# assignment3

Izzy Southon

03/09/2024

## Assignment 3 Time Series

### Question 1 (17 Marks) Unemployment

```
# setting up
# load in Household Labour Force Survey (HLFS) spreadsheet
data <- read.csv("hlfs-jun24qtr-csv.csv")
```

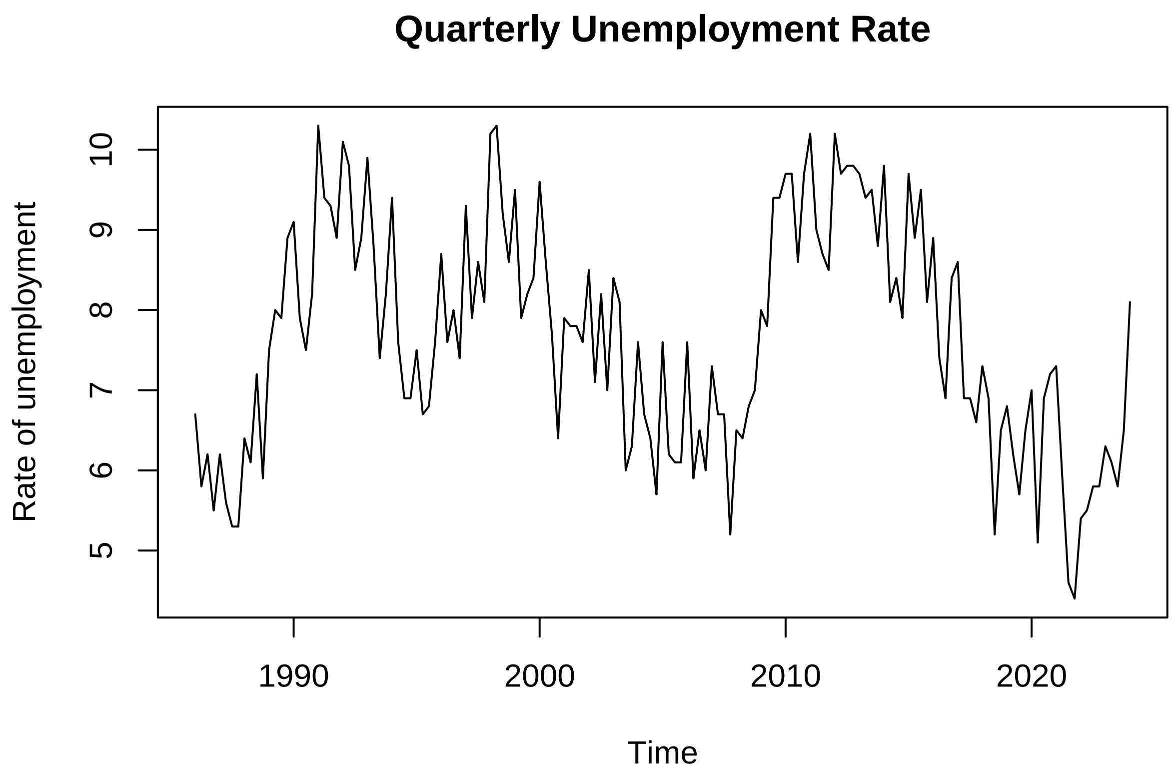
```
# save series 'HLFQ.SIQ3' to a data frame
rate <- data %>% filter(Series_reference == "HLFQ.SIQ3")
```

```
# save series 'HLFQ.SIK3' to a data frame
persons <- data %>% filter(Series_reference == "HLFQ.SIK3")
```

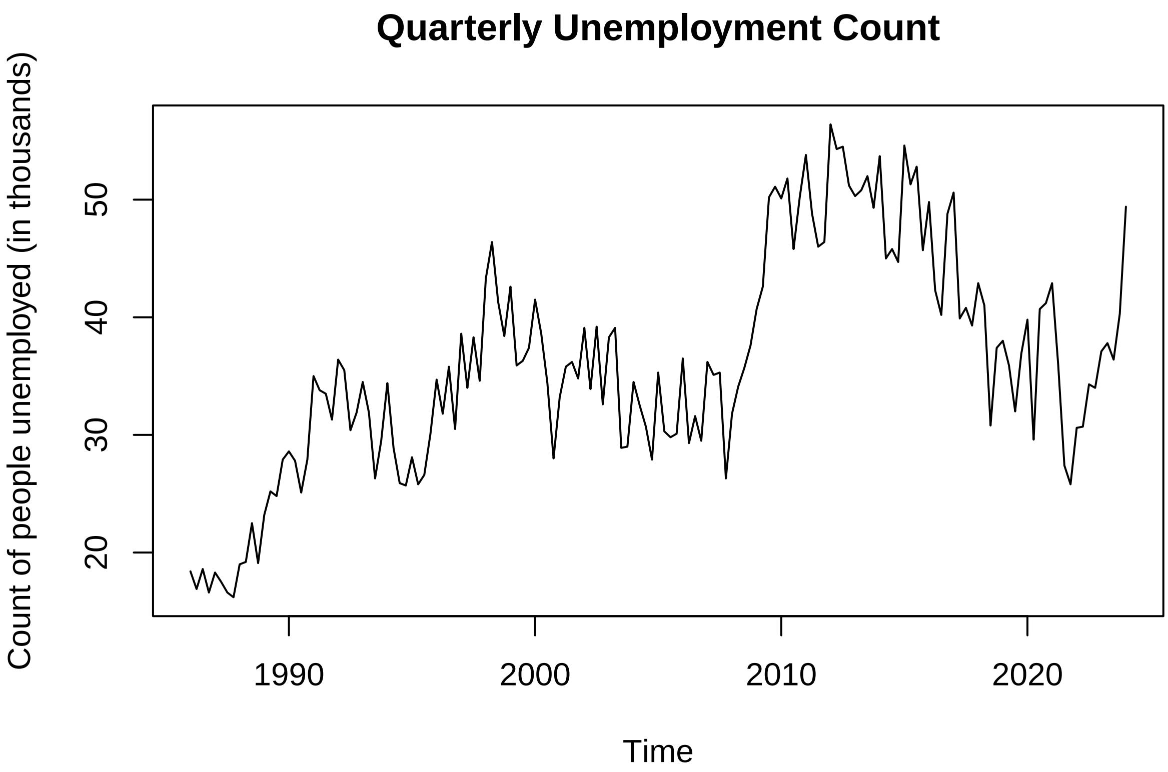
**a) Convert the unemployment rate and unemployment counts to time series objects and plot them. Make sure you derive year and quarter of each observation from the Period column. The Period value is a combination of the year + the number of the month at the end of the quarter. e.g. 1990.09 is the September (3rd) quarter of 1990.**

```
# subsetting
#rate$Period <- sub("\\.03$", " Q1", rate$Period)
#rate$Period <- sub("\\.06$", " Q2", rate$Period)
#rate$Period <- sub("\\.09$", " Q3", rate$Period)
#rate$Period <- sub("\\.12$", " Q4", rate$Period)

#persons$Period <- sub("\\.03$", " Q1", persons$Period)
#persons$Period <- sub("\\.06$", " Q2", persons$Period)
#persons$Period <- sub("\\.09$", " Q3", persons$Period)
#persons$Period <- sub("\\.12$", " Q4", persons$Period)

# define the quarters in a list
replacements <- list("\\.03$" = " Q1", "\\.06$" = " Q2", "\\.09$" = " Q3", "\\.12$" = " Q4")

# Apply the replacements using a for loop
for (pattern in names(replacements)) {
  replacement <- replacements[[pattern]]
  rate$Period <- sub(pattern, replacement, rate$Period)
  persons$Period <- sub(pattern, replacement, persons$Period)
}
```

```r
# derive year and quarter from Period column for both dataframes
rate$Period <- yearquarter(rate$Period)
persons$Period <- yearquarter(persons$Period)

# plot the unemployment rate
ts_rate <- ts(rate$Data_value, start=1986, end=2024, frequency=4)
plot(ts_rate, ylab="Rate of unemployment",
     main="Quarterly Unemployment Rate")
```

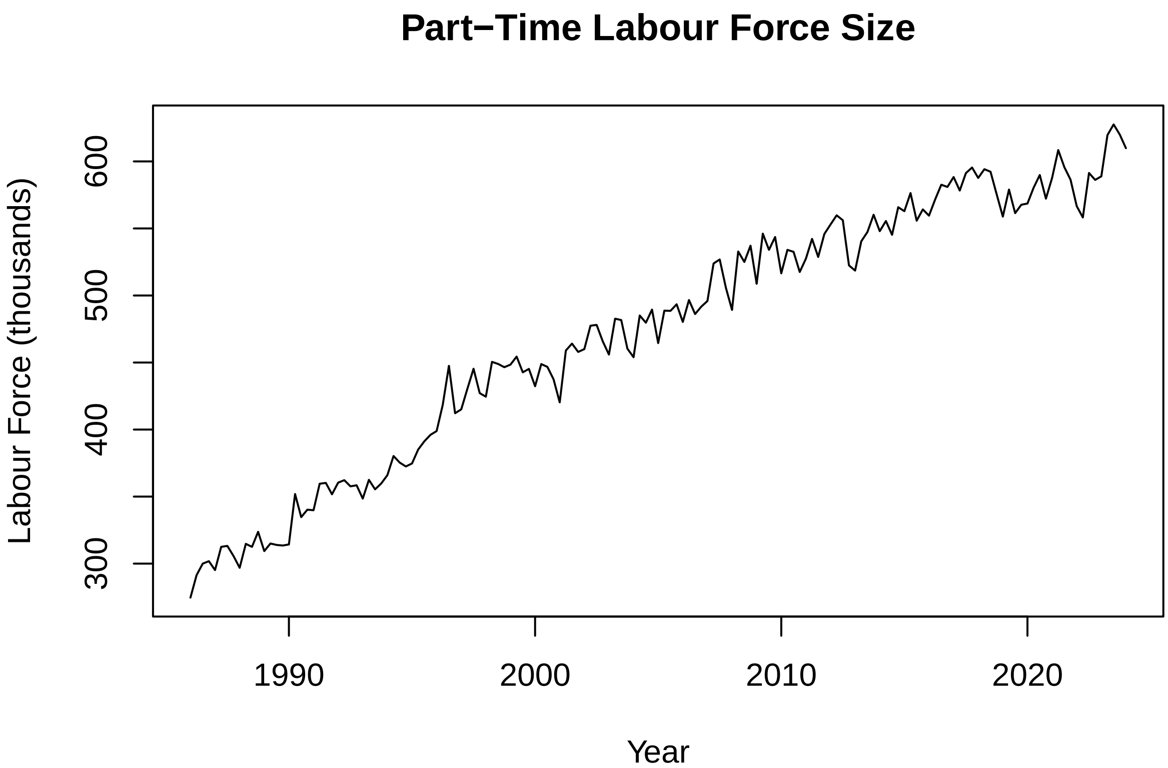



```r
# plot the unemployment count
ts_count <- ts(persons$Data_value, frequency = 4, start=1986, end=2024)
plot(ts_count, ylab="Count of people unemployed (in thousands)",
     main="Quarterly Unemployment Count")
```




**b) Estimate the size of the part time labour force for each quarter from these two time series. Plot and comment on what you find.**

```
labour_force <- ts_count / (ts_rate / 100)
plot(labour_force, main = "Part-Time Labour Force Size", xlab = "Year", ylab = "Labour Force (thousands)
```




**Comments on plot:**

- The plot above shows that broadly speaking, overtime from years 1986 to 2024, the estimated part-time labour force increased in size.
- There is a clear, increasing trend.
- At a particular point of each year, the estimated part-time labour force decreases in size, this could be due to students graduating university and beginning full-time roles instead of part-time roles.
- During some quarters, the estimated part-time labour force decreased in size for a short period, till increasing again.

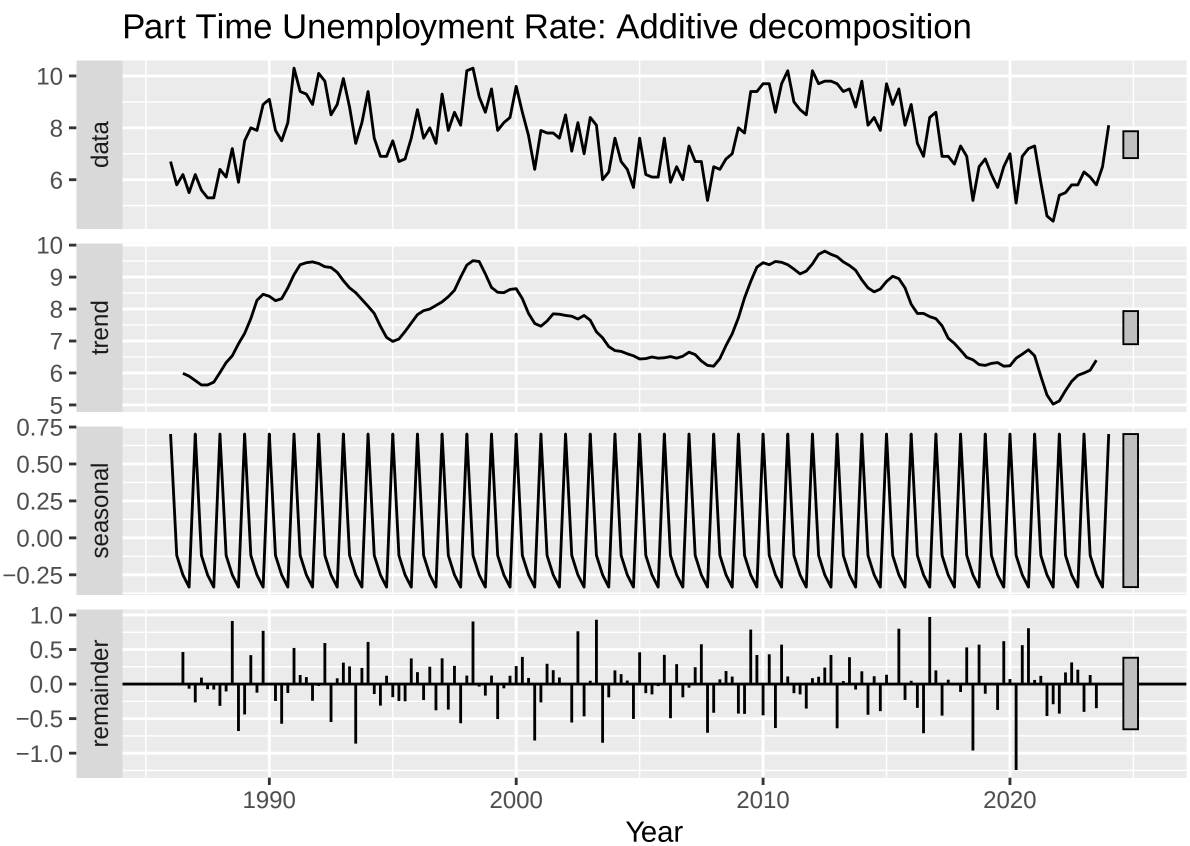
**c) Carry out a seasonal decomposition of the part time unemployment rate, and graph the result.**

Spilt a time series Yt into

- a trend, Tt
- a seasonal component, St
- a residual irregular component, It

Additive: Yt = Tt + St + It

```r
# convert to tsibble
ts_rate_tsibble <- ts_rate %>%
  as_tsibble(index = Period)

ts_rate_tsibble %>%
  select(value) %>% as.ts() %>%
  decompose(type="additive") %>%
  autoplot() + xlab("Year") +
  ggtitle("Part Time Unemployment Rate: Additive decomposition")
```




**d) Test the count of unemployed people for stationarity, also test the first and second differences. Comment on what you find.**

```r
# original
unitroot_kpss(ts_count)
```

```
##   kpss_stat kpss_pvalue
##    1.548602    0.010000
```

```r
# first difference
unitroot_kpss(diff(ts_count))
```

```
##   kpss_stat kpss_pvalue
##  0.05390228  0.10000000
```

```r
# second difference
unitroot_kpss(diff(diff(ts_count)))
```

```
##   kpss_stat kpss_pvalue
##  0.03197416  0.10000000
```

```r
#unitroot_kpss(diff(ts_count, differences = 2))
```

**Comments on results:**

The null hypothesis of the KPSS test is that the series is stationary, while the alternative hypothesis is that the series is non-stationary. Here the original data fails but the first and second differenced data passes the test for stationarity. This is because with the original data, the p-value is less than the significance level of 0.05, thus indicating that the time series is non-stationary. The high p-values and low KPSS statistics for the first and second difference suggest that the time series is stationary.

**e) Fit a seasonal ARIMA model to the count of unemployed people, and interpret the result - referring to your answer regarding stationarity.**

```r
model <- ts_count %>% auto.arima() # use time series not tibble
model
```

```
## Series: .
## ARIMA(0,1,2)(0,0,2)[4]
##
## Coefficients:
##           ma1      ma2    sma1    sma2
##       -0.3487  -0.1322  0.1660  0.1458
## s.e.   0.0817   0.0782  0.0911  0.0835
##
## sigma^2 = 16.6:  log likelihood = -427.4
## AIC=864.79   AICc=865.21   BIC=879.91
```

**Interpret the result referring to your answer regarding stationarity:**

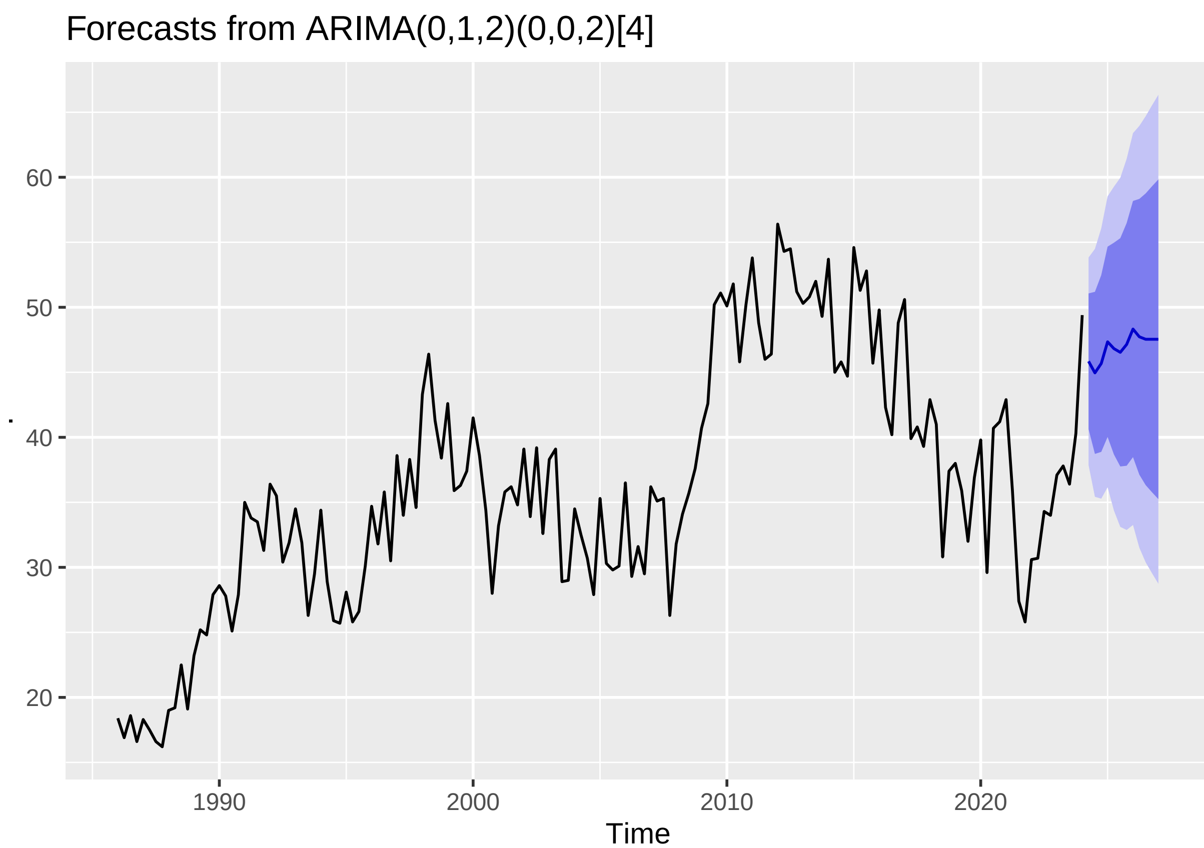
Originally, the time series was non-stationarity, indicated by the KPSS test. The ARIMA(0,1,2)(0,0,2)[4] model captures the neccessary differencing to make the series stationary. The combination of the KPPS test results and the ARIMA model confirms that the original time series needed to be differenced to achieve stationarity.

ARIMA(0,1,2) captures the non-seasonal part. There are no autoregressive terms, the series is differenced once to achieve stationarity, and there are two moving average terms.

(0,0,2)[4] is the seasonal part. There are no seasonal autoregressive terms, there is no seasonal differencing, and there are two seasonal moving average terms. The [4] indicates quarterly seasonality. Lower values of AIC, corrected AIC (AICc), and BIC indicate a better model fit.

**f) Forecast the ARIMA fit three years (12 quarters) ahead. Plot the result, and give a 95% prediction interval for the number of part time unemployed in the March Quarter of 2027, commenting on the quality of the result.**

```r
# 'model' is the. ARIMA model
forecast_fit <- forecast(model, h = 12)

# plot the forecast
autoplot(forecast_fit)
```




```r
# extract the forecasted value for the March Quarter of 2027
march_2027_forecast <- forecast_fit$mean[1]

# extract the 95% prediction interval
march_2027_lower <- forecast_fit$lower[1]  # Lower bound
march_2027_upper <- forecast_fit$upper[1]  # Upper bound

# print the results
cat("Forecast for March Quarter 2027 (Q1 2027):\n")
```

```
## Forecast for March Quarter 2027 (Q1 2027):
```

```r
cat("Point Estimate: ", march_2027_forecast, "\n")
```

```
## Point Estimate:  45.8454
```

```r
cat("95% Prediction Interval: [", march_2027_lower, ",", march_2027_upper, "]\n")
```

```
## 95% Prediction Interval: [ 40.62337 , 51.06743 ]
```

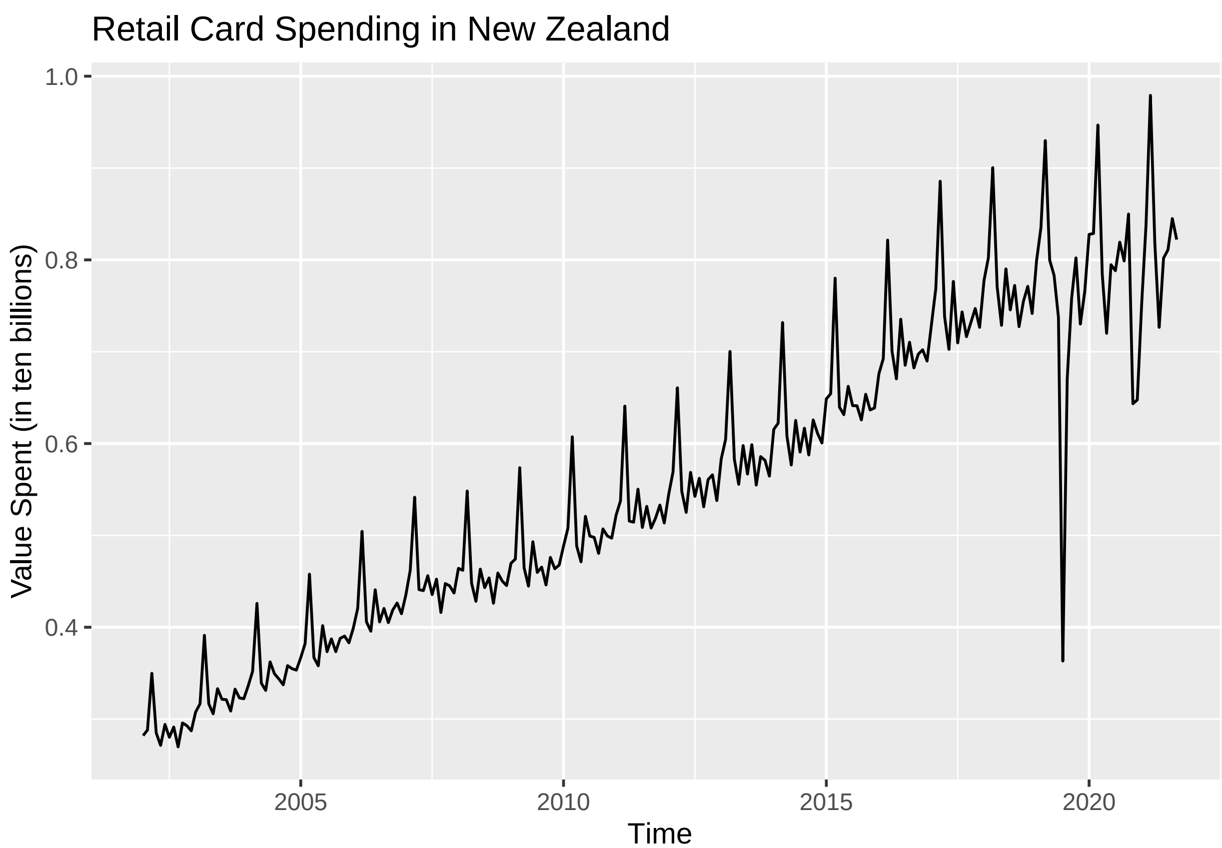
**Comment on the quality of the result:**

This is a relatively narrow prediction interval, indicating more certainty when forecasting the estimated part-time unemployed labour force in the March quarter of 2027. The 95% prediction interval ranges from 40.62 to 51.07. This means that there's a 95% chance that the actual number of part-time unemployed in the March Quarter of 2027 will fall within this range.

**Question 2 (10 Marks) Card Spending**

**a) Plot the data and comment on the general features of it.**

```
cards <- read.csv("card.csv")
#cards$Period <- as.Date(cards$Period)
ts_cards <- ts(cards$Value/10000000000, start = c(2002, 1), frequency = 12) # Starts in 2002, 1st quart

ts_cards %>%
 autoplot() +
  ggtitle("Retail Card Spending in New Zealand") +
  xlab("Time") + ylab("Value Spent (in ten billions)")
```

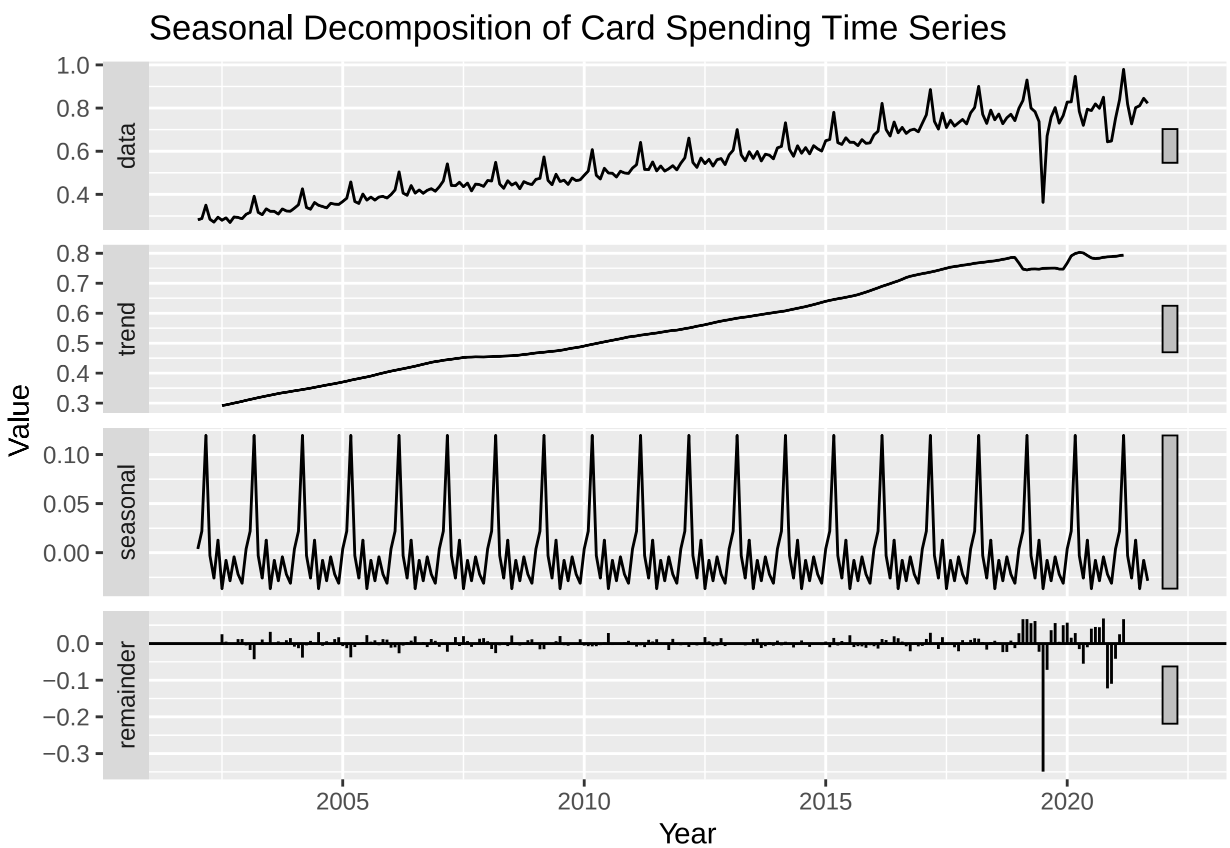



**Comment on the general features:**

- Clear increasing trend.
- There is also a strong seasonal pattern that increases in size as the level of the series increases.
- The sudden spike in retail card spending at the beginning of each year could be a result of new years sales and end of season sales in retail stores.
- Huge decrease in retail card spending in 2019. This could have been due to the impending Covid-19 lockdown in NZ that began 25 March 2020. Inflation, people have less money to spend due to job loss.
- Around 2022 there was a slight decrease in retail card spending. This might be due to inflation and people having less money to spend.

**b) Carry out a seasonal decomposition, and comment on the features of each of the components.**

```r
# Perform the seasonal decomposition
decomp <- ts_cards %>%
  decompose(type = "additive")

# Plot the decomposed components
autoplot(decomp) +
  ggtitle("Seasonal Decomposition of Card Spending Time Series") +
  xlab("Year") +
  ylab("Value")
```

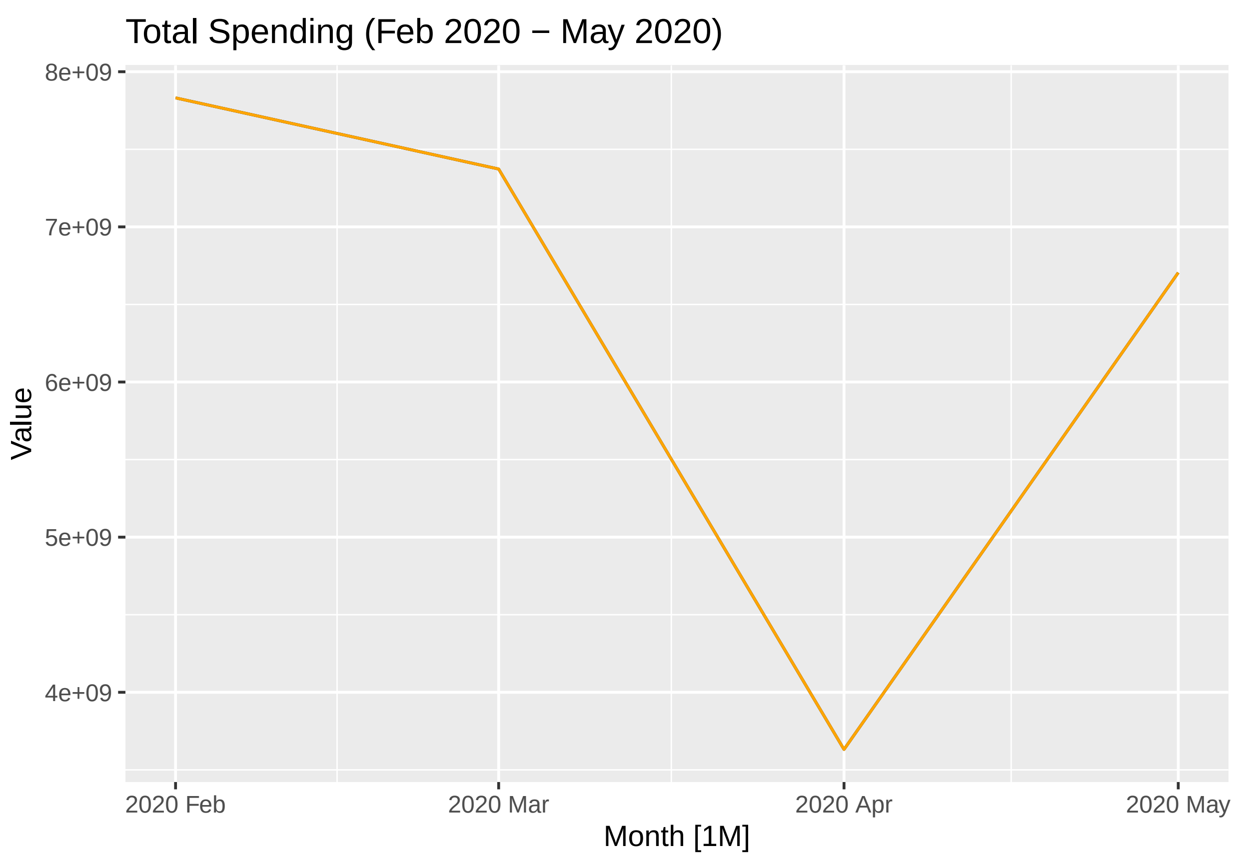



- **data:** Shows the original data. Retail card spending increases overtime with a drop in 2019.
- **trend:** Shows an increase in retail card spending overtime, with a decrease in card spending in 2019.
- **seasonal:** Repeating patterns of retail card spending can be observed. Higher card spending in NZ occurs in particular months (increases the summer season months and decreases in the winter months).
- **remainder:** A consistent pattern of positive and negative groups can be observed. There is a decrease in retail card spending in 2019.

The grey bars on the right of each panel show the relative scales of the components. Since the plots are on different scales, the bars vary in size when they actually represent the same length.

**c) Make a rough estimate the amount of card spending volume that was lost during the month of April 2020. (Give your answer in billions of dollars lost.) There are several ways you could do this: briefly explain the method you choose to use.**

```r
# convert Period to yearmonth format and create a tsibble
cards <- cards %>%
  mutate(Month = yearmonth(Period)) %>%
  as_tsibble(index = Month)

# filter data for specific months
filtered_cards <- cards %>%
  filter(Month %in% yearmonth(c("2020 Feb", "2020 Mar", "2020 Apr", "2020 May"))) %>%
  as_tsibble()

# plot the filtered data
filtered_cards %>%
  autoplot(Value) +
  geom_line(aes(y = Value), colour = "orange") +
  labs(y = "Value", title = "Total Spending (Feb 2020 - May 2020)")
```




I chose to use a method that shows the difference in card spending volume between March 2020 and April 2020.

- In March 2020, \$7372900000 was spent.
- In April 2020, \$3631600000 was spent.

$7372900000 - 3631600000 = 3741300000$ dollars lost. 3.741 billion dollars were lost during April 2020 from March 2020. I chose to use this method as you can visibly see the amount of volume spent decrease. Comparing April 2020 with March 2020 (the month prior) shows how much spending volume decreased over the period of one month. I chose to look at the raw data from the 'cards' dataframe to get the value of retail card spending for March 2020 and April 2020. I also used a graph to visualise the card spending volume lost during the month of April 2020.

### References

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