Recommendation of an Investment of a Stock by Analyzing Historical Data

Prepared by Tee Tian En

Table of Contents

1.0 Introduction

In this assignment, we want to recommend an investment of a stock by analyzing historical data. We first set up for the working directory and get all needed library in R.

```
# Setting my working directory
setwd("C:/Users/ASUS/OneDrive/Documents/Personal/private/GitHub")

# Ready library
library(fpp3)
```

2.0 Data Set Chosen

Yahoo Finance is a website that provides information of stocks. We have selected a data set from the website to carry out the analysis.

In this assignment, we use the stock prices of **Natural Grocers by Vitamin Cottage, Inc.** from 1 March 2023 to 30 September 2023.

Natural Grocers by Vitamin Cottage, Inc. has been categories in consumer defensive sector and grocery stores industry. The company was founded in 1955 and is headquartered in Lakewood, Colorado. It offer grocery products, paper products, cleaning products, dairy products, beverages, and other supplements.

We download the data set from Yahoo Finance, and save it in a csv file. We also add the counting day using trading days (Day).

The data is saved into three sets:

- NGVC: the seven-months time-frame data (from 1 March 2023 to 30 September 2023), used in main analysis;

- NGVC_historical: the 6 years historical data (from 1 January 2018 to 30 September 2023), used in detecting seasonality of the investment;

- NGVC_oct: the data covered until 14 consecutive trading days (from 1 March 2023 to 19 October 2023), used in forecasting and the choice of best forecasting method.

```r
# Read the data into R
# Save it as tsibble object
NGVC_data <- read.csv("NGVC.csv")

# Data from 1 March 2023 to 30 September 2023
NGVC <- NGVC_data |>
  filter(Date >= "2023-03-01" & Date <= "2023-09-30") |>
  mutate(Day = row_number()) |>
  mutate(Date = as_date(Date)) |>
  as_tsibble(index = Day)

# Data from 1 January 2018 to 30 September 2023
NGVC_historical <- NGVC_data |>
  filter(Date >="2018-01-01" & Date <= "2023-09-30") |>
  mutate(Day = row_number()) |>
  mutate(Date = as_date(Date)) |>
  as_tsibble(index = Day)

# Data from 1 March 2023 to 19 October 2023
NGVC_oct <- NGVC_data |>
  filter(Date >= "2023-03-01" & Date <= "2023-10-19") |>
  mutate(Day = row_number()) |>
  mutate(Date = as_date(Date)) |>
  as_tsibble(index = Day)

# Check the data
NGVC
```

```
## # A tsibble: 148 x 8 [1]
##    Date        Open  High   Low Close Adj.Close Volume   Day
##    <date>     <dbl> <dbl> <dbl> <dbl>     <dbl>  <int> <int>
##  1 2023-03-01  10.9  11.0  10.7  10.8      10.7 121700     1
##  2 2023-03-02  10.9  11.0  10.8  11        10.8  93900     2
##  3 2023-03-03  11.1  11.2  10.9  11.2      11.0  91700     3
##  4 2023-03-06  11.2  11.2  10.8  10.9      10.7 151800     4
##  5 2023-03-07  10.9  10.9  10.7  10.9      10.7  59400     5
##  6 2023-03-08  10.8  11.2  10.8  11.1      11.0  43600     6
##  7 2023-03-09  11.2  11.2  10.9  11        10.8  77600     7
##  8 2023-03-10  11.0  11.1  10.8  11        10.8  51600     8
##  9 2023-03-13  10.8  10.9  10.4  10.4      10.3  81000     9
## 10 2023-03-14  10.6  10.8  10.4  10.5      10.3  93600    10
## # ℹ 138 more rows
```

```
NGVC_historical
```

```
## # A tsibble: 1,446 x 8 [1]
##    Date        Open  High   Low Close Adj.Close Volume   Day
##    <date>     <dbl> <dbl> <dbl> <dbl>     <dbl>  <int> <int>
##  1 2018-01-02  8.99  9.04  8.86  8.91      7.01 126400     1
##  2 2018-01-03  8.91  9     8.85  8.87      6.98  66100     2
##  3 2018-01-04  8.87  8.9   8.34  8.67      6.83 108000     3
##  4 2018-01-05  8.63  8.85  8.48  8.51      6.70 118800     4
##  5 2018-01-08  8.47  8.5   8.15  8.42      6.63 159100     5
##  6 2018-01-09  8.46  8.5   8.26  8.3       6.53  65500     6
##  7 2018-01-10  8.21  8.22  7.92  8.15      6.42 113700     7
##  8 2018-01-11  8.15  8.44  8.05  8.42      6.63  51700     8
##  9 2018-01-12  8.41  8.62  8.28  8.33      6.56  59300     9
## 10 2018-01-16  8.44  8.72  8.29  8.56      6.74 466700    10
## # ℹ 1,436 more rows
```

```
NGVC_oct
```

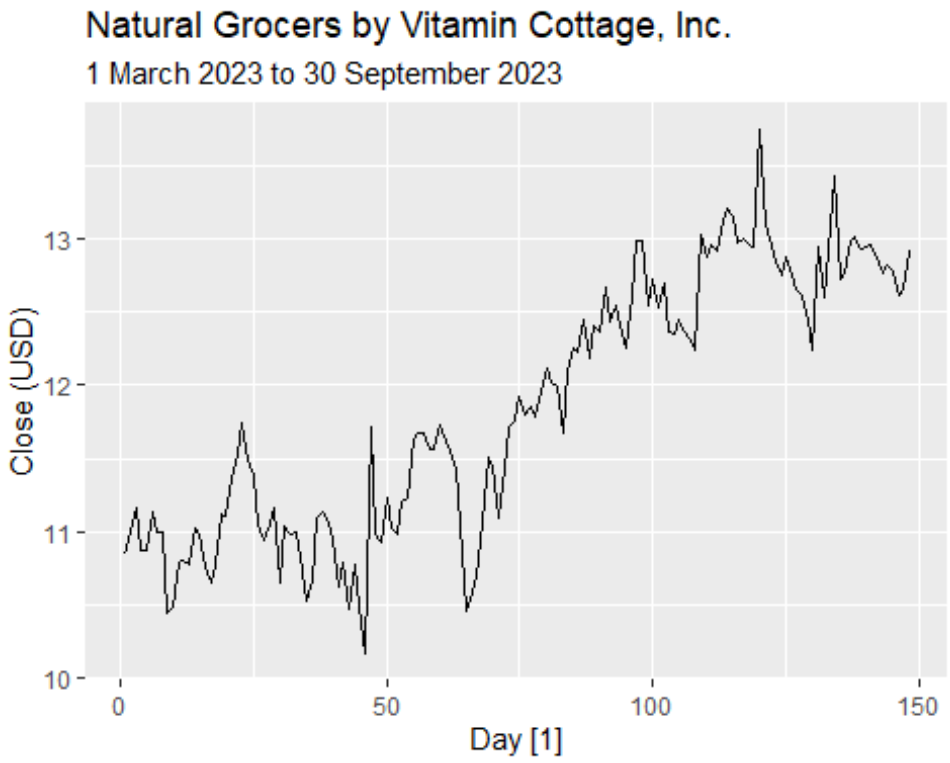
```
## # A tsibble: 162 x 8 [1]
##    Date        Open  High   Low Close Adj.Close Volume   Day
##    <date>     <dbl> <dbl> <dbl> <dbl>     <dbl>  <int> <int>
##  1 2023-03-01  10.9  11.0  10.7  10.8      10.7 121700     1
##  2 2023-03-02  10.9  11.0  10.8  11        10.8  93900     2
##  3 2023-03-03  11.1  11.2  10.9  11.2      11.0  91700     3
##  4 2023-03-06  11.2  11.2  10.8  10.9      10.7 151800     4
##  5 2023-03-07  10.9  10.9  10.7  10.9      10.7  59400     5
##  6 2023-03-08  10.8  11.2  10.8  11.1      11.0  43600     6
##  7 2023-03-09  11.2  11.2  10.9  11        10.8  77600     7
##  8 2023-03-10  11.0  11.1  10.8  11        10.8  51600     8
##  9 2023-03-13  10.8  10.9  10.4  10.4      10.3  81000     9
## 10 2023-03-14  10.6  10.8  10.4  10.5      10.3  93600    10
## # ℹ 152 more rows
```

3.0 Analysis

3.1 Seven-month Time-frame Data

a) Provide relevant time series plots illustrating the chosen investment and highlight any notable fluctuations in price movement. Additionally, offer an explanation for the peaks and troughs observed over the selected seven-month time-frame.

```
# Produce time series plots over the selected seven-month time-frame
NGVC |>
  autoplot(Close) +
  labs(title = "Natural Grocers by Vitamin Cottage, Inc.",
       subtitle = "1 March 2023 to 30 September 2023",
       y = "Close (USD)")
```



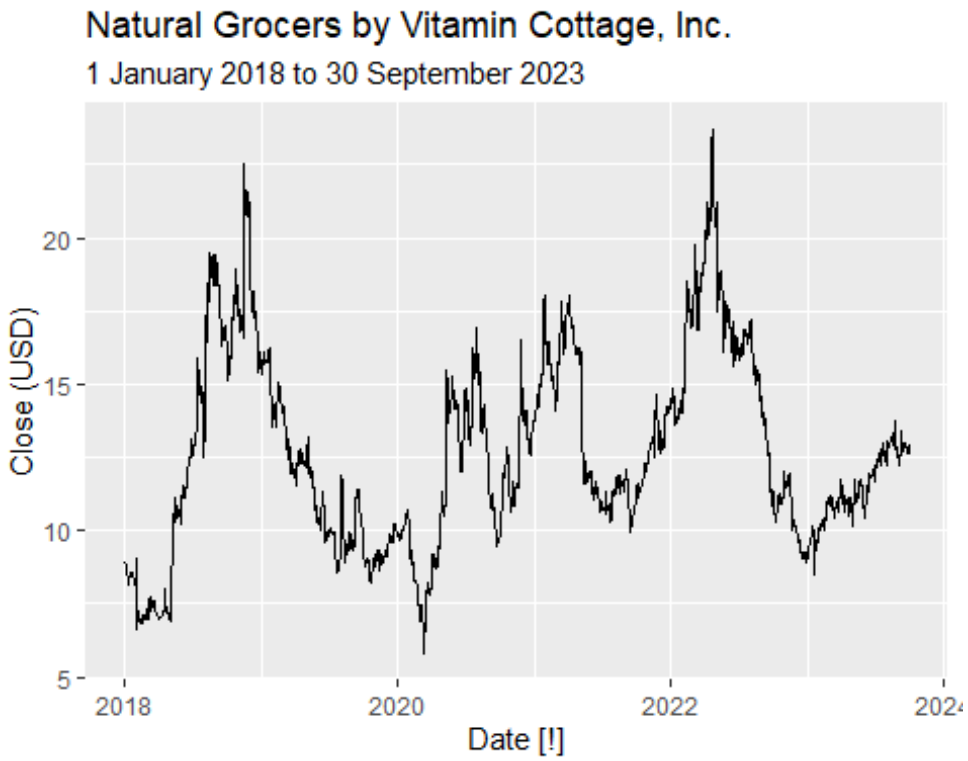
From the graph above, we can observe that there is an increasing trend throughout the seven-month time-frame. However, the seven-period time-frame does not give a good insight on seasonality or cycle.

The closing price was fluctuated around USD 11 from March 2023 until May 2023, and it started increased to USD 13 from June 2023 to July 2023. The price then fluctuated around USD 13 from August 2023 to September 2023. There was a period of reduced price at the beginning and the end of May 2023.

3.2 Historical Data

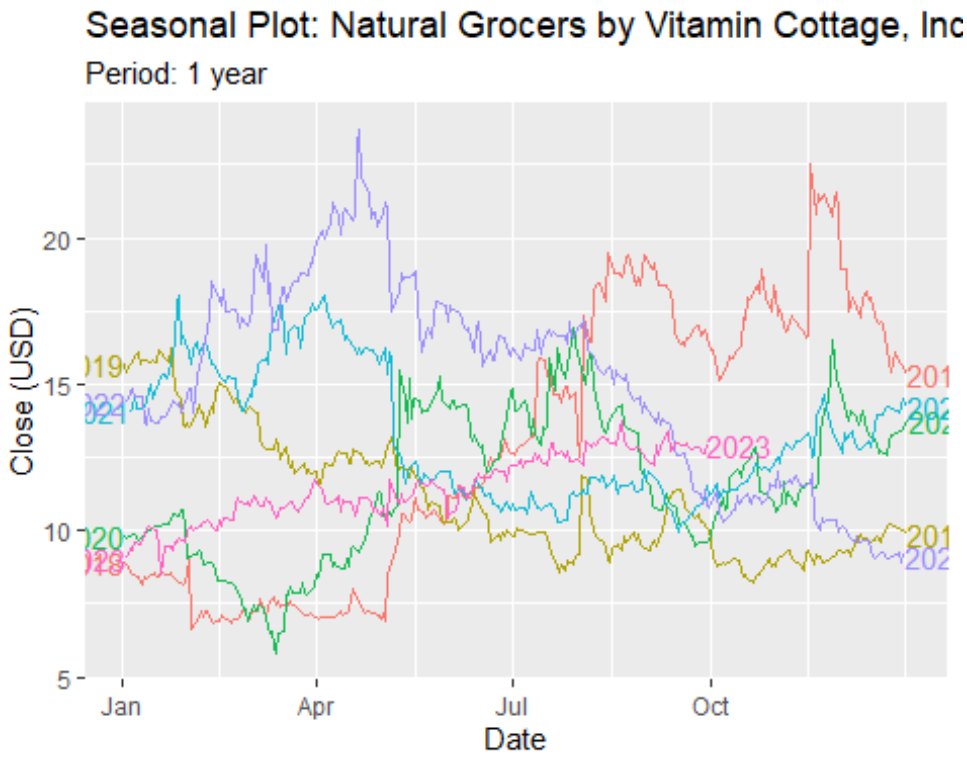
b) Analyze historical data from previous years, covering an appropriate time-frame, to detect any patterns of seasonality influencing the pricing of the selected investment, with a specific emphasis on the stock index where the investment is listed. Does this analysis provide insights for investors regarding optimal timing for buying and selling decisions?

```
# Plot of the historical data
NGVC_historical |>
  update_tsibble(index = Date, regular = FALSE) |>
  autoplot(Close) +
  labs(title = "Natural Grocers by Vitamin Cottage, Inc.",
       subtitle = "1 January 2018 to 30 September 2023",
       y = "Close (USD)")
```

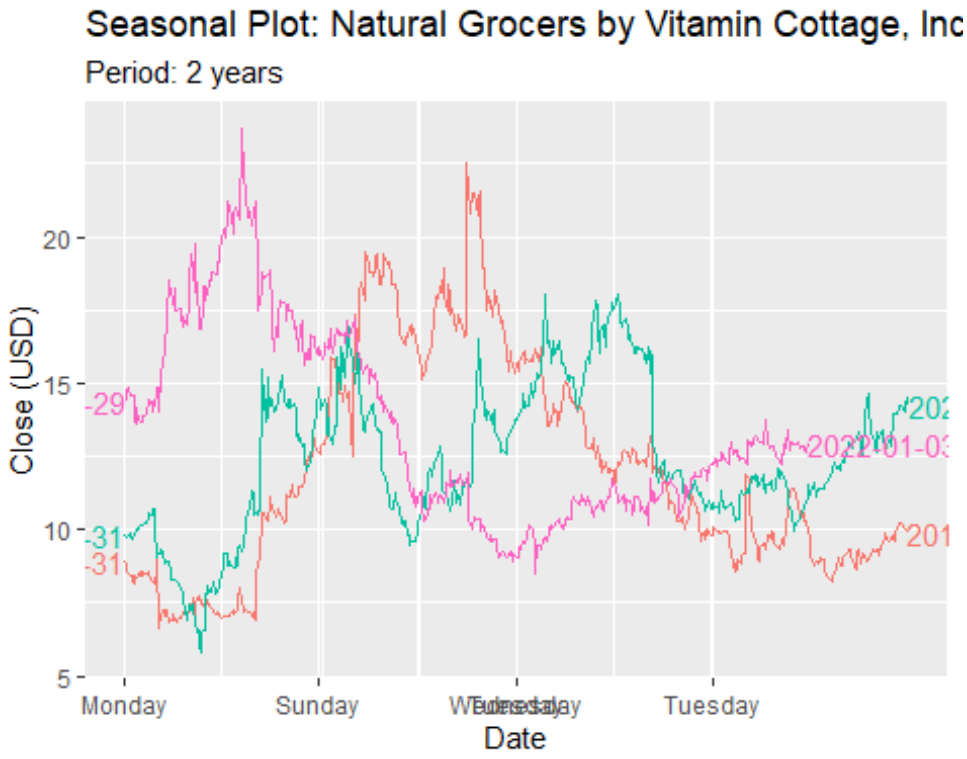


The graph above shows some peaks around the end of 2018, mid-year of 2020, the start of 2021, and mid-year of 2022. The need in groceries during the COVID-19 pandemic may be a reason of the rise in 2020 and 2021. Post-pandemic period is also a point for the increase in the year 2022. A sudden fall at the start of 2020 may cause by the COVID-19 pandemic. In the mid-year of 2021, the sudden drop occurs. The large decrease in 2022 may cause by the start of Ukraine war.

```r
# Detect patterns of seasonality
NGVC_historical |>
  update_tsibble(index = Date, regular = FALSE) |>
  gg_season(Close, labels = "both") +
  labs(title = "Seasonal Plot: Natural Grocers by Vitamin Cottage, Inc.",
       subtitle = "Period: 1 year",
       y = "Close (USD)")
```

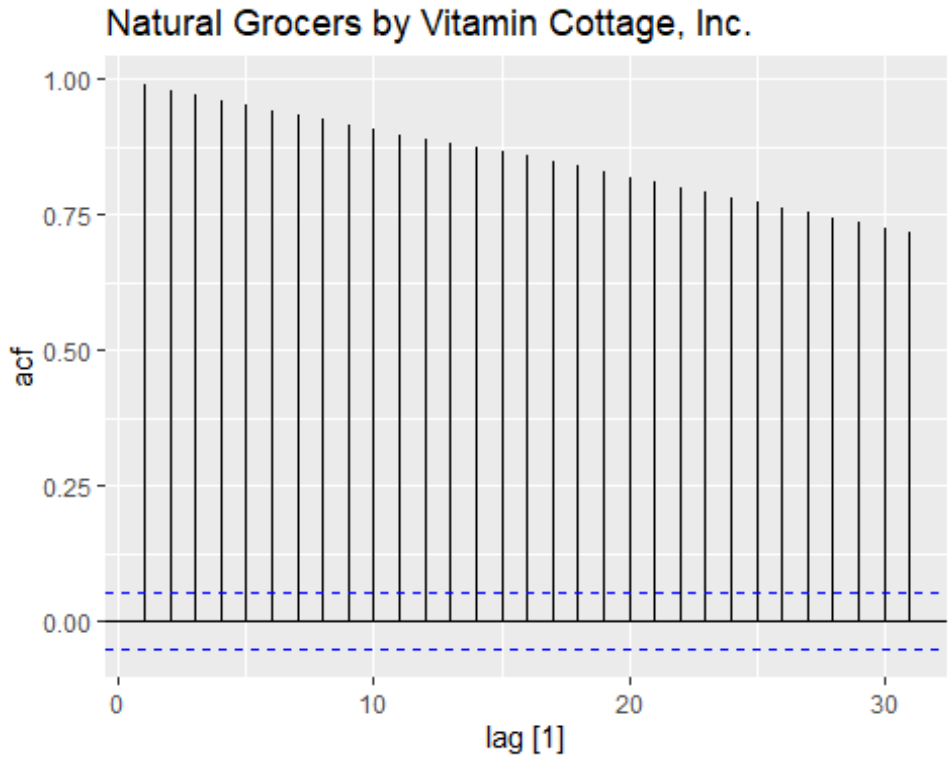


```
NGVC_historical |>
  update_tsibble(index = Date, regular = FALSE) |>
  gg_season(Close, labels = "both", period = "2 year") +
  labs(title = "Seasonal Plot: Natural Grocers by Vitamin Cottage, Inc.",
       subtitle = "Period: 2 years",
       y = "Close (USD)")
```



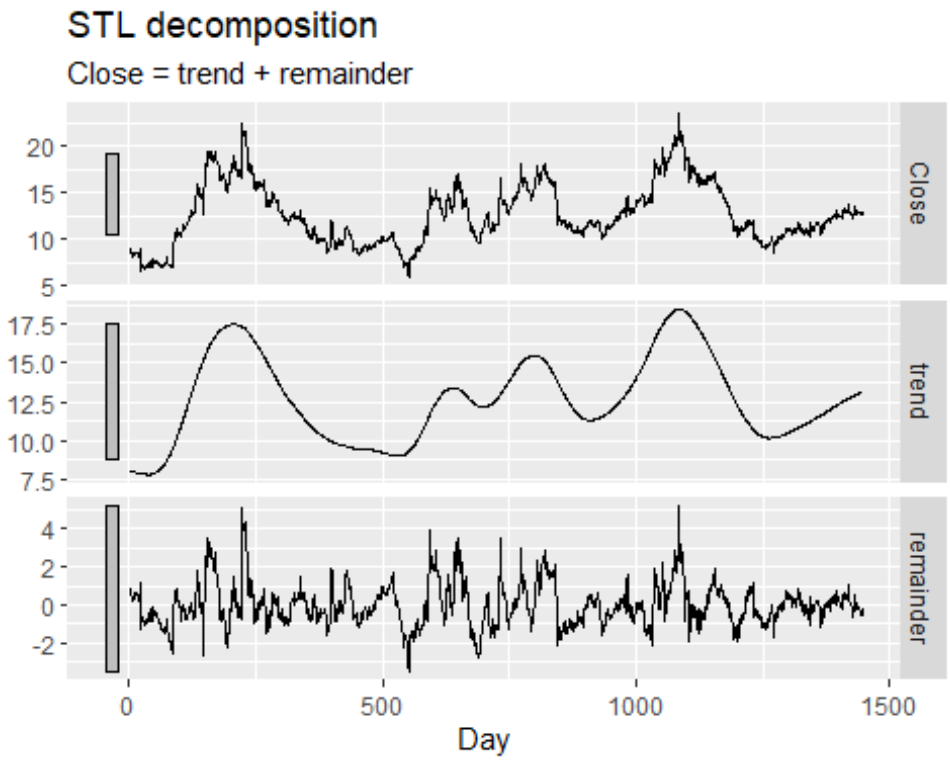
From the seasonal plot, we do not detect any seasonality neither using a period of 1 year nor 2 years. There are peaks and troughs in the graph, but they are not in a pattern of season. We continue to produce ACF plot and STL decomposition plot.

```
NGVC_historical |>
  ACF(Close) |>
  autoplot() +
  labs(title = "Natural Grocers by Vitamin Cottage, Inc.")
```



The slow decrease in the ACF plot as the lags increase is due to the trend, but there is no pattern that can be observed and thus seasonality is not shown.

```
dcmp <- NGVC_historical |>
  model(stl = STL(Close))
components(dcmp) |>
  autoplot()
```



The STL decomposition plot does not show the seasonal plot, suggests that there is no seasonality. The trend plot suggests a slightly increase trend, with a possibility of cycle occurring.

Next, we specifically emphasis on stock index.

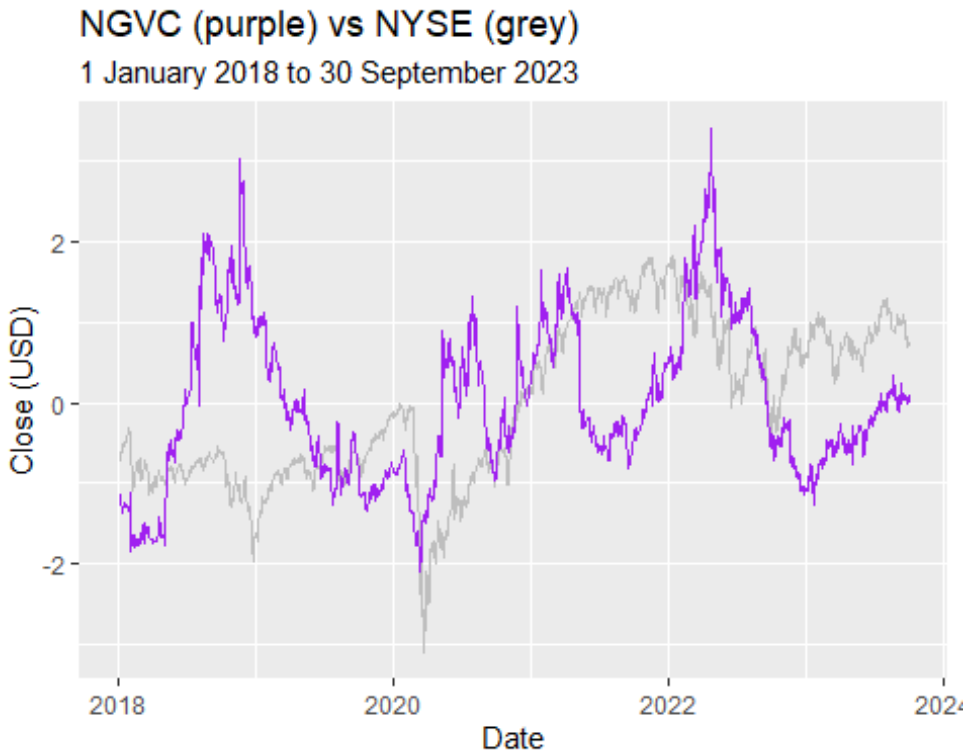
```
# stock index - NYSE
# Read in NYSE data
NYSE_data <- read.csv("^NYA.csv")

# Data from 1 January 2018 to 30 September 2023
NYSE <- NYSE_data |>
  filter(Date >="2018-01-01" & Date <= "2023-09-30") |>
  mutate(Day = row_number()) |>
  mutate(Close = as.numeric(Close)) |>
  mutate(Date = as_date(Date)) |>
  as_tsibble(index = Day)

# Standardization
NGVC_historical <- NGVC_historical |>
  mutate(norm_Close = (NGVC_historical$Close -
mean(NGVC_historical$Close))/sd(NGVC_historical$Close))
```

```r
NYSE <- NYSE |>
  mutate(norm_Close = (NYSE$Close - mean(NYSE$Close))/sd(NYSE$Close))

# Plot of NYSE and NGVC in 6 years period
ggplot() +
  geom_line(NYSE, mapping = aes(x = Date, y = norm_Close), color = "grey") +
  geom_line(NGVC_historical, mapping = aes(x = Date, y = norm_Close), color =
"purple") +
  labs(title = "NGVC (purple) vs NYSE (grey)",
       subtitle = "1 January 2018 to 30 September 2023",
       y = "Close (USD)")
```



From the stock index, we observed that there is a large drop in 2020, but it shows an increasing trend after that. NGVC behaves similarly to the NYSE during the mid-year of 2019 until the mid-year of 2021. It also shows similarity in 2022 and 2023, where they increase and decrease at a similar time. Overall, by observing the graph above, NGVC has a good performance as it is higher the than stock index price most of the time and suggests that an investment is worth.

However, there is no seasonality shown in the NGVC price, suggests that it does not provide insights for investor regarding optimal timing for buying and selling decisions. The increasing trend suggests the price will keep increasing for at least a period of time.

3.3 Training and Test Sets

c) (i) Split the data into training and test sets using an appropriate ratio to identify the best basic forecasting method. Justify the choice of the forecasting method.

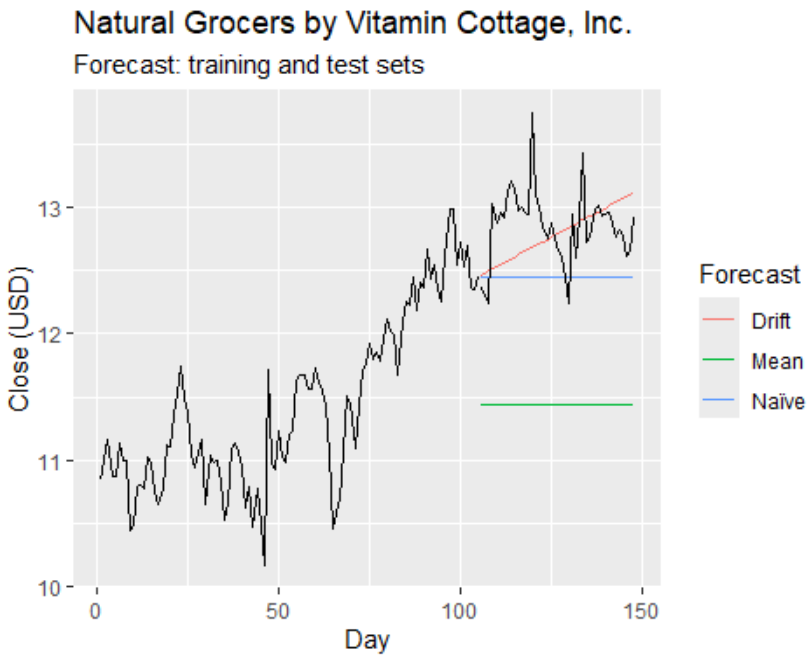
```
# Split data into training and test sets
# Training set
NGVC_training <- NGVC |>
  filter(Date < "2023-08-01")

# Test set
NGVC_test <- NGVC |>
  filter(Date >= "2023-08-01")

# Basic forecasting method
NGVC_fit <- NGVC_training |>
  model(
    Mean = MEAN(Close),
    `Naïve` = NAIVE(Close),
    Drift = RW(Close ~ drift())
  )

NGVC_fc <- NGVC_fit |>
  forecast(new_data = NGVC_test)
```

```
NGVC_fc |>
  autoplot(NGVC_training, level = NULL) +
  autolayer(NGVC_test, Close, colour = "black") +
  guides(colour = guide_legend(title = "Forecast"))+
  labs(title = "Natural Grocers by Vitamin Cottage, Inc.",
       subtitle = "Forecast: training and test sets",
       y = "Close (USD)")
```



We then split the seven-month time frame data into training and test sets to identity the best forecasting method. The ratio of the split data is 5:2 months, which are approximately 70% in training set and 30% in test set. This is a appropriate ratio for this data sets as there covered most of the data in training set.

We use the basic forecasting methods to continue, that is mean method, naïve method and drift method. The seasonal naïve method is not consider because we cannot observed any seasonality in this data set. The models are fitted into the training set and forecast had been done and compared with the actual values in the test set.

In this case, the mean forecasts do not close to the actual values, suggests that mean method is not suitable. The drift forecasts are more closer to the actual values compared to the naïve forecasts, suggests that the drift method might be the best forecasting method in this case. We will choose the drift method as our choice of the best forecasting method in the following sessions.

Due to the lack of basic forecast, the drift method may give a large prediction interval as it may not capture enough evidences for the forecast. We need to look into it and consider other forecasting methods in this instance.

c) (ii) Based on the findings from part (c)(i), forecast the stock price for 14 consecutive trading days starting from 2 October 2023. Present both the actual and forecasted values on the same time series plot.

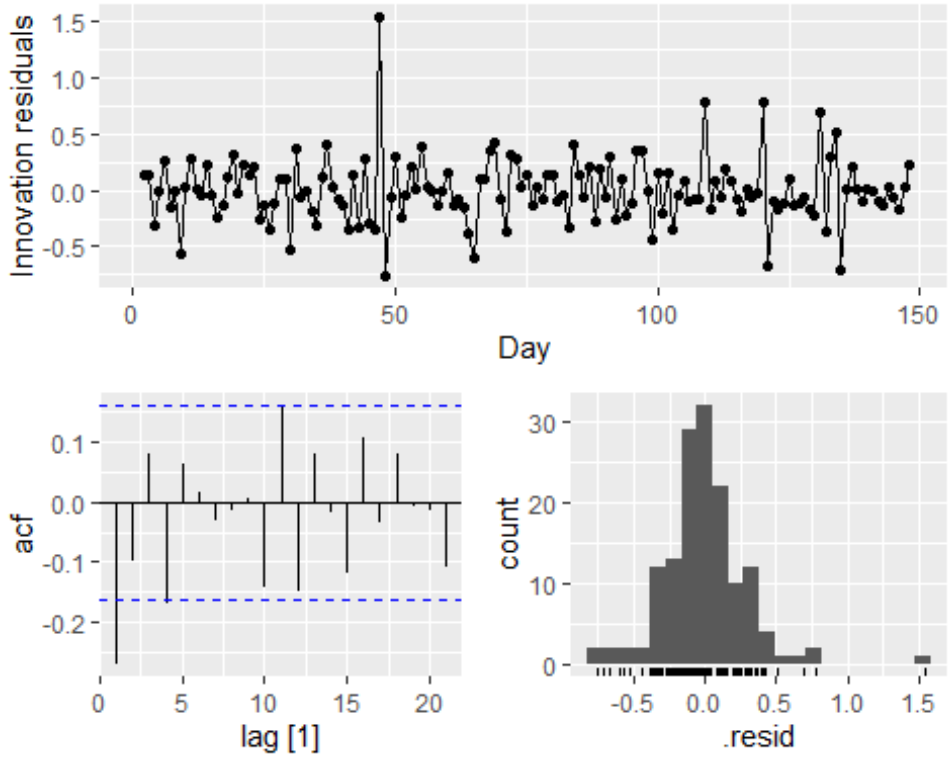
We consider the **drift** method and forecast the stock price for 14 consecutive trading days.

First, we fit the model again using the seven-month time-frame data and check the residuals.

```
# Best forecast method - Drift

# Forecast for 14 consecutive trading days
NGVC_fit14 <- NGVC |>
  model(RW(Close ~ drift()))

# Check the residuals
NGVC_fit14 |>
  gg_tsresiduals()
```

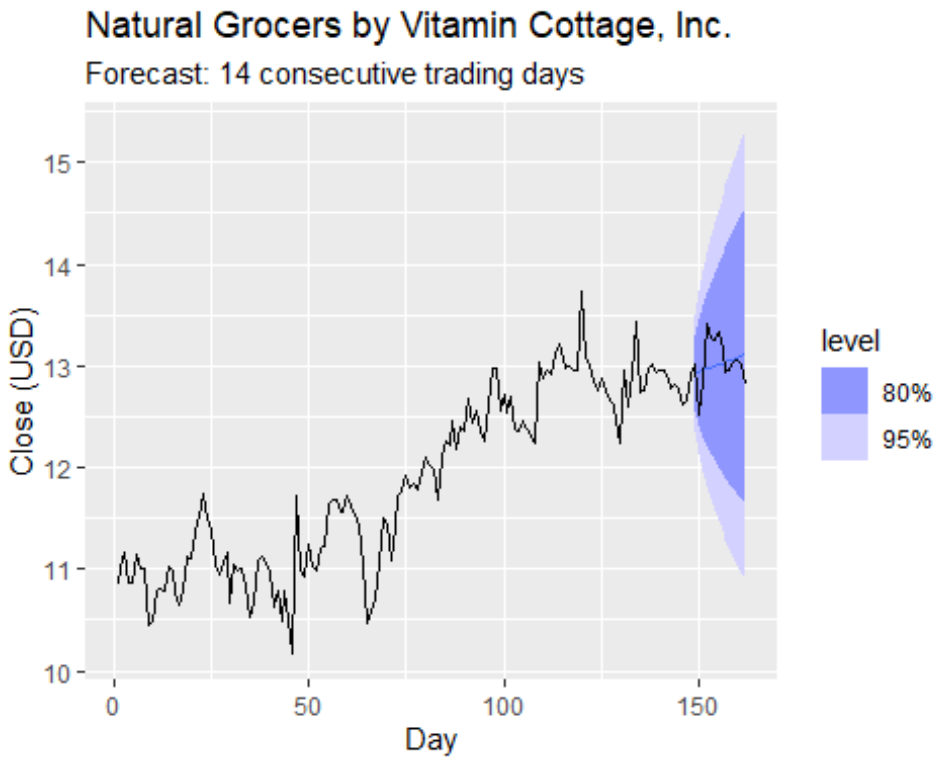


From the residuals plot, we can see that most of the residuals are around 0, except of some points, especially point 47 which spikes out and suggests that it may be an outlier. The acf plot looks good, except for the lag 1 which is significantly different from 0. It suggests that the residuals may be uncorrelated. The histogram shows a long right tail, suggests that the residuals may not be normal. However, more analysis need to consider as there might be a presence of outlier and affects the residuals.

We then continue to forecast the stock price as follows.

```
# Forecast for 14 consecutive trading days (Cont.)
NGVC_fc14 <- NGVC_fit14 |>
  forecast(h = 14)

# Plot of both actual and forecasted values
NGVC_fc14 |>
  autoplot(NGVC_oct) +
  autolayer(NGVC, Close, colour = "black") +
  labs(title = "Natural Grocers by Vitamin Cottage, Inc.",
       subtitle = "Forecast: 14 consecutive trading days",
       y = "Close (USD)")
```



Using the drift method, we then forecast for 14 consecutive trading days. The graph of forecasts and actual values is plotted above. The large interval of prediction does suggest that it may not capture the exact pattern of the data set. However, the forecasts seems to be close to the actual values.

3.4 Forecasted Values

d) Explain the underlying reasons for the similarities or differences between the actual and forecasted values.

The forecasts only capture the trend of the data set, but not the pattern. The actual values fluctuated around the forecasts, between the prediction interval. Due to the changes in the underlying patterns of the data set that were not adequately captured, there is differences between the actual and forecasted values.

We can study the forecasting more by looking into the accuracy.

```
# Accuracy
accuracy(NGVC_fit14)
```

```
## # A tibble: 1 × 10
##   .model            .type        ME  RMSE   MAE     MPE  MAPE  MASE RMSSE
ACF1
##   <chr>             <chr>     <dbl> <dbl> <dbl>   <dbl> <dbl> <dbl> <dbl>
<dbl>
## 1 RW(Close ~ drift… Trai… -2.78e-16 0.287 0.203 -0.0307  1.73  1.01 0.999
-0.269
```

```
accuracy(NGVC_fc14, NGVC_oct)
```

```
## # A tibble: 1 × 10
##   .model              .type     ME  RMSE   MAE   MPE  MAPE  MASE RMSSE
ACF1
##   <chr>               <chr>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
<dbl>
## 1 RW(Close ~ drift()) Test  0.0301 0.238 0.195 0.200  1.49 0.965 0.827
0.388
```

The accuracy of both fitted and forecasted values are close. There is a noticeable difference in the ME. However, the overall accuracy is close, suggests that the drift method may be used in forecasting.

4.0 Conclusion

e) Conclude by stating your future performance expectations for this stock and the reasons behind your recommendation.

NGVC is a potential stock which it is worth invest in the following period. The best method of our choice is the drift method with positive slope, suggests that the future performance for this stock is expected to be increase.

From the historical data, we observe that there might be a cycle, and now is just the start of a new cycle. This will be another reason why this stock is recommended to buy.

Furthermore, after the COVID-19 pandemic, fears of recession, Russia-Ukraine war and many other factors, the need of daily products has increased as people need these products to survive. Thus, the stock of NGVC is expected to perform better in the future.

In short, NGVC is worth to invest and it is expected to have a better performance in the future.

5.0 References

Natural Grocers by Vitamin Cottage, Inc. (NGVC). Retrieved from https://finance.yahoo.com/quote/NGVC?p=NGVC&.tsrc=fin-srch

NYSE COMPOSITE (DJ) (^NYA). Retrieved from https://finance.yahoo.com/quote/%5ENYA?p=^NYA&.tsrc=fin-srch

Hyndman, R.J., & Athanasopoulos, G. (2021) Forecasting: principles and practice, 3rd edition, OTexts: Melbourne, Australia. Retrieved from https://otexts.com/fpp3/