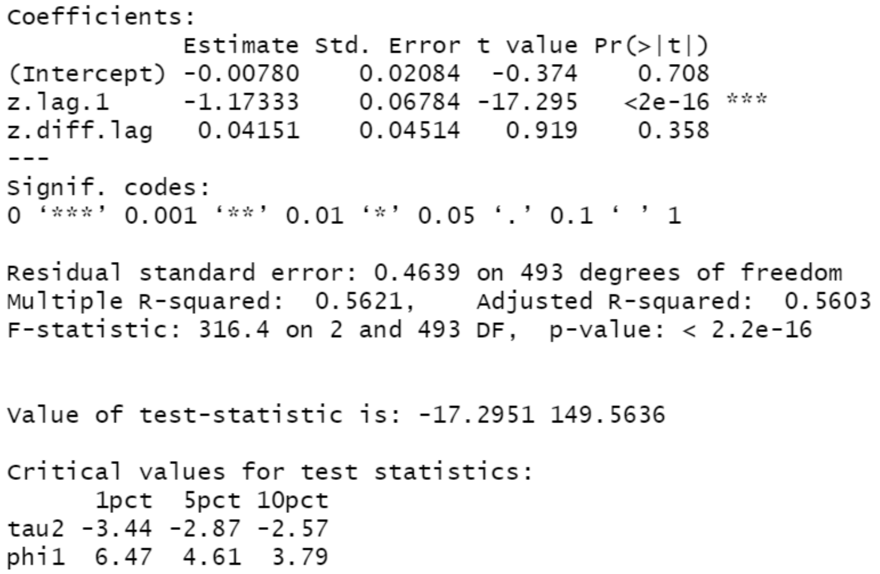


```
Value of test-statistic is: -17.3063 

Critical values for test statistics: 
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

$$|T_{stat}| = 17.3063 > crit \Rightarrow reject\ H_0 \Rightarrow \begin{cases} scd\ non -\ stationary \\ \Delta scd\ stationary \end{cases} \Rightarrow scd\colon I(1)$$

Unit Root Test for Different Series





```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.00780    0.02084  -0.374    0.708    
z.lag.1     -1.17333    0.06784 -17.295   <2e-16 ***
z.diff.lag   0.04151    0.04514   0.919    0.358    
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4639 on 493 degrees of freedom
Multiple R-squared:  0.5621,	Adjusted R-squared:  0.5603 
F-statistic: 316.4 on 2 and 493 DF,  p-value: < 2.2e-16


Value of test-statistic is: -17.2951 149.5636 

Critical values for test statistics: 
      1pct  5pct 10pct
tau2 -3.44 -2.87 -2.57
phi1  6.47  4.61  3.79
```

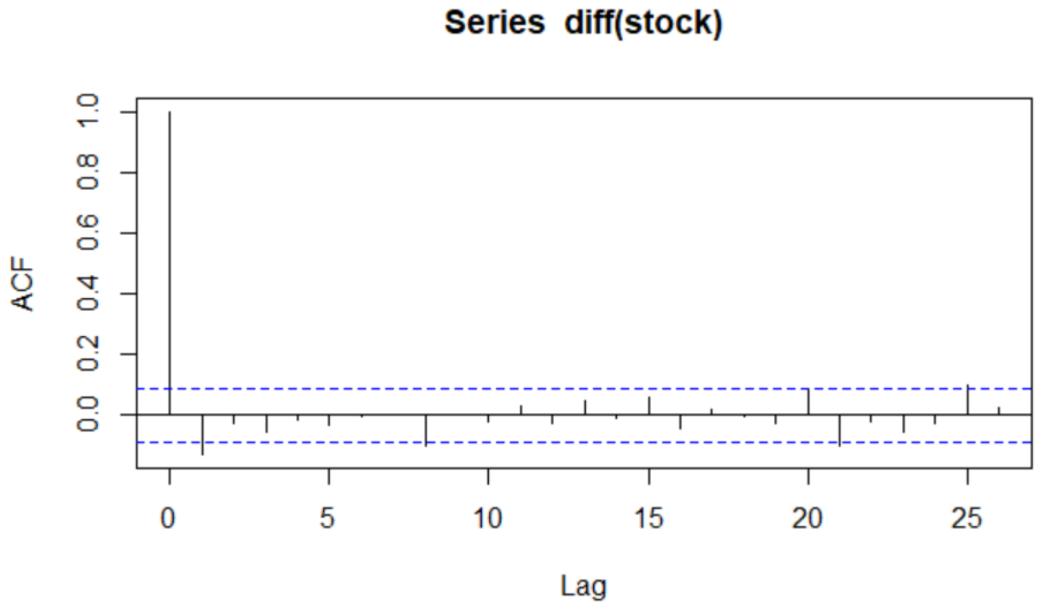
$$\begin{cases} H_0: \delta = 0: \text{SCD has unit root} \\ H_1: \delta < 0: \text{SCD has not unit root} \end{cases}$$

$$|\tau_{ur}| = 17.2951 > 3.44 = \tau_{0.05} \Rightarrow Reject\ H_0$$

$$Therefore, SCD\ series\ has\ not\ unit\ root\ and\ it\ is\ trend\ stationary$$

b. ACF and PACF

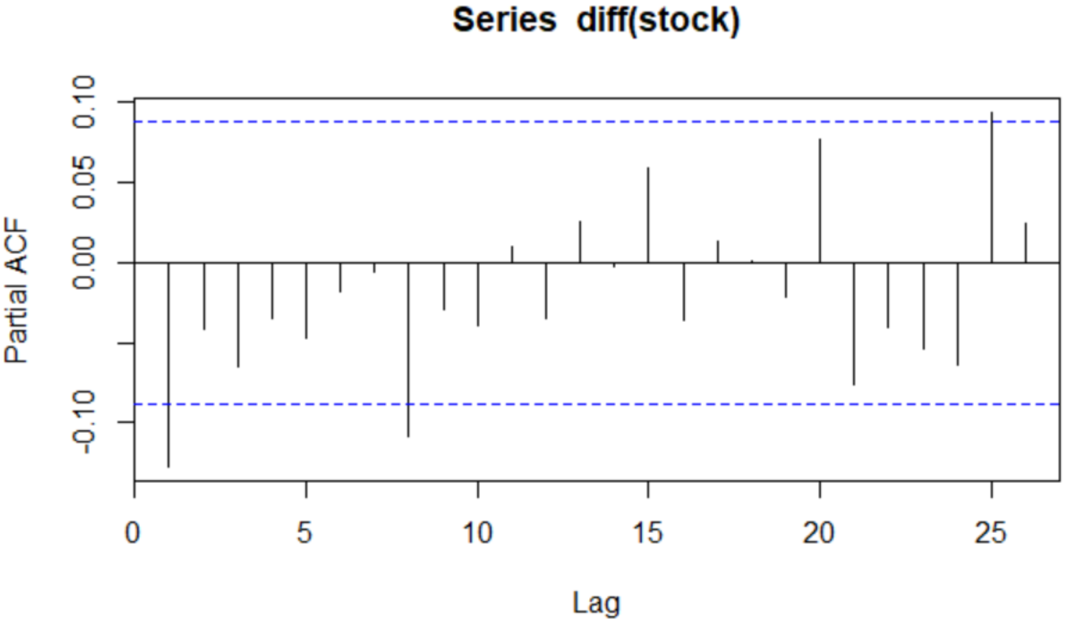
ACF: Autocorrelation is the correlation between two values in a time series. In other words, the time series data correlate with themselves—hence, the name. We talk about these correlations using the term “lags.” Analysts record time-series data by measuring a characteristic at evenly spaced intervals—such as daily, monthly, or yearly. The number of intervals between the two observations is the lag. The autocorrelation function (ACF) assesses the correlation between observations in a time series for a set of lags.



The x-axis corresponds to the different lags of the residuals. Whereas the y-axis shows the correlation of each lag. Finally, the dashed blue line represents the significance level.

After the lag-0 correlation, the subsequent correlations drop quickly to zero and stay (mostly) between the limits of the significance level (dashed blue lines). Therefore, we can conclude that the residuals of this model meet the assumption of no autocorrelation.

PACF: The partial autocorrelation function is similar to the ACF except that it displays only the correlation between two observations that the shorter lags between those observations do not explain. The partial autocorrelation function (PACF) is more useful during the specification process for an autoregressive model.

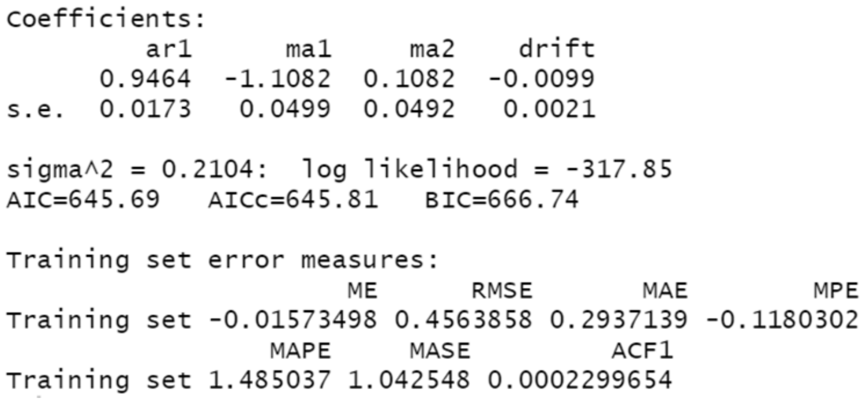


c. ARIMA (1, 1, 2) model

ARIMA is a method for forecasting or predicting future outcomes based on a historical time series. It is based on the statistical concept of serial correlation, where past data points influence future data points.

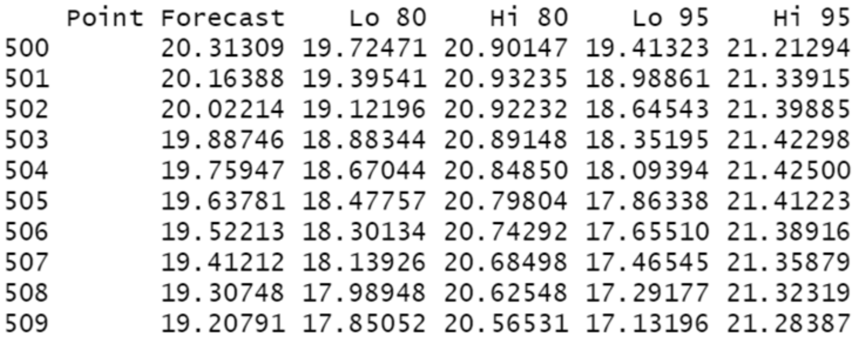
An ARIMA model can be understood by outlining each of its components as follows:

- Autoregression (AR): refers to a model that shows a changing variable that regresses on its own lagged, or prior, values.
- Integrated (I): represents the differencing of raw observations to allow the time series to become stationary (i.e., data values are replaced by the difference between the data values and the previous values).
- Moving average (MA): incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations.



```
Coefficients:
         ar1      ma1     ma2    drift
      0.9464  -1.1082  0.1082  -0.0099
s.e.  0.0173   0.0499  0.0492   0.0021

sigma^2 = 0.2104:  log likelihood = -317.85
AIC=645.69   AICc=645.81   BIC=666.74

Training set error measures:
                      ME      RMSE       MAE        MPE
Training set -0.01573498 0.4563858 0.2937139 -0.1180302
                 MAPE     MASE          ACF1
Training set 1.485037 1.042548 0.0002299654
```

d. Forecast for the first 10 observations in 2023



```
    Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
500       20.31309 19.72471 20.90147 19.41323 21.21294
501       20.16388 19.39541 20.93235 18.98861 21.33915
502       20.02214 19.12196 20.92232 18.64543 21.39885
503       19.88746 18.88344 20.89148 18.35195 21.42298
504       19.75947 18.67044 20.84850 18.09394 21.42500
505       19.63781 18.47757 20.79804 17.86338 21.41223
506       19.52213 18.30134 20.74292 17.65510 21.38916
507       19.41212 18.13926 20.68498 17.46545 21.35879
508       19.30748 17.98948 20.62548 17.29177 21.32319
509       19.20791 17.85052 20.56531 17.13196 21.28387
```

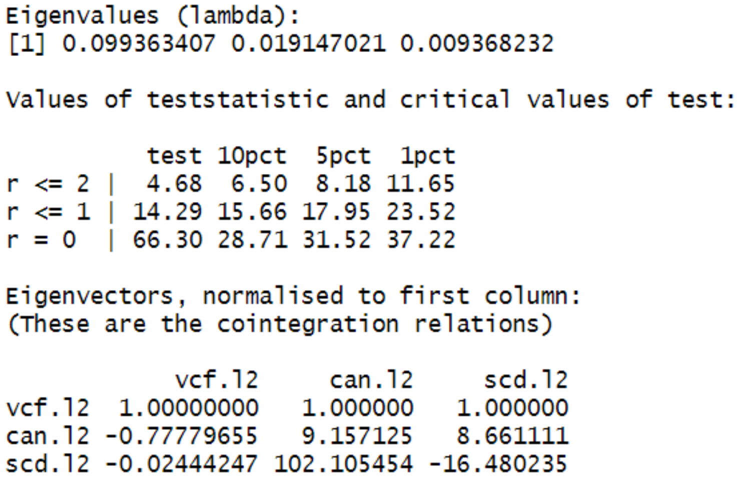
B. Group

I. Test for cointegration

Cointegration: A relationship between two non-stationary time series is possible. Two-time series that are impacted by the same underlying source, such as market forces, may “move together” over time. This relationship is referred to as co-integration. We will use Johansen test to do this test

1. Johansen test

a. Johansen test using “trace” criteria



```
Eigenvalues (lambda):
[1] 0.099363407 0.019147021 0.009368232

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 2 |  4.68  6.50  8.18 11.65
r <= 1 | 14.29 15.66 17.95 23.52
r = 0  | 66.30 28.71 31.52 37.22

Eigenvectors, normalised to first column:
(These are the cointegration relations)

            vcf.l2     can.l2     scd.l2
vcf.l2  1.00000000   1.000000   1.000000
can.l2 -0.77779655   9.157125   8.661111
scd.l2 -0.02444247 102.105454 -16.480235
```

The number of cointegrate is 0 (r = 0)

$$\{H_0: Number\ of\ cointegrate = 0\ H_1: Number\ of\ cointegrate > 0$$

$|\lambda_{stat}| = 66.3 > |\lambda_{0.05}| = 31.52$ Reject Ho Number of cointegrate > 0

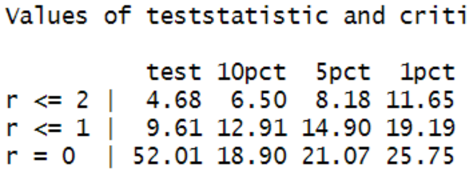
The number of cointegrate is 1 (r = 0)

$$\{H_0: Number\ of\ cointegrate \leq 1\ H_1: Number\ of\ cointegrate > 1$$

$|\lambda_{stat}| = 14.29 < \ |\lambda_{0.05}| = 17.95$ Not reject Ho Number of cointegrate = 1

 There are 1 cointegrate between 3 series

b. Johansen test using “eigen value” criteria



```
Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 2 |  4.68  6.50  8.18 11.65
r <= 1 |  9.61 12.91 14.90 19.19
r = 0  | 52.01 18.90 21.07 25.75
```

The number of cointegrate is 0 (r = 0)

$$\{H_0: Number\ of\ cointegrate = 0\ H_1: Number\ of\ cointegrate > 0$$

$|\lambda_{stat}| = 52.01 > |\lambda_{0.05}| = 21.07$ Reject Ho Number of cointegrate > 0

The number of cointegrate is 1 (r = 0)

$$\{H_0: Number\ of\ cointegrate \leq 1\ H_1: Number\ of\ cointegrate > 1$$

$|\lambda_{stat}| = 9.61 < \ |\lambda_{0.05}| = 14.90$ Not reject Ho Number of cointegrate = 1

 There are 1 cointegrate between 3 series

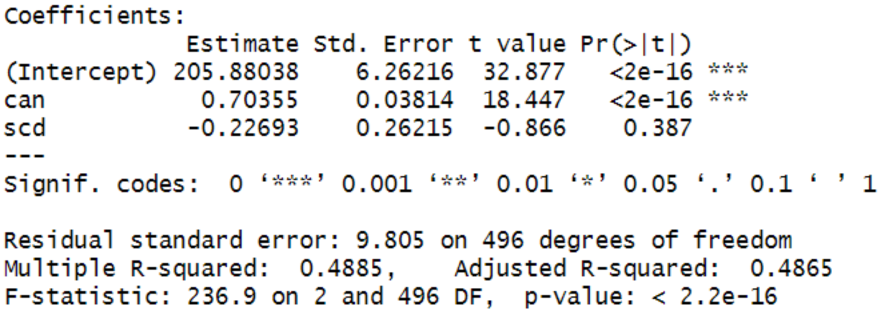
After using Johansen test with two criteria “trace” and “eigen value” we got the same results that there is one relationship among 3 series.

2. Unit root test

Based on the results of the individual part, we have the result that all 3 series have unit root, and the difference of 3 series has no unit root.

3. Test cointegrate

Series are cointegrated when their trends are not too far apart and are in some sense similar. This vague statement, though, can be made precise by conducting a cointegration test, which tests whether the residuals from regressing one series on the other one are stationary. If they are, the series are cointegrated. Thus, a cointegration test is in fact a Dickey-Fuller stationarity test on residuals, and its null hypothesis is of non-cointegration.

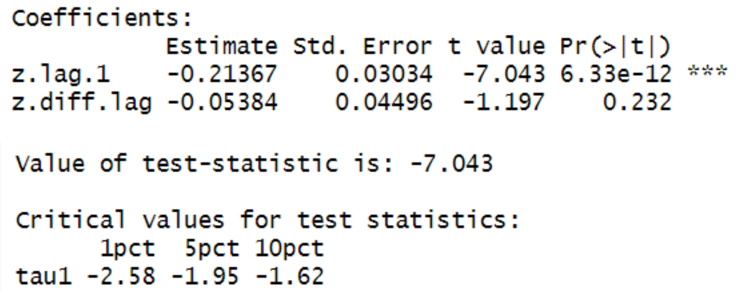


```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 205.88038    6.26216  32.877   <2e-16 ***
can           0.70355    0.03814  18.447   <2e-16 ***
scd          -0.22693    0.26215  -0.866    0.387    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.805 on 496 degrees of freedom
Multiple R-squared:  0.4885,	Adjusted R-squared:  0.4865 
F-statistic: 236.9 on 2 and 496 DF,  p-value: < 2.2e-16
```

$$VCF_t = \ \beta_o + \ \beta_1 * CAN_t + \ \beta_2 * SCD_t + u_t$$

$$VCF_t = 205.88 + 0.7035 * CAN_t - 0.2269 * SCD_t + e_t$$

Stationary residual test



```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
z.lag.1     -0.21367    0.03034  -7.043 6.33e-12 ***
z.diff.lag  -0.05384    0.04496  -1.197    0.232    

Value of test-statistic is: -7.043 

Critical values for test statistics: 
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

$$\{H_0: u_t\ has\ unit\ root\ H_1: u_t\ has\ no\ unit\ root$$

$$|\tau_{stat}| = 7.043 > |\tau_{\alpha}| \rightarrow \quad Reject\ H_o \rightarrow \quad u_t\ is\ stationary$$

In absolute terms the test statistic value of 7.043 is higher than of 3 critical values (2.58, 1.95 and 1.62). Because residues are stationary, then series are coinciding, and it means that there exists some long term relationship between variables.

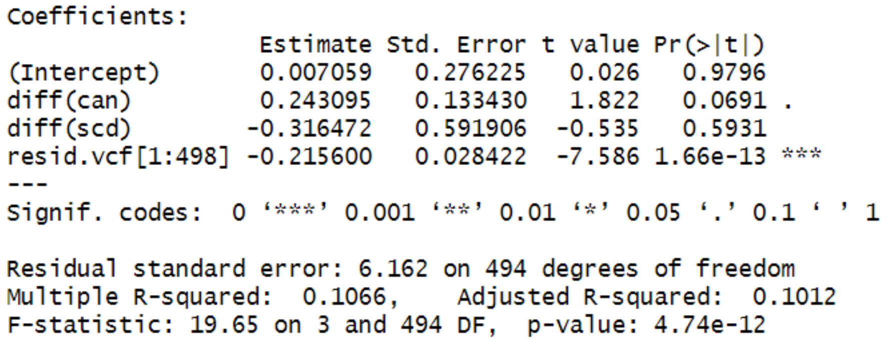
4. Error correction model

Cointegration implies that time series will be connected through an error correction model. The error correction model is important in time series analysis because it allows us to better understand long-run dynamics. Additionally, failing to properly model cointegrated variables can result in biased estimates. The error correction model:

- Reflects the long-run equilibrium relationships of variables.

- Includes a short-run dynamic adjustment mechanism that describes how variables adjust when they are out of equilibrium.

- Uses adjustment coefficients to measure the forces that push the relationship towards long-run equilibrium.



```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)        0.007059   0.276225   0.026   0.9796    
diff(can)          0.243095   0.133430   1.822   0.0691 .  
diff(scd)         -0.316472   0.591906  -0.535   0.5931    
resid.vcf[1:498] -0.215600   0.028422  -7.586 1.66e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.162 on 494 degrees of freedom
Multiple R-squared:  0.1066,	Adjusted R-squared:  0.1012 
F-statistic: 19.65 on 3 and 494 DF,  p-value: 4.74e-12
```

$$\Delta VCF_t = \alpha_o + \alpha_1 * \Delta CAN_t + \alpha_2 * \Delta SCD_t + \gamma * u_{t-1} + v_t \quad : \quad ECM$$

$$\Delta VCF_t = 0.007 + 0.243 * \Delta CAN_t - 0.316 * \Delta SCD_t - 0.2156 * u_{t-1} + e_t$$

Correction coefficient is -0.2156. So we can conclude that this is a weak correction. The adjusted speed of the VCF series is 21.6%. It means that after each period, VCF adjusts 21.6% of error.

II. VECTOR AUTOREGRESSIVE (VAR)

1. Granger causality test

The Granger causality test is a statistical hypothesis test for determining whether one time series is a factor and offer useful information in forecasting another time series.

Let $\mathit{\Delta} scd_t$ and $\mathit{\Delta} vcf_t$ be stationary time series. To test the null hypothesis that $\mathit{\Delta} vcf_t$ does not Granger-cause $\mathit{\Delta} scd_t$, one first finds the proper lagged values of $\mathit{\Delta} scd_t$ to include in an univariate autoregression of $\mathit{\Delta} SCD_t$:

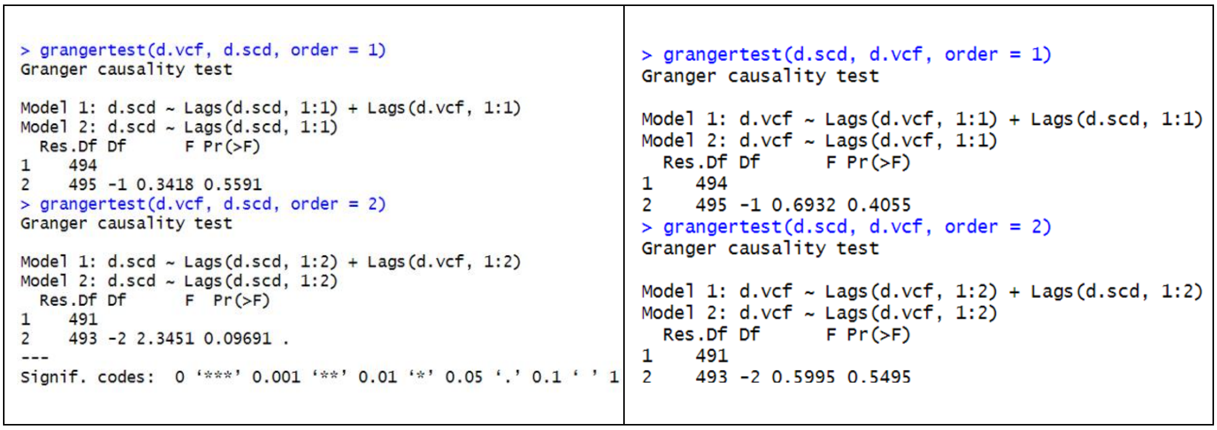
$$\mathit{\Delta} SCD_t = \beta_{20} + \beta_{221}\mathit{\Delta} SCD_{t-1} + v_{2t}$$

Next, the autoregression is augmented by including lagged values $\mathit{\Delta} vcf_t$:

$$\mathit{\Delta} SCD_t = \beta_{20} + \beta_{211}\mathit{\Delta} VCF_{t-1} + \beta_{221}\mathit{\Delta} SCD_{t-1} + v_{2t}$$

One retains in this regression all lagged values of $\mathit{\Delta} vcf_t$that are individually significant according to their t-statistics, provided that collectively they add explanatory power to the regression according to an F-test (whose null hypothesis is no explanatory power jointly added by the $\mathit{\Delta} vcf_t$'s). In the notation of the above augmented regression, *p* is the shortest, and *q* is the longest, lag length for which the lagged value of $\mathit{\Delta} vcf_t$is significant.

The null hypothesis that $\mathit{\Delta} vcf_t$ does not Granger-cause $\mathit{\Delta} scd_t$is accepted if and only if no lagged values of $\mathit{\Delta} vcf_t$are retained in the regression.



```
> grangertest(d.vcf, d.scd, order = 1)
Granger causality test

Model 1: d.scd ~ Lags(d.scd, 1:1) + Lags(d.vcf, 1:1)
Model 2: d.scd ~ Lags(d.scd, 1:1)
  Res.Df Df      F Pr(>F)
1    494
2    495 -1 0.3418 0.5591
> grangertest(d.vcf, d.scd, order = 2)
Granger causality test

Model 1: d.scd ~ Lags(d.scd, 1:2) + Lags(d.vcf, 1:2)
Model 2: d.scd ~ Lags(d.scd, 1:2)
  Res.Df Df      F  Pr(>F)
1    491
2    493 -2 2.3451 0.09691 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> grangertest(d.scd, d.vcf, order = 1)
Granger causality test

Model 1: d.vcf ~ Lags(d.vcf, 1:1) + Lags(d.scd, 1:1)
Model 2: d.vcf ~ Lags(d.vcf, 1:1)
  Res.Df Df      F Pr(>F)
1    494
2    495 -1 0.6932 0.4055
> grangertest(d.scd, d.vcf, order = 2)
Granger causality test

Model 1: d.vcf ~ Lags(d.vcf, 1:2) + Lags(d.scd, 1:2)
Model 2: d.vcf ~ Lags(d.vcf, 1:2)
  Res.Df Df      F Pr(>F)
1    491
2    493 -2 0.5995 0.5495
```

The hypothesis:

$$\{H_0 : \ \beta_{211} = 0 : \ \ \mathit{\Delta} vcf_t \ is \ not \ cause \ to \ \mathit{\Delta} scd_t \ \ H_1 : \ \beta_{211} \neq 0 \\ : \ \ \mathit{\Delta} vcf_t \ is \ \ cause \ to \ \mathit{\Delta} scd_t$$

p-value = 0.5591, Null hypothesis is not rejected hence $\mathit{\Delta} vcf_t \ is \ not \ cause \ to \ \mathit{\Delta} scd_t$

And $\mathit{\Delta} scd_t = \beta_{20} + \beta_{221}\mathit{\Delta} scd_{t-1} + v_{2t}$

$\mathit{\Delta} vcf_t$ explains to $\mathit{\Delta} scd_t$ and vice versa:

$$\mathit{\Delta} vcf_t = \beta_{10} + \beta_{111}\mathit{\Delta} vcf_{t-1} + v_{1t}$$

$$\mathit{\Delta} vcf_t = \beta_{10} + \beta_{111}\mathit{\Delta} vcf_{t-1} + \beta_{121}\mathit{\Delta} scd_{t-1} + v_{1t}$$

The hypothesis:

$$\{H_0 : \beta_{121} = 0 : \ \Delta scd_t \text{ is not cause to } \Delta vcf_t \ \ H_1 : \beta_{121} \neq 0 \\ : \ \Delta scd_t \text{ is cause to } \Delta vcf_t$$

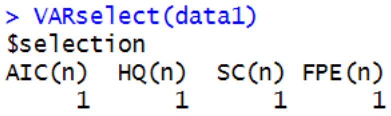
p-value = 0.4055, Null hypothesis is not rejected hence Δscd_t *is not cause to* Δvcf_t

And $\Delta vcf_t = \beta_{10} + \beta_{111}\Delta vcf_{t-1} + v_{1t}$

2. VAR (1) Model

VAR models (vector autoregressive models) are used for multivariate time series. The structure is that each variable is a linear function of past lags of itself and past lags of the other variables.

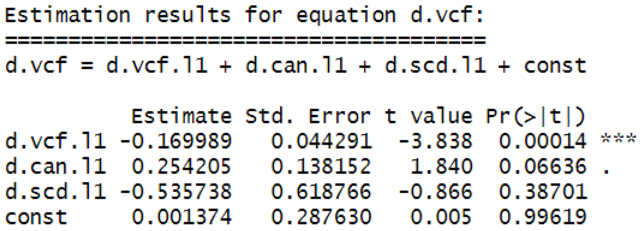
The vector autoregressive model of order 1, denoted as VAR(1), each variable is a linear function of the lag 1 values for all variables in the set.



```
> VARselect(data1)
$selection
AIC(n)  HQ(n)  SC(n) FPE(n)
     1      1      1      1
```

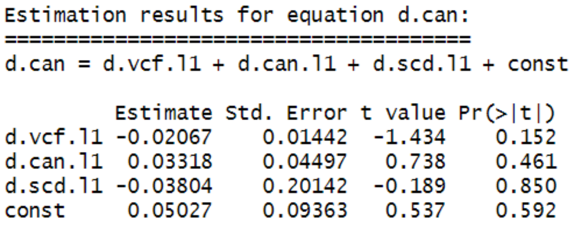
a. Estimate

var1 <- VAR (data1, p =1, type = "const")



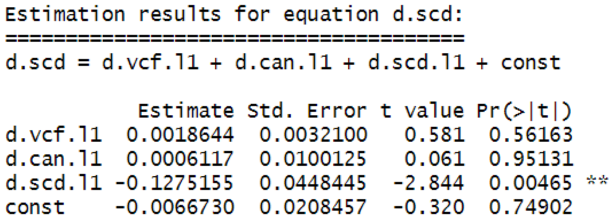
```
Estimation results for equation d.vcf:
======================================
d.vcf = d.vcf.l1 + d.can.l1 + d.scd.l1 + const

          Estimate Std. Error t value Pr(>|t|)    
d.vcf.l1 -0.169989   0.044291  -3.838  0.00014 ***
d.can.l1  0.254205   0.138152   1.840  0.06636 .  
d.scd.l1 -0.535738   0.618766  -0.866  0.38701    
const     0.001374   0.287630   0.005  0.99619    
```

$$\Delta vcf_t = 0.0014 - 0.1699\Delta vcf_{t-1} + 0.2542\Delta can_{t-1} - 0.5357\Delta scd_{t-1} + v_{1t}$$



```
Estimation results for equation d.can:
======================================
d.can = d.vcf.l1 + d.can.l1 + d.scd.l1 + const

         Estimate Std. Error t value Pr(>|t|)
d.vcf.l1 -0.02067    0.01442  -1.434    0.152
d.can.l1  0.03318    0.04497   0.738    0.461
d.scd.l1 -0.03804    0.20142  -0.189    0.850
const     0.05027    0.09363   0.537    0.592
```

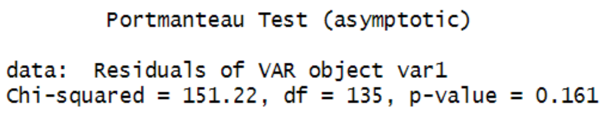
$$\Delta can_t = 0.0503 - 0.0381\Delta scd_{t-1} + 0.0332\Delta can_{t-1} - 0.0207\Delta vcf_{t-1} + v_{2t}$$



```
Estimation results for equation d.scd:
======================================
d.scd = d.vcf.l1 + d.can.l1 + d.scd.l1 + const

           Estimate Std. Error t value Pr(>|t|)   
d.vcf.l1  0.0018644  0.0032100   0.581  0.56163   
d.can.l1  0.0006117  0.0100125   0.061  0.95131   
d.scd.l1 -0.1275155  0.0448445  -2.844  0.00465 **
const    -0.0066730  0.0208457  -0.320  0.74902   
```

$$\Delta scd_t = -0.0067 - 0.1275\Delta scd_{t-1} + 0.0006\Delta can_{t-1} + 0.00187\Delta vcf_{t-1} + v_{3t}$$

b. Serial correlation of Residual test

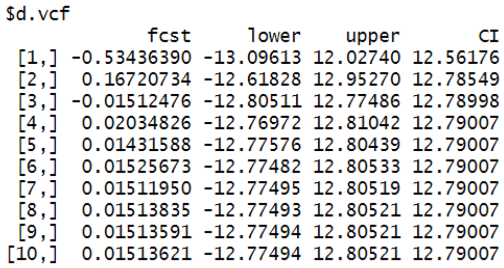


```
        Portmanteau Test (asymptotic)

data:  Residuals of VAR object var1
Chi-squared = 151.22, df = 135, p-value = 0.161
```

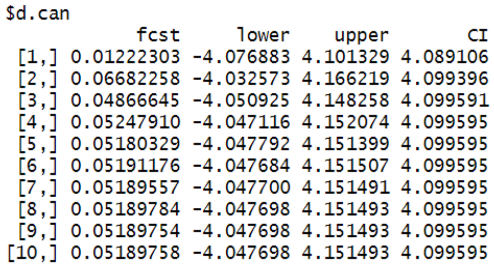
$$\{H_0 : \ No\ serial\ correlation \qquad H_1 : has\ serial\ corelation$$

P_value = 0.161, Null hypothesis is not rejected hence *there is no serial correlation*

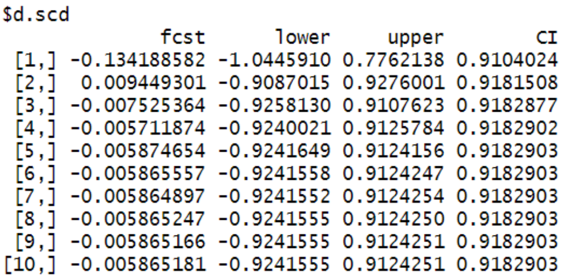
c. Forecast with VAR (1)



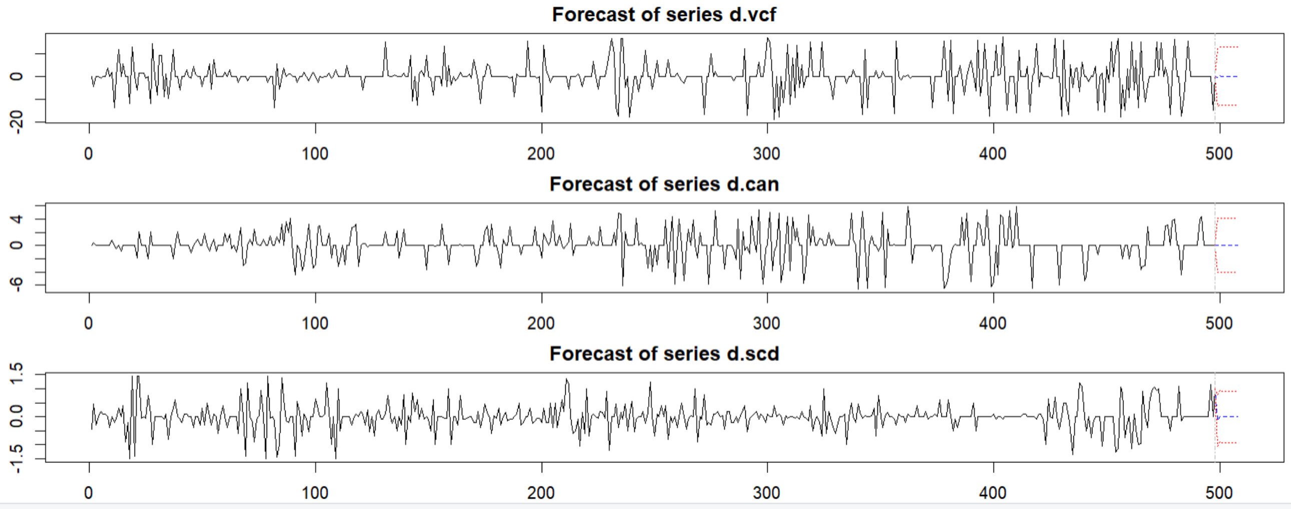
```
$d.vcf
             fcst     lower    upper       CI
 [1,] -0.53436390 -13.09613 12.02740 12.56176
 [2,]  0.16720734 -12.61828 12.95270 12.78549
 [3,] -0.01512476 -12.80511 12.77486 12.78998
 [4,]  0.02034826 -12.76972 12.81042 12.79007
 [5,]  0.01431588 -12.77576 12.80439 12.79007
 [6,]  0.01525673 -12.77482 12.80533 12.79007
 [7,]  0.01511950 -12.77495 12.80519 12.79007
 [8,]  0.01513835 -12.77493 12.80521 12.79007
 [9,]  0.01513591 -12.77494 12.80521 12.79007
[10,]  0.01513621 -12.77494 12.80521 12.79007
```



```
$d.can
            fcst     lower    upper       CI
 [1,] 0.01222303 -4.076883 4.101329 4.089106
 [2,] 0.06682258 -4.032573 4.166219 4.099396
 [3,] 0.04866645 -4.050925 4.148258 4.099591
 [4,] 0.05247910 -4.047116 4.152074 4.099595
 [5,] 0.05180329 -4.047792 4.151399 4.099595
 [6,] 0.05191176 -4.047684 4.151507 4.099595
 [7,] 0.05189557 -4.047700 4.151491 4.099595
 [8,] 0.05189784 -4.047698 4.151493 4.099595
 [9,] 0.05189754 -4.047698 4.151493 4.099595
[10,] 0.05189758 -4.047698 4.151493 4.099595
```

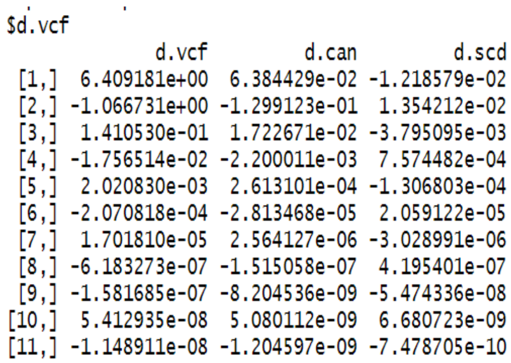


```
$d.scd
              fcst      lower     upper        CI
 [1,] -0.134188582 -1.0445910 0.7762138 0.9104024
 [2,]  0.009449301 -0.9087015 0.9276001 0.9181508
 [3,] -0.007525364 -0.9258130 0.9107623 0.9182877
 [4,] -0.005711874 -0.9240021 0.9125784 0.9182902
 [5,] -0.005874654 -0.9241649 0.9124156 0.9182903
 [6,] -0.005865557 -0.9241558 0.9124247 0.9182903
 [7,] -0.005864897 -0.9241552 0.9124254 0.9182903
 [8,] -0.005865247 -0.9241555 0.9124250 0.9182903
 [9,] -0.005865166 -0.9241555 0.9124251 0.9182903
[10,] -0.005865181 -0.9241555 0.9124251 0.9182903
```

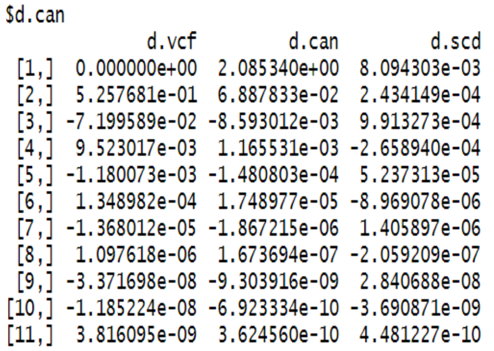


4. Impulse response function

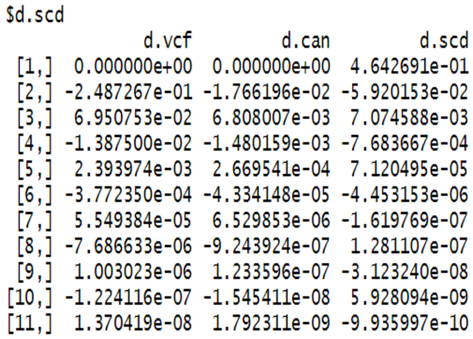
Impulse response coefficients



```
$d.vcf
              d.vcf         d.can         d.scd
 [1,]  6.409181e+00  6.384429e-02 -1.218579e-02
 [2,] -1.066731e+00 -1.299123e-01  1.354212e-02
 [3,]  1.410530e-01  1.722671e-02 -3.795095e-03
 [4,] -1.756514e-02 -2.200011e-03  7.574482e-04
 [5,]  2.020830e-03  2.613101e-04 -1.306803e-04
 [6,] -2.070818e-04 -2.813468e-05  2.059122e-05
 [7,]  1.701810e-05  2.564127e-06 -3.028991e-06
 [8,] -6.183273e-07 -1.515058e-07  4.195401e-07
 [9,] -1.581685e-07 -8.204536e-09 -5.474336e-08
[10,]  5.412935e-08  5.080112e-09  6.680723e-09
[11,] -1.148911e-08 -1.204597e-09 -7.478705e-10
```

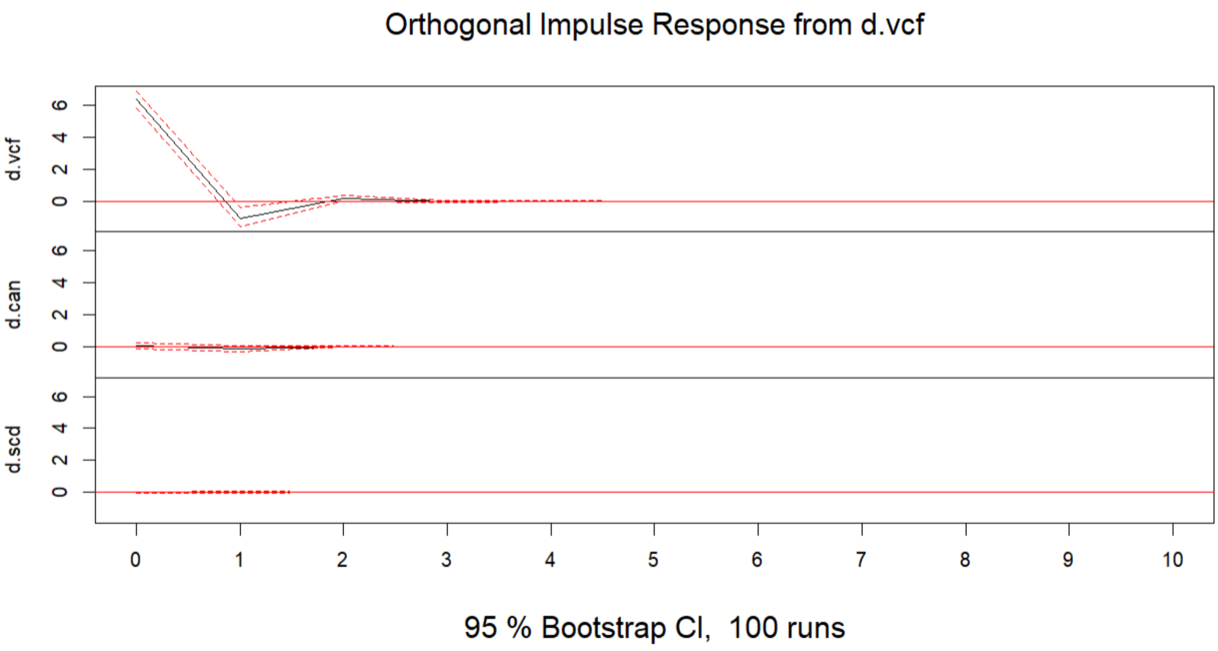


```
$d.can
              d.vcf         d.can         d.scd
 [1,]  0.000000e+00  2.085340e+00  8.094303e-03
 [2,]  5.257681e-01  6.887833e-02  2.434149e-04
 [3,] -7.199589e-02 -8.593012e-03  9.913273e-04
 [4,]  9.523017e-03  1.165531e-03 -2.658940e-04
 [5,] -1.180073e-03 -1.480803e-04  5.237313e-05
 [6,]  1.348982e-04  1.748977e-05 -8.969078e-06
 [7,] -1.368012e-05 -1.867215e-06  1.405897e-06
 [8,]  1.097618e-06  1.673694e-07 -2.059209e-07
 [9,] -3.371698e-08 -9.303916e-09  2.840688e-08
[10,] -1.185224e-08 -6.923334e-10 -3.690871e-09
[11,]  3.816095e-09  3.624560e-10  4.481227e-10
```



```
$d.scd
              d.vcf         d.can         d.scd
 [1,]  0.000000e+00  0.000000e+00  4.642691e-01
 [2,] -2.487267e-01 -1.766196e-02 -5.920153e-02
 [3,]  6.950753e-02  6.808007e-03  7.074588e-03
 [4,] -1.387500e-02 -1.480159e-03 -7.683667e-04
 [5,]  2.393974e-03  2.669541e-04  7.120495e-05
 [6,] -3.772350e-04 -4.334148e-05 -4.453153e-06
 [7,]  5.549384e-05  6.529853e-06 -1.619769e-07
 [8,] -7.686633e-06 -9.243924e-07  1.281107e-07
 [9,]  1.003023e-06  1.233596e-07 -3.123240e-08
[10,] -1.224116e-07 -1.545411e-08  5.928094e-09
[11,]  1.370419e-08  1.792311e-09 -9.935997e-10
```

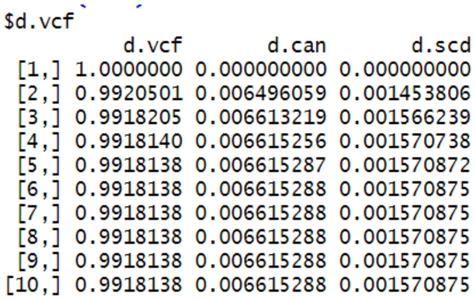
Impulse responses are best represented in graphs showing the responses of a VAR endogenous variable in time.



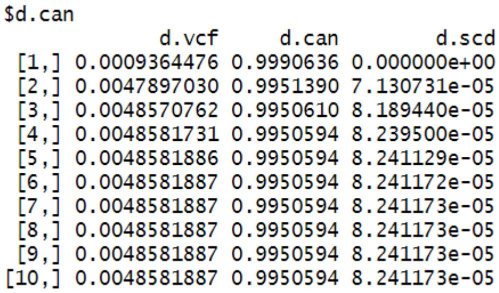
The interpretation here is straightforward: an impulse (shock) to DC at time zero has large effects the next period, but the effects become smaller and smaller as the time passes. The dotted lines show the 95 percent interval estimates of these effects. The VAR function prints the values corresponding to the impulse response graphs.

5. Forecast error variance decomposition – fevd

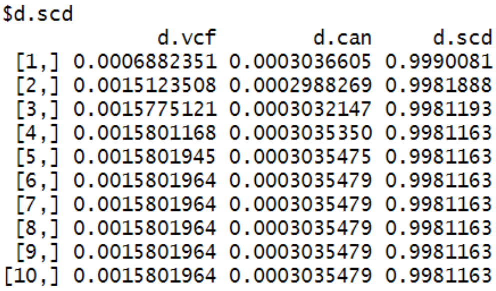
The forecast error variance decomposition is based upon the orthogonalized impulse response coefficient matrices and allow the user to analyze the contribution of variable j to the h-step forecast error variance of variable k. Forecast error variance decomposition (FEVD) is an econometric tool used by many economists in the vector autoregression (VAR) context for assessing the driving forces of business cycles.



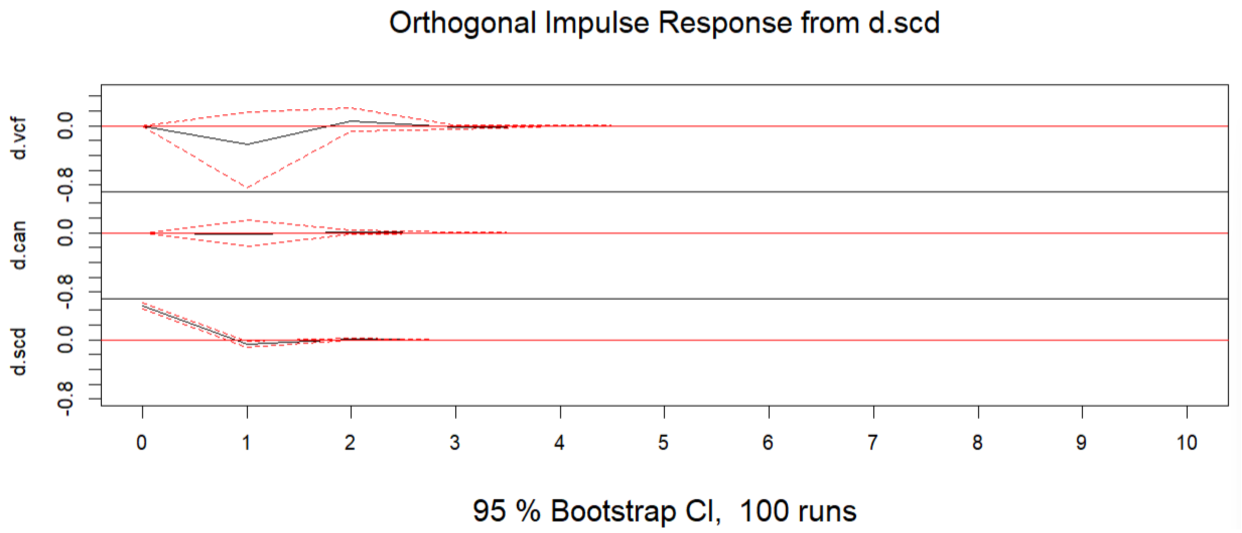
```
$d.vcf
          d.vcf       d.can       d.scd
 [1,] 1.0000000 0.000000000 0.000000000
 [2,] 0.9920501 0.006496059 0.001453806
 [3,] 0.9918205 0.006613219 0.001566239
 [4,] 0.9918140 0.006615256 0.001570738
 [5,] 0.9918138 0.006615287 0.001570872
 [6,] 0.9918138 0.006615288 0.001570875
 [7,] 0.9918138 0.006615288 0.001570875
 [8,] 0.9918138 0.006615288 0.001570875
 [9,] 0.9918138 0.006615288 0.001570875
[10,] 0.9918138 0.006615288 0.001570875
```



```
$d.can
             d.vcf     d.can        d.scd
 [1,] 0.0009364476 0.9990636 0.000000e+00
 [2,] 0.0047897030 0.9951390 7.130731e-05
 [3,] 0.0048570762 0.9950610 8.189440e-05
 [4,] 0.0048581731 0.9950594 8.239500e-05
 [5,] 0.0048581886 0.9950594 8.241129e-05
 [6,] 0.0048581887 0.9950594 8.241172e-05
 [7,] 0.0048581887 0.9950594 8.241173e-05
 [8,] 0.0048581887 0.9950594 8.241173e-05
 [9,] 0.0048581887 0.9950594 8.241173e-05
[10,] 0.0048581887 0.9950594 8.241173e-05
```



```
$d.scd
             d.vcf        d.can     d.scd
 [1,] 0.0006882351 0.0003036605 0.9990081
 [2,] 0.0015123508 0.0002988269 0.9981888
 [3,] 0.0015775121 0.0003032147 0.9981193
 [4,] 0.0015801168 0.0003035350 0.9981163
 [5,] 0.0015801945 0.0003035475 0.9981163
 [6,] 0.0015801964 0.0003035479 0.9981163
 [7,] 0.0015801964 0.0003035479 0.9981163
 [8,] 0.0015801964 0.0003035479 0.9981163
 [9,] 0.0015801964 0.0003035479 0.9981163
[10,] 0.0015801964 0.0003035479 0.9981163
```



CONCLUSION

Time series are constructed using data measured over time at evenly spaced intervals. This paper mainly introduces the basic concepts of time series and time series processing models. The main method for forecasting that time is the ARIMA model. However, the ARIMA model is only suitable for stationary and linear time series, so time series with fast variation or short historical data series give inaccurate results. The time series in economics, due to the characteristics of economic development, depends a lot on different factors, so it has many variations and is nonlinear. Therefore, the ARIMA model cannot handle well in the economic field. In addition, in our report, we also use both cointegration and VAR models to test whether the series are correlated with each other and to predict the factors in the past that will affect the present and the future.