INDIVIDUAL ASSESSMENT COVER SHEET
Faculty of Design and Creative Technologies

AUT
TE WĀNANGA ARONUI O TĀMAKI MAKAU RAU

First Name	Nikisha	**Family Name**	Chhima	**Student ID No**	22168889
Paper Name	Forecasting	**Paper Code:**	STAT603	**Assignment Due Date**	04/09/2024
Lecturer:	Dr Victor Miranda	**Tutorial Day**	Friday	**Date Submitted**	03/09/2024
Tutor:	Dr Victor Miranda	**Tutorial Time**	12pm-12pm	**No.Words/Pages**	19 pages

In order to ensure fair and honest assessment results for all students, it is a requirement that the work that you hand in for assessment is your own work. If you are uncertain about any of these matters then please discuss them with your lecturer.

Plagiarism and Dishonesty are methods of cheating for the purposes of General Academic Regulations (GAR) http://www.aut.ac.nz/calendar

Assignments will not be accepted if this section is not completed and signed.

Please read the following and **tick** ✓ to indicate your understanding:

1. I understand it is my responsibility to keep a copy of my assignment. ☑ **Yes** ☐ **No**
2. I have signed and read the **Student's Statement below**. ☑ **Yes** ☐ **No**
3. I understand that a software programme (Turnitin) that detects plagiarism and copying may be used on my assignment. ☑ **Yes** ☐ **No**

Student's Statement:

This assessment is entirely my own work and has not been submitted in any other course of study. I have submitted a copy of this assessment to Turnitin, if required.
In this assessment I have acknowledged, to the best of my ability:

- The source of direct quotes from the work of others.
- The ideas of others (includes work from private or professional services, past assessments, other students, books, journals, cut/paste from internet sites and/or other materials).
- The source of diagrams or visual images.

Student's Signature: Nikisha **Date: 03/09/2024**

The information on this form is collected for the primary purpose of submitting your assignment for assessment. Other purposes of collection include receiving your acknowledgement of plagiarism polices and attending to administrative matters. If you choose not to complete all questions on this form, it may not be possible for the Faculty of Design and Creative Technologies to accept your assignment.

Assignment 1 STAT603 - Answers

3/09/2024

Nikisha Chhima, Student ID: 22168889

{NOTE:} Write down your solutions / answers into this Rmd file. Knit the file to PDF and submit.

Total Possible Marks:

90 marks, contributing 25% towards your final grade.

Due date

23:59hr, 4 September 2024.

Page limit

Max 20 pages including `R` code, plots, and your answers.

Data:

- Quarterly unemployed population (thousands) in New Zealand, from Quarter 1, 2001, to Quarter 4, 2022. (Filename: `NZ_Unemployment_Quarterly.xlsx`)
- Quarterly inflation (percent) in New Zealand from Quarter 1, 2001 to Quarter 4, 2022 (Filename: `NZ_Inflation_Quarterly.xlsx`)
- Quarterly real national disposable income (Billion NZ dollars) in New Zealand from Quarter 1, 2010 to Quarter 3, 2022 (Filename: `NZ_DispIncome_Quarterly.xlsx`)

QUESTION 1.

(a) Extract the data from the `New South Wales` state. Aggregate the variable `turnover` (sum over `Turnover`) using `summarise()` and briefly discuss the observed frequencies {(5 marks)}.

Answer:

```
#Importing library
library("fpp3")
```

```
## Registered S3 method overwritten by 'tsibble':
##   method               from
##   as_tibble.grouped_df dplyr
```

```
## -- Attaching packages --------------------------------------------- fpp3 1.0.0 --

## v tibble      3.2.1     v tsibble     1.1.5
## v dplyr       1.1.4     v tsibbledata 0.4.1
## v tidyr       1.3.1     v feasts      0.3.2
## v lubridate   1.9.3     v fable       0.3.4
## v ggplot2     3.5.1     v fabletools  0.4.2

## -- Conflicts ------------------------------------------------- fpp3_conflicts --
## x lubridate::date()    masks base::date()
## x dplyr::filter()      masks stats::filter()
## x tsibble::intersect() masks base::intersect()
## x tsibble::interval()  masks lubridate::interval()
## x dplyr::lag()         masks stats::lag()
## x tsibble::setdiff()   masks base::setdiff()
## x tsibble::union()     masks base::union()
```

```r
#Extracting data from New South Wales state
NSW_turnover <- aus_retail %>%
  filter(State == "New South Wales") %>%
  summarise(Turnover = sum(Turnover))
NSW_turnover
```

```
## # A tsibble: 441 x 2 [1M]
##       Month Turnover
##       <mth>    <dbl>
##  1 1982 Apr    2322
##  2 1982 May    2397.
##  3 1982 Jun    2292.
##  4 1982 Jul    2357.
##  5 1982 Aug    2266.
##  6 1982 Sep    2308.
##  7 1982 Oct    2354.
##  8 1982 Nov    2538.
##  9 1982 Dec    3430
## 10 1983 Jan    2294.
## # i 431 more rows
```

I can see that in 1982 we only have records from month April to December. However, in the following years up to 2018 we have months from January to December. There is a steady increase in turnover in between the years, with a significant increase which takes place in the month of December each year. This increase could be due to upcoming major holidays such as Christmas. While families and children may have time off work and school, they often spend more money on gifts, food, and home essentials. Businesses also tend to promote sales with specials and deals, to entice more customers, leading to a higher turnover.

(b) Use `slice()` or `filter()` to create three training subsets from this data, as follows: excluding the last year of data (first subset), then two (second subset) and three (third subset) years of data {(6 marks)}.

Answer:

```r
#First training subset - excluding last year of data
NSW_train_1 <- NSW_turnover %>% slice(1:(n()-12))
NSW_train_1
```

```
## # A tsibble: 429 x 2 [1M]
##       Month Turnover
##       <mth>    <dbl>
##  1 1982 Apr    2322
##  2 1982 May    2397.
##  3 1982 Jun    2292.
##  4 1982 Jul    2357.
##  5 1982 Aug    2266.
##  6 1982 Sep    2308.
##  7 1982 Oct    2354.
##  8 1982 Nov    2538.
##  9 1982 Dec    3430
## 10 1983 Jan    2294.
## # i 419 more rows
```

```r
#Second training subset - excluding the last two years of data
NSW_train_2 <- NSW_turnover %>% slice(1:(n()-24))
NSW_train_2
```

```
## # A tsibble: 417 x 2 [1M]
##       Month Turnover
##       <mth>    <dbl>
##  1 1982 Apr    2322
##  2 1982 May    2397.
##  3 1982 Jun    2292.
##  4 1982 Jul    2357.
##  5 1982 Aug    2266.
##  6 1982 Sep    2308.
##  7 1982 Oct    2354.
##  8 1982 Nov    2538.
##  9 1982 Dec    3430
## 10 1983 Jan    2294.
## # i 407 more rows
```

```r
#Third training subset - excluding the last three years of data
NSW_train_3 <- NSW_turnover %>% slice(1:(n()-36))
NSW_train_3
```

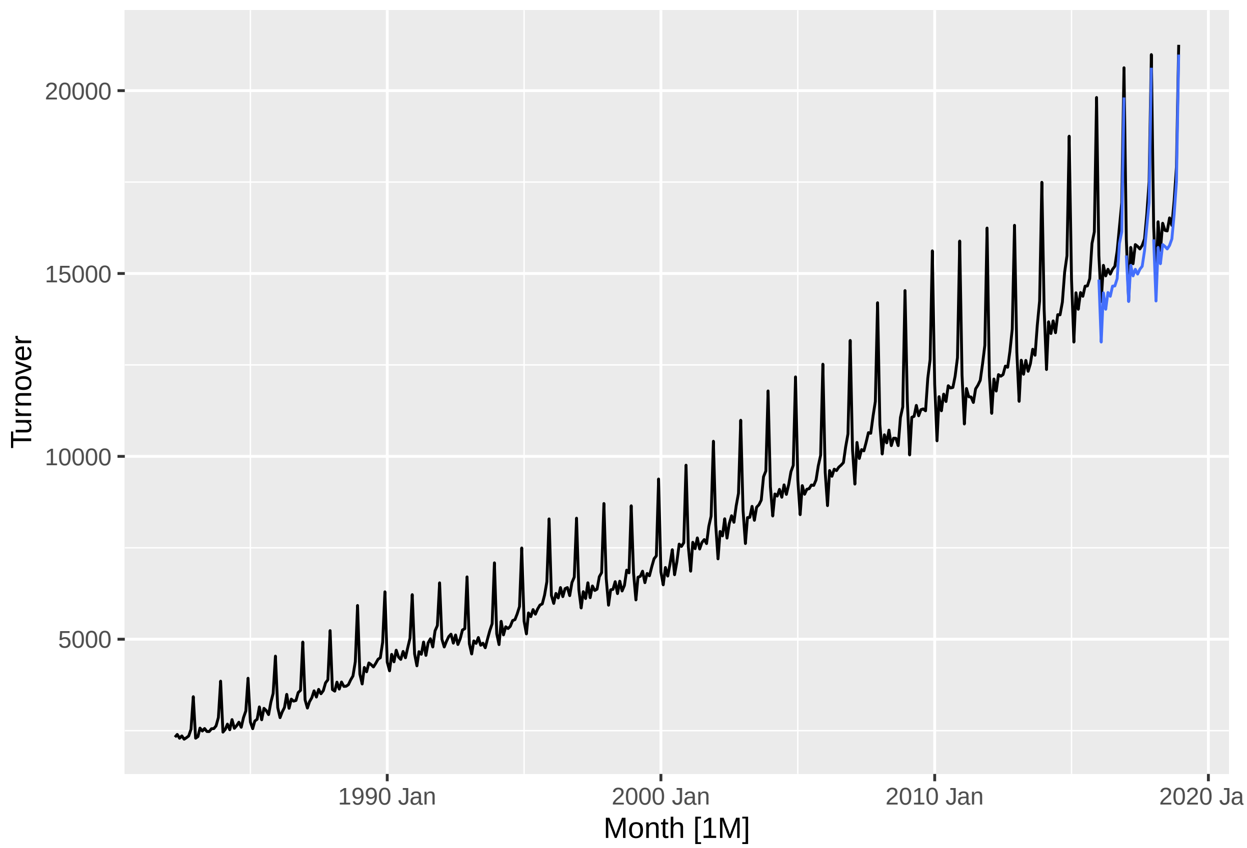
```
## # A tsibble: 405 x 2 [1M]
##       Month Turnover
##       <mth>    <dbl>
##  1 1982 Apr    2322
##  2 1982 May    2397.
##  3 1982 Jun    2292.
##  4 1982 Jul    2357.
##  5 1982 Aug    2266.
##  6 1982 Sep    2308.
##  7 1982 Oct    2354.
```

```
## 8 1982 Nov    2538.
## 9 1982 Dec    3430
## 10 1983 Jan    2294.
## # i 395 more rows
```

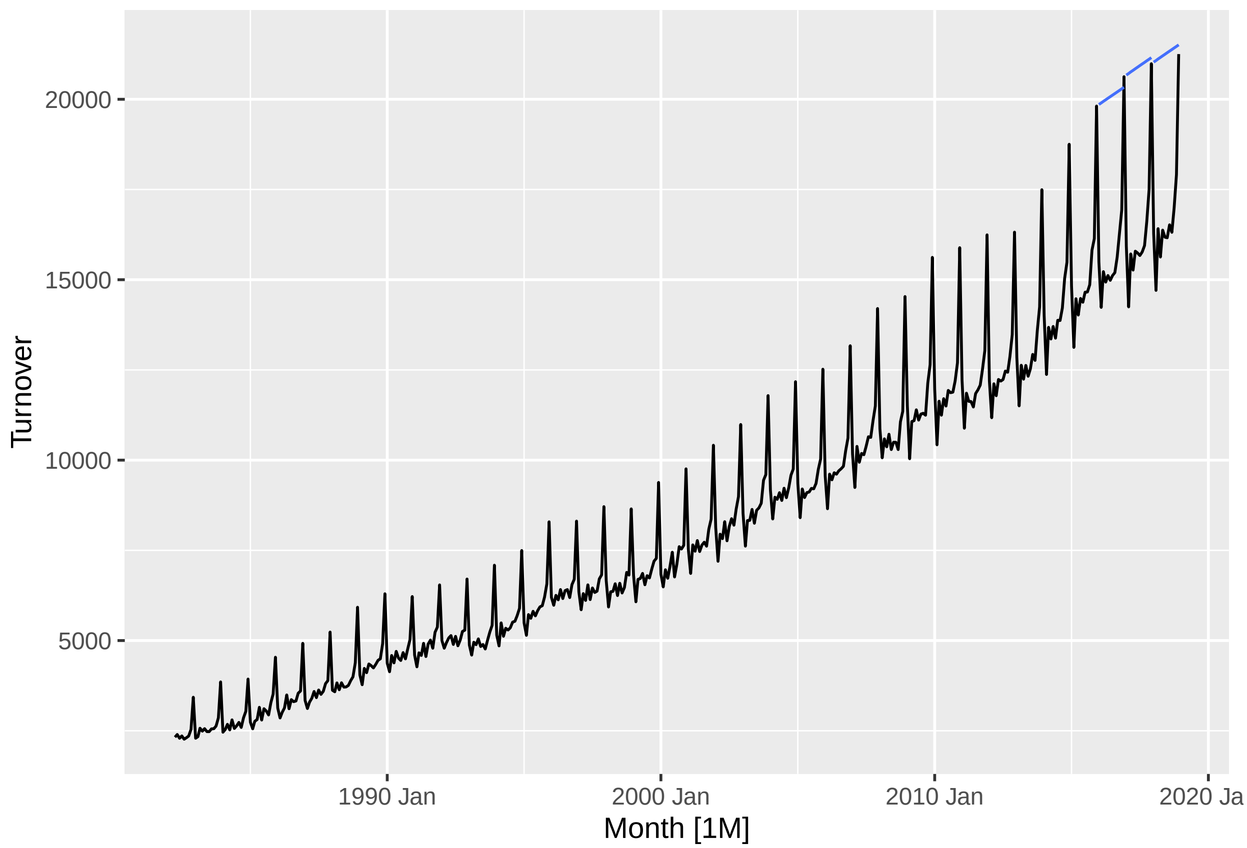
(c) Compute one year of forecasts for each training set using (a) the seasonal naive method, and (b) the drift method. Make a plot of the original series with the forecasts from the three sets in each case {(8 marks)}.

Answer:

```
#Seasonal Naive Method
#Using first training set
NSW_sn_1 <- NSW_train_1 %>%
model(SNAIVE_model_1 = SNAIVE(Turnover))
NSW_fc_sn_1 <- NSW_sn_1 %>%
forecast(h = "1 year")

#Using second training set
NSW_sn_2 <- NSW_train_2 %>%
model(SNAIVE_model_2 = SNAIVE(Turnover))
NSW_fc_sn_2 <- NSW_sn_2 %>%
forecast(h = "1 year")

#Using third training set
NSW__sn_3 <- NSW_train_3 %>%
model(SNAIVE_model_3 = SNAIVE(Turnover))
NSW_fc_sn_3 <- NSW__sn_3 %>%
forecast(h = "1 year")

#Plotting with seasonal naive method
NSW_turnover %>%
autoplot(Turnover) +
autolayer(NSW_fc_sn_1, level = NULL) +
autolayer(NSW_fc_sn_2, level = NULL) +
autolayer(NSW_fc_sn_3, level = NULL)
```




```r
#Drift Method
#Using first training set
NSW_d_1 <- NSW_train_1 %>%
model(RW_model_1 = RW(Turnover ~ drift()))
NSW_fc_d_1 <- NSW_d_1 %>%
forecast(h = "1 year")

#Using second training set
NSW_d_2 <- NSW_train_2 %>%
model(RW_model_2 = RW(Turnover ~ drift()))
NSW_fc_d_2 <- NSW_d_2 %>%
forecast(h = "1 year")

#Using third training set
NSW_d_3 <- NSW_train_3 %>%
model(RW_model_3 = RW(Turnover ~ drift()))
NSW_fc_d_3 <- NSW_d_3 %>%
forecast(h = "1 year")

#Plotting with drift method
NSW_turnover %>%
autoplot(Turnover) +
autolayer(NSW_fc_d_1, level = NULL) +
autolayer(NSW_fc_d_2, level = NULL) +
autolayer(NSW_fc_d_3, level = NULL)
```



(d) Compare the forecasts from (c) for each TEST data set. Use forecast accuracy functions in R. Which method performs best forecasting-wise in this case? {(8 marks)}.

Answer:

```
#Comparing Seasonal Naive and Drift Methods
bind_rows(
  accuracy(NSW_fc_sn_1, NSW_turnover),
  accuracy(NSW_fc_sn_2, NSW_turnover),
  accuracy(NSW_fc_sn_3, NSW_turnover),
  accuracy(NSW_fc_d_1, NSW_turnover),
  accuracy(NSW_fc_d_2, NSW_turnover),
  accuracy(NSW_fc_d_3, NSW_turnover)
  )
```

```
## # A tibble: 6 x 10
##   .model         .type     ME  RMSE   MAE    MPE  MAPE  MASE RMSSE    ACF1
##   <chr>          <chr>  <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl>   <dbl>
## 1 SNAIVE_model_1 Test    463.  484.  463.   2.81  2.81  1.13 0.977 -0.151
## 2 SNAIVE_model_2 Test    453.  490.  453.   2.79  2.79  1.11 0.990  0.100
## 3 SNAIVE_model_3 Test    708.  731.  708.   4.52  4.52  1.77 1.50   0.0734
## 4 RW_model_1     Test  -4542. 4764. 4542. -28.0  28.0  11.1  9.62   0.249
## 5 RW_model_2     Test  -4648. 4881. 4648. -29.6  29.6  11.4  9.86   0.280
## 6 RW_model_3     Test  -4282. 4534. 4331. -28.1  28.3  10.8  9.33   0.290
```

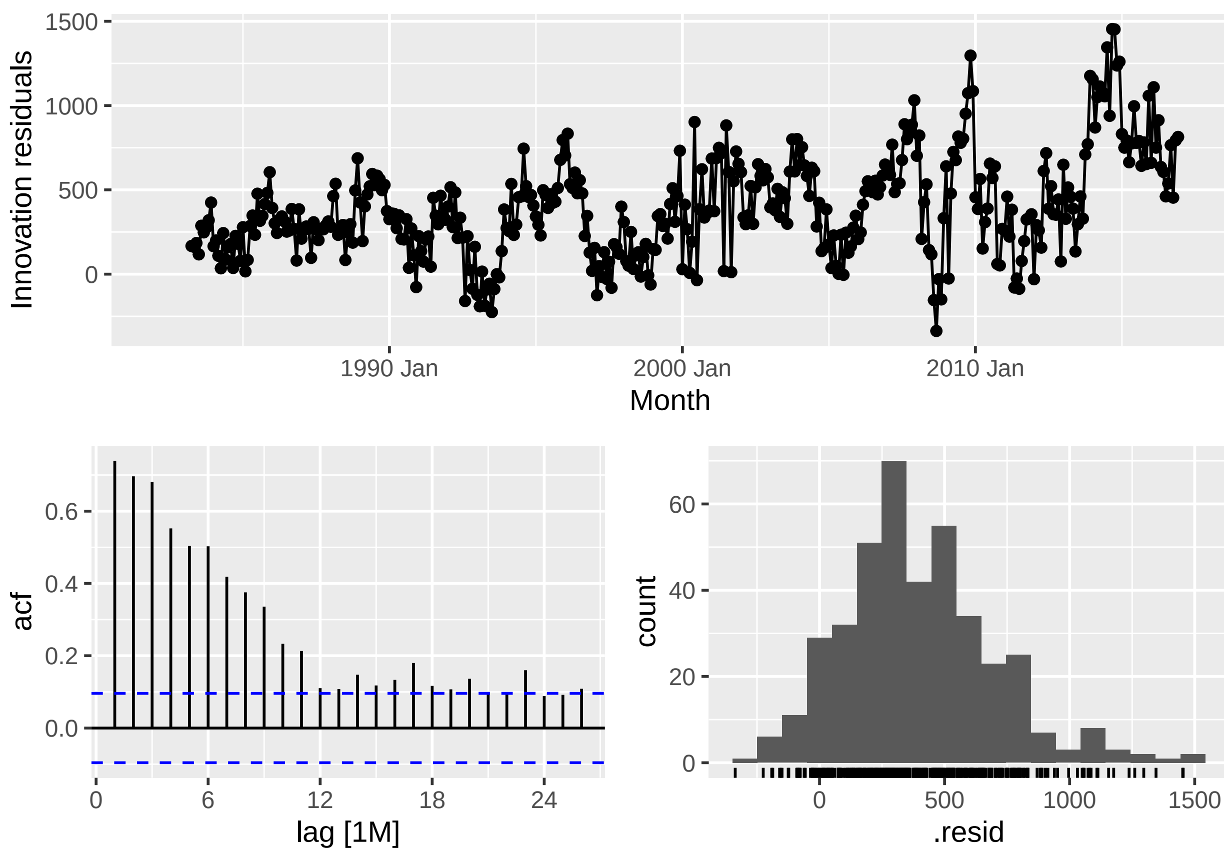
Seasonal Naive Method performs the best forecasting, in this case, in comparison to the Drift method. Seasonal Naive Method provides the lowest RMSE (Root Square Error) and MAE (Mean Absolute Error) values. NSW_fc_sn_1 provides the lowest value of 483.7 in RMSE, whereas NSW_fc_sn_2 provides the lowest value of 452.6 in MAE. MAE is less sensitive to larger outliers, making it a more balanced measure of the average forecast. Which is why NSW_fc_sn_2 performs the best out of all tests.

(e) Do the residuals from the best method resemble white noise? Briefly discuss your answer. {(3 marks)}

Answer:

```
#Checking for white noise
NSW_sn_2 %>% gg_tsresiduals()
```

```
## Warning: Removed 12 rows containing missing values or values outside the scale range
## (‘geom_line()‘).

## Warning: Removed 12 rows containing missing values or values outside the scale range
## (‘geom_point()‘).

## Warning: Removed 12 rows containing non-finite outside the scale range
## (‘stat_bin()‘).
```



The ACF plot displays multiple lags that exceed the significance threshold with majority being from lags 1 to 12, suggesting that information still exists in the residuals (indicating no white noise resemblance). The distribution of the residuals does not appear perfectly normal as it is not centered around zero, with slight skewness to the left, again not resembling white noise. I can also see an increase in variance over time, with the data being displayed in a slight funnel shape. Therefore, the residuals do not resemble white noise.

QUESTION 2

(a) Import the data into `R` and convert the data into a `tsibble`. Then, plot the series and discuss the main features, including variability and patterns {(5 marks)}.

Answer:

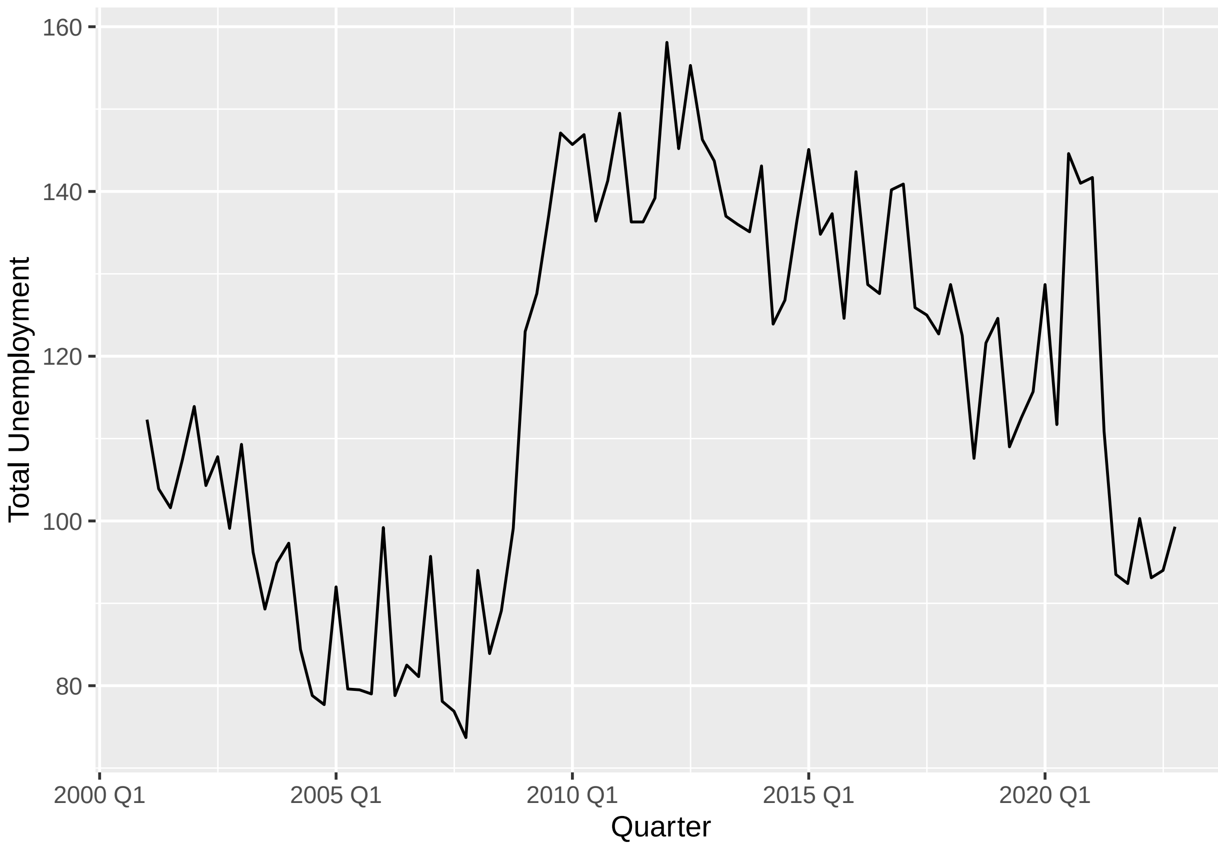
```
#Importing library
library("readxl")

#Importing data
nz_dataset <- read_excel("C:/Users/nikis/OneDrive - AUT University/Year 2 A+/Sem 2 2024/STAT603/Assignme
                         skip = 2, col_names = FALSE)
```

```
## New names:
## * `` -> `...1`
## * `` -> `...2`
```

```
#Naming columns correctly
names(nz_dataset)[1] <- "Quarter"
names(nz_dataset)[2] <- "Total Unemployment"

#Converting into a tsibble
nz_quartely_unemployment <- nz_dataset %>%
  mutate(Quarter = yearquarter(Quarter)) %>%
  as_tsibble(index = Quarter)

nz_quartely_unemployment
```

```
## # A tsibble: 88 x 2 [1Q]
##    Quarter `Total Unemployment`
##      <qtr>                <dbl>
##  1 2001 Q1                112.
##  2 2001 Q2                104.
##  3 2001 Q3                102.
##  4 2001 Q4                107.
##  5 2002 Q1                114.
##  6 2002 Q2                104.
##  7 2002 Q3                108.
##  8 2002 Q4                 99.1
##  9 2003 Q1                109.
## 10 2003 Q2                 96.2
## # i 78 more rows
```

```
#Plotting series
nz_quartely_unemployment %>%
  autoplot(`Total Unemployment`) + xlab("Quarter") + ylab("Total Unemployment")
```



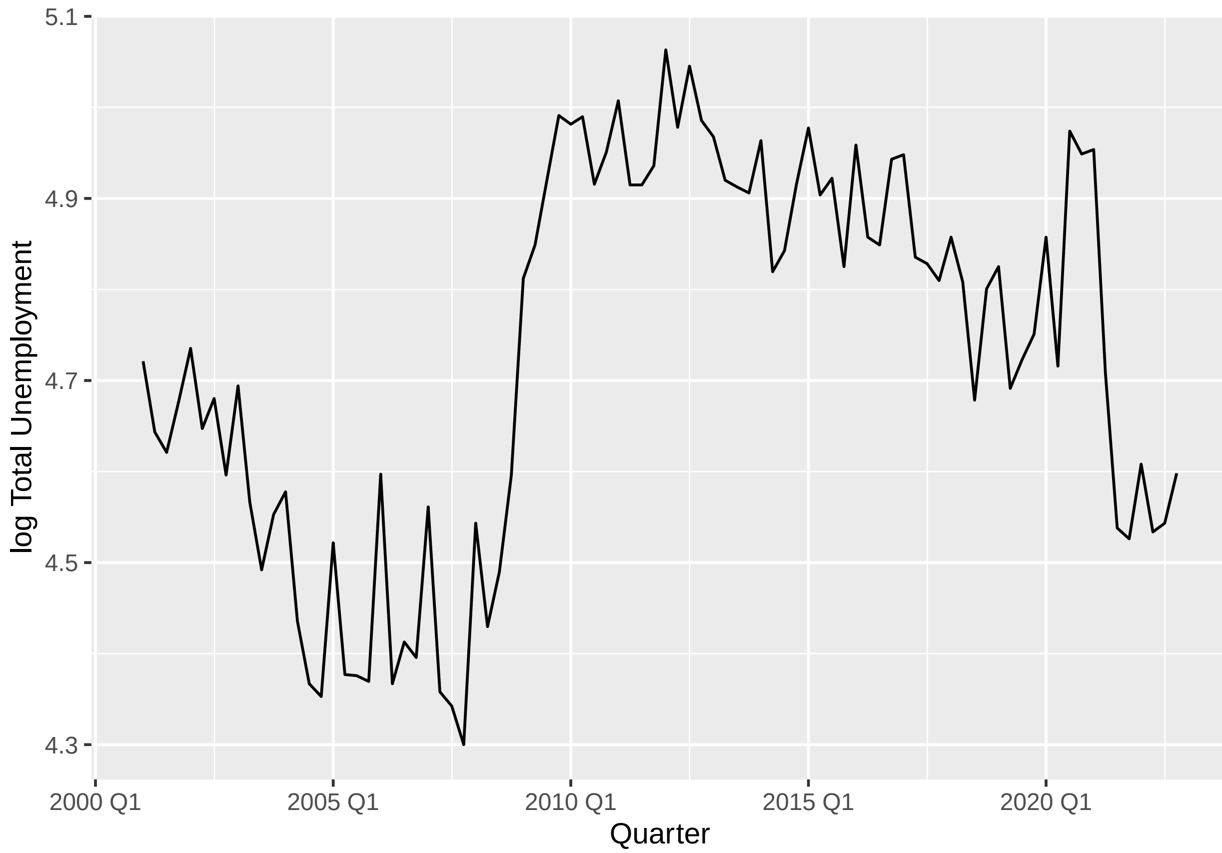
The data generally follows an upward trend, however there are some cycles where the total unemployment changes. This could be due to economic reasoning within society such as recession, expansion, peak or trough. We see a recession from the 2001 Q1 to 2007 Q4, from there it increases until 2011 Q1 where is again decreases and continues a similar cycle. The variance in the data is non-constant throughout the quarters as well.

(b) Decide whether or not a transformation is necessary to stabilize the variance. {Explain} your answer. If your decision is `yes`, then apply the transformation and describe its effect {(6 marks)}.

Answer:

The plot above shows that there is non-constant variance. The purpose of a log transformation is to stabilize the variance in a series. Which is why I think it is necessary to transform this data.

```
#Transforming data
log_nz_quartely_unemployment <- nz_quartely_unemployment %>%
  mutate(log_data = log(`Total Unemployment`))

#Plotting series
log_nz_quartely_unemployment %>%
autoplot(log_data) + xlab("Quarter") + ylab("log Total Unemployment")
```



The effect of the transformation has changed the variance to become more constant. Especially, from 2010 Q1 to 2020 Q1.

(c) Create a training data set by holding the last two years of the series. This means that the TEST data set contains the last two years of data. With the training data set, compute two years of forecasts using the (a) mean, (b) naive, (c) seasonal naive, and (d) drift methods {(8 marks)}.

Answer:

```
#Creating a training data set by holding the last two years of the series
#Using the transformed data
unemployment_train <- log_nz_quartely_unemployment %>% slice(1:(n()-8))

#Forecast using mean, naive, snaive and drift method
fit <- unemployment_train %>%
model(
mean = MEAN(log_data),
naive = NAIVE(log_data),
snaive = SNAIVE(log_data),
drift = RW(log_data ~ drift())
)

#Computing forecast with two years
fc <- fit %>% forecast(h="2 years")
```

(d) Now, compute the RMSE and the MAPE for your forecasts from (d). Which method performs best based on these measures? Explain your answer {(6 marks)}.

Answer:

```
#Computing the RMSE and the MAPE for my forecasts from above
fc %>%
  accuracy(log_nz_quartely_unemployment) %>%
  select(.model, .type, RMSE, MAPE)
```

```
## # A tibble: 4 x 4
##   .model .type  RMSE  MAPE
##   <chr>  <chr> <dbl> <dbl>
## 1 drift  Test  0.364  7.35
## 2 mean   Test  0.177  3.62
## 3 naive  Test  0.350  7.09
## 4 snaive Test  0.313  5.95
```

Based on RMSE and MAPE values, the mean method performs the best based on its measures. It has the lowest value in RMSE at 0.18 and lowest value in MAPE at 3.62%. The lower the RMSE value, indicates more precise predictions with smaller average errors. For lower MAPE values, it means smaller average percentage errors in predictions. Overall, meaning the lower the RMSE and MAPE values, the better the performance for forecasting.

QUESTION 3

(a) Fit a regression model to the quarterly unemployment data with a linear trend and seasonal dummies. Discuss the results, specifically, the coefficients of the `trend` and the `seasonal dummies`, as well as the R^2 {(8 marks)}.

Answer:

```
#Importing library
library(ggplot2)

#Fitting regression model with linear trend and seasonal dummies
myfit <- nz_quartely_unemployment %>%
model(reg = TSLM(`Total Unemployment` ~ trend() + season()))

#Viewing results
report(myfit)
```

```
## Series: Total Unemployment
## Model: TSLM
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -39.465 -17.015   3.059  17.803  42.034
##
## Coefficients:
```

```
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)   107.97052    5.96615  18.097  < 2e-16 ***
## trend()         0.37405    0.08969   4.171 7.43e-05 ***
## season()year2 -12.48768    6.43825  -1.940   0.0558 .
## season()year3 -12.28446    6.44013  -1.907   0.0599 .
## season()year4 -10.86305    6.44325  -1.686   0.0956 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.35 on 83 degrees of freedom
## Multiple R-squared: 0.2097,  Adjusted R-squared: 0.1716
## F-statistic: 5.506 on 4 and 83 DF, p-value: 0.00055596
```

From the results, I can see that the linear trend is a significant coefficient whereas seasonal dummies of quarter 2, quarter 3, and quarter 4 are not significant coefficients due to their p-values being over 0.05. The linear trend coefficient of 0.374, indicates that the unemployment increases by approximately 374 per quarter, suggesting an upward trend over time. For the seasonal dummies, the quarter 2 (season()year2) coefficient with a value of -12.488, indicates that the unemployment decreases by 12,488 in comparison to quarter 1 (baseline year). For the quarter 3 (season()year3) coefficient, the value is -12.284, indicating that the unemployment decreases by 12,284 in comparison to quarter 1. Finally, for the quarter 4 (season()year4) coefficient, the value is -10.863, indicating that the unemployment decreases by 10,863 in comparison to quarter 1. The adjusted R^2 value is at 0.1716. The low value indicates that this model does not fit well with the data.

(b) Plot the `quarterly unemployment data`, along with the `quarterly inflation` series, and the `real national disposable income` data. Conduct a correlation analysis. Discuss the results {(6 marks)}.

Answer:

```
#Importing data - Inflation data set
nz_inflation_dataset <-read_excel("C:/Users/nikis/OneDrive - AUT University/Year 2 A+/Sem 2 2024/STAT60
                         skip = 2, col_names = FALSE)
```

```
## New names:
## * `` -> `...1`
## * `` -> `...2`
```

```
#Naming the columns correctly
names(nz_inflation_dataset)[1] <- "Quarter"
names(nz_inflation_dataset)[2] <- "Inflation"

#Converting into a tsibble
nz_quartely_inflation <- nz_inflation_dataset %>%
  mutate(Quarter = yearquarter(Quarter)) %>%
  as_tsibble(index = Quarter)

nz_quartely_inflation
```

```
## # A tsibble: 88 x 2 [1Q]
##    Quarter Inflation
```

```
##      <qtr>     <dbl>
##  1 2001 Q1      -0.2
##  2 2001 Q2       0.9
##  3 2001 Q3       0.6
##  4 2001 Q4       0.6
##  5 2002 Q1       0.6
##  6 2002 Q2       1
##  7 2002 Q3       0.5
##  8 2002 Q4       0.6
##  9 2003 Q1       0.4
## 10 2003 Q2       0
## # i 78 more rows
```

```r
#Importing data - DispIncome data set
nz_dispIncome_dataset <-read_excel("C:/Users/nikis/OneDrive - AUT University/Year 2 A+/Sem 2 2024/STAT6
                        skip = 2, col_names = FALSE)
```

```
## New names:
## * `` -> `...1`
## * `` -> `...2`
```

```r
#Naming the columns correctly
names(nz_dispIncome_dataset)[1] <- "Quarter"
names(nz_dispIncome_dataset)[2] <- "Real National Disposable Income"

#Converting into a tsibble
nz_quartely_dispIncome <- nz_dispIncome_dataset %>%
  mutate(Quarter = yearquarter(Quarter)) %>%
  as_tsibble(index = Quarter)

nz_quartely_dispIncome
```

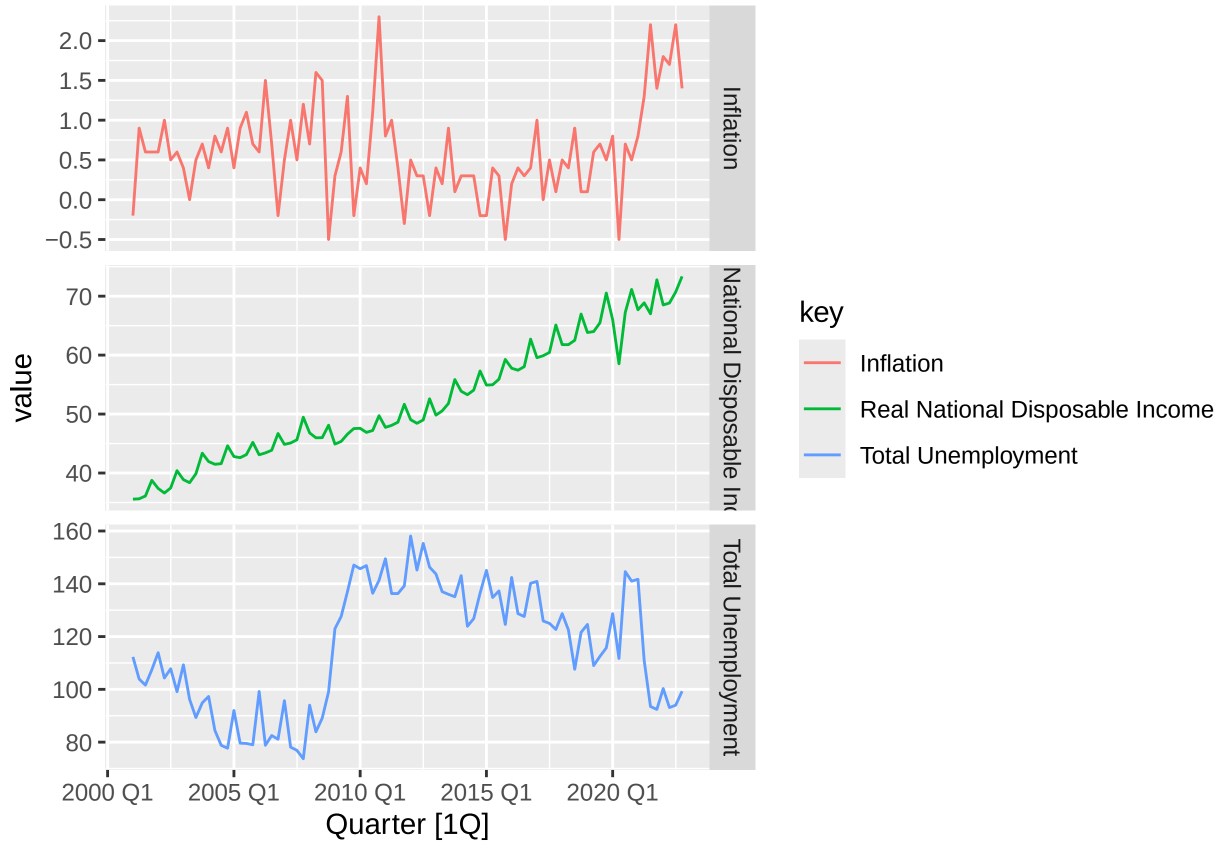
```
## # A tsibble: 88 x 2 [1Q]
##    Quarter `Real National Disposable Income`
##      <qtr>                             <dbl>
##  1 2001 Q1                              35.6
##  2 2001 Q2                              35.6
##  3 2001 Q3                              36.1
##  4 2001 Q4                              38.8
##  5 2002 Q1                              37.4
##  6 2002 Q2                              36.6
##  7 2002 Q3                              37.5
##  8 2002 Q4                              40.4
##  9 2003 Q1                              38.9
## 10 2003 Q2                              38.4
## # i 78 more rows
```

```r
#Importing libraries
library(dplyr)
library(tsibble)

#Combining data set together
```

```r
combined_quartely_nz_data <- nz_quartely_unemployment %>%
  left_join(nz_quartely_inflation, by = "Quarter") %>%
  left_join(nz_quartely_dispIncome, by = "Quarter")

combined_quartely_nz_data
```

```
## # A tsibble: 88 x 4 [1Q]
##    Quarter `Total Unemployment` Inflation `Real National Disposable Income`
##      <qtr>                <dbl>     <dbl>                             <dbl>
##  1 2001 Q1                112.       -0.2                              35.6
##  2 2001 Q2                104.        0.9                              35.6
##  3 2001 Q3                102.        0.6                              36.1
##  4 2001 Q4                107.        0.6                              38.8
##  5 2002 Q1                114.        0.6                              37.4
##  6 2002 Q2                104.        1                                36.6
##  7 2002 Q3                108.        0.5                              37.5
##  8 2002 Q4                 99.1       0.6                              40.4
##  9 2003 Q1                109.        0.4                              38.9
## 10 2003 Q2                 96.2       0                                38.4
## # i 78 more rows
```

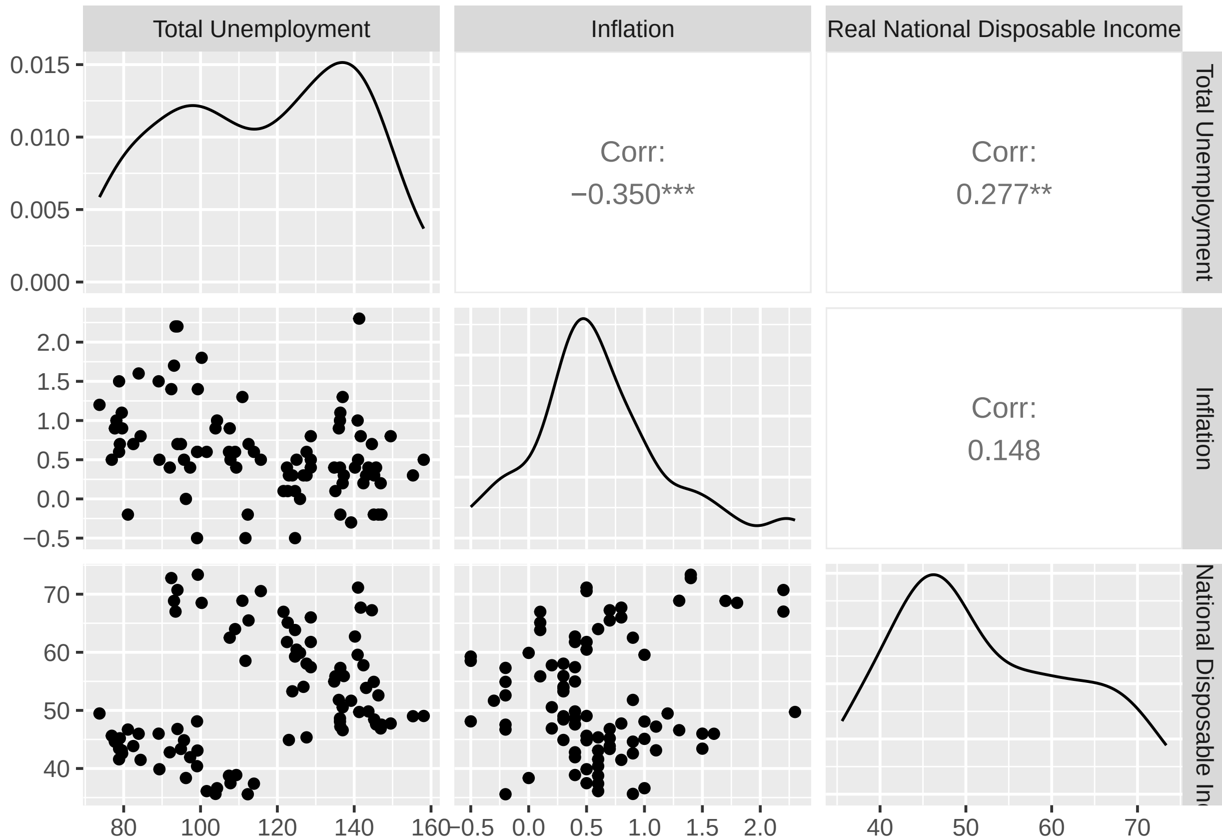
```r
#Plotting combined series
combined_quartely_nz_data %>%
  gather("key", "value", `Total Unemployment`, Inflation, `Real National Disposable Income`) %>%
  autoplot(.vars = value) +
  facet_grid(vars(key), scales = "free_y")
```




```r
#Importing library
library("GGally")
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```r
#Conducting a correlation matrix
combined_quartely_nz_data |> GGally::ggpairs(columns = 2:4)
```



From the correlation matrix above, we can see how Inflation has a highly significant negative correlation with Total Unemployment, meaning as the Inflation increase the Total Unemployment is likely to decrease. The Real National Disposable Income coefficient has a moderately high significant correlation with Total Unemployment, so as Real National Disposable Income increases the Total Unemployment tends to increase as well.

(c) Fit a regression model to the `quarterly unemployment` data using the `quarterly inflation` and the `national    disposable income` series as explanatory variables. Discuss the coefficients of both explanatories. Based on your answer to (b), are these effects reasonable? Explain briefly your answer {(6 marks)}.

Answer:

```
#Importing library
library(fable)

#Fitting regression model
myfit2 <- combined_quartely_nz_data %>%
model(reg2 = TSLM(`Total Unemployment` ~ Inflation + `Real National Disposable Income`))

#Displaying results
report(myfit2)
```

```
## Series: Total Unemployment
```

```
## Model: TSLM
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -43.4130 -15.7628  -0.7174  14.7076  54.8476
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        85.5847    11.5065   7.438 7.55e-11 ***
## Inflation                         -16.1476     3.8760  -4.166 7.41e-05 ***
## `Real National Disposable Income`   0.7643     0.2184   3.499 0.000746 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.78 on 85 degrees of freedom
## Multiple R-squared: 0.2332,  Adjusted R-squared: 0.2151
## F-statistic: 12.92 on 2 and 85 DF, p-value: 1.2594e-05
```

Both Inflation and Real National Disposable Income are good predictors of Total Unemployment due to their p-values being lower than 0.05. Which is reflected in the correlation matrix making these effects reasonable. The Inflation coefficient with a value of -16.148, means the unemployment is expected to decrease by approximately 16,148 per percentage change from previous period (%). The Real National Disposable Income coefficient with a value of 0.764, for each unit increase, the unemployment is expected to increase by 764. The adjusted R^2 value is at 0.2151. This low value indicates that this model does not fit too well with the data.

(d) Do we need to include the linear trend and seasonal dummies in the regression model in (c)? Perform an appropriate relevant analysis and discuss {(5 marks)}.

Answer:

```
#Model without linear trend and seasonal dummies
data_without_trend_seasonal <- combined_quartely_nz_data %>%
     model(TSLM(`Total Unemployment` ~ Inflation + `Real National Disposable Income`))
glance(data_without_trend_seasonal) |> select(.model:AICc)
```

```
## # A tibble: 1 x 10
##   .model    r_squared adj_r_squared sigma2 statistic p_value    df log_lik   AIC
##   <chr>         <dbl>         <dbl>  <dbl>     <dbl>   <dbl> <int>   <dbl> <dbl>
## 1 TSLM(`To~     0.233         0.215   432.      12.9 1.26e-5     3   -390.  539.
## # i 1 more variable: AICc <dbl>
```

```
#Model with linear trend and seasonal dummies
data_with_trend_seasonal <- combined_quartely_nz_data %>%
      model(TSLM(`Total Unemployment` ~ trend() + season() + Inflation + `Real National Disposable Inco
glance(data_with_trend_seasonal) |> select(.model:AICc)
```

```
## # A tibble: 1 x 10
##   .model   r_squared adj_r_squared sigma2 statistic  p_value    df log_lik   AIC
##   <chr>        <dbl>         <dbl>  <dbl>     <dbl>    <dbl> <int>   <dbl> <dbl>
## 1 "TSLM(`~     0.606         0.577   233.      20.8 1.38e-14     7   -361.  488.
## # i 1 more variable: AICc <dbl>
```

Above I have decided to compare the AICc values of both models, one model excluding the linear trend and seasonal dummies and the other including them. AICc balances model fit, it penalizes models that use more parameters to avoid overfitting, so the lower the AICc value the better it's performance. From both models, I can see that the AICc for the model without linear trend and seasonal dummies has a value of 539, whereas the model with them has a value of 490. The model that includes the linear trend and seasonal dummies performs better as the AICc has a lower value in comparison.

END OF ASSINGMENT 1