

# MATH1318 Time Series Analysis

## Assignment 2

Joyal Joy Madeckal - s3860476

### Introduction

Autoregressive integrated moving average (ARIMA) model is one of the important statistic analysis model for modelling time series data for better understanding of the data set and to predict the future trends. For this assignment we are given a data set which shows the Antarctica land ice mass in billion metric tons relative to the ice mass in 2001. We have the following objectives for the assignment.

- Perform descriptive analysis on the time series data
- Analyse the time series data with ACF and PACF plots and perform the unit root tests for stationarity.
- Deal with non-stationarity of the data if exists and perform the unit root tests and ACF-PACF analysis
- Use ACF-PACF, EACF and BIC table to find the suitable ARIMA models for the time series data.

### Descriptive Analysis

- First, we will load the data and check data statistics.
- We will load the required packages for loading the data

```
library(TSA)
```

```
##  
## Attaching package: 'TSA'  
  
## The following objects are masked from 'package:stats':  
##  
##   acf, arima  
  
## The following object is masked from 'package:utils':  
##  
##   tar
```

```
library(tseries)
```

```
## Registered S3 method overwritten by 'quantmod':  
##   method      from  
##   as.zoo.data.frame zoo
```

```
# Loading the required data  
raw_ice_mass <- read.csv("assignment2Data2022.csv", col.names = c("Year", "Ice_Mass"))  
# Checking the summary statistics of the data  
summary(raw_ice_mass)
```

```
##      Year      Ice_Mass
## Min.   :2002   Min.    :-2539.055
## 1st Qu.:2006   1st Qu. :-1785.137
## Median :2011   Median  : -938.991
## Mean   :2011   Mean    :-1063.774
## 3rd Qu.:2016   3rd Qu. : -289.390
## Max.   :2020   Max.    :   -7.547
```

- We can see that all the values in the data are negative. So, if we need to do transformation then we have to make all the values positive.
- Now we will convert the data into time series object

```
ice_mass_ts <- ts(raw_ice_mass$Ice_Mass, start = 2002, end = 2020)
# Plotting the time series data
plot(ice_mass_ts, xlab = "Year", ylab = "Annual Antarctica land ice mass",
     main = "Time series plot for yearly change in Antarctica Land Ice Mass", type = "o")
```

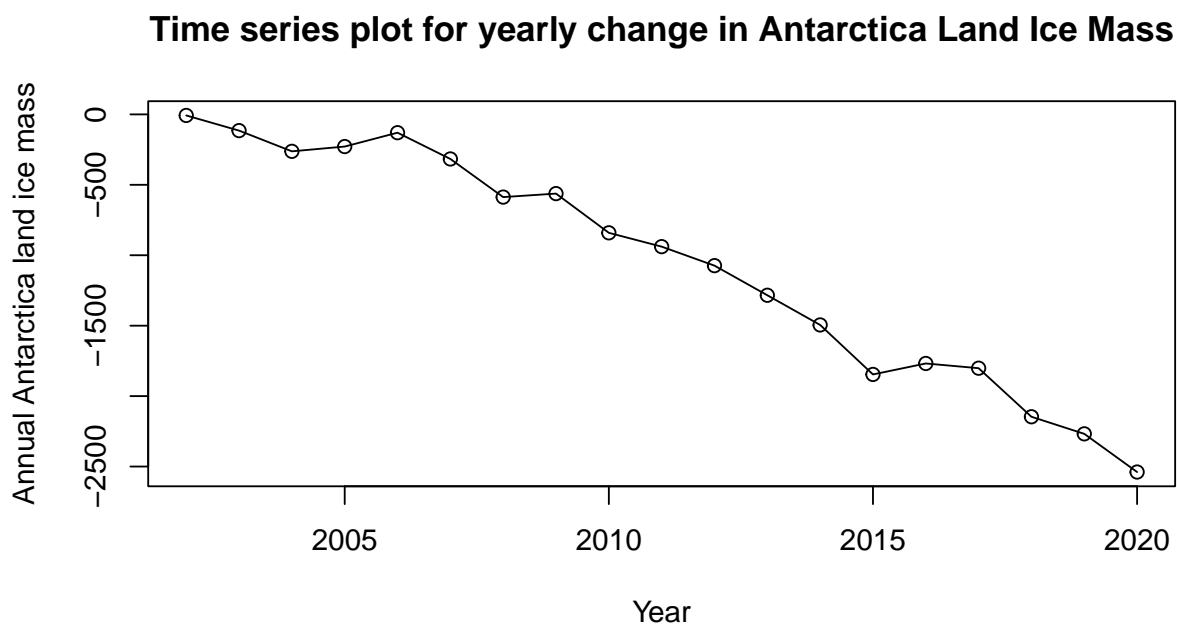


Figure 1: Time series plot for yearly change in Antarctica Land Ice Mass

- The time series plot looks like there are no patterns within the series.
- The plot appears to be an auto regressive (AR) process.

We will analyse the time series plot based on the following statistics:

- **Trend** - We can clearly see that there is a downward trend through out the series. From year 2000 till 2020 the plot is going downwards and it aligns with the global fact that the ice in Antarctica is melting.

- **Seasonality** - The plot is not showing any seasonality. But if we closely analyse we can say that the pattern after 2005 appear to be repeating after 2015. But as we have fewer data points it is not possible to conclude that. So, we will take it as there is no repeating pattern in the series.
- **Changing Variance** - We cannot see any changing variance in the series.
- **Behaviour** - The data is mostly showing AR characteristics. We can't find any moving average (MA) characteristics in the data.
- **Change Point** - We cannot see any change point for the time series data.

## Stationarity Validation

As there is a clear downward trend and based on the above statistics we can clearly say that the series is non-stationary. We have to deal with this first and we need to ensure the series is stationary before proceeding forward. We will confirm the series is non-stationary using Auto correlation function (ACF) and Partial auto correlation function (PACF) plots.

```
# Function for ACF and PACF plots
plot_acf_pacf <- function(acf,pacf) {
  par(mfrow=c(1,2))
  acf(ice_mass_ts, main = "ACF plot of
    Antarctica Land Ice Mass")
  pacf(ice_mass_ts, main = "PACF plot of
    Antarctica Land Ice Mass")
}

plot_acf_pacf("ACF plot of
  Antarctica Land Ice Mass", "PACF plot of
  Antarctica Land Ice Mass")
```

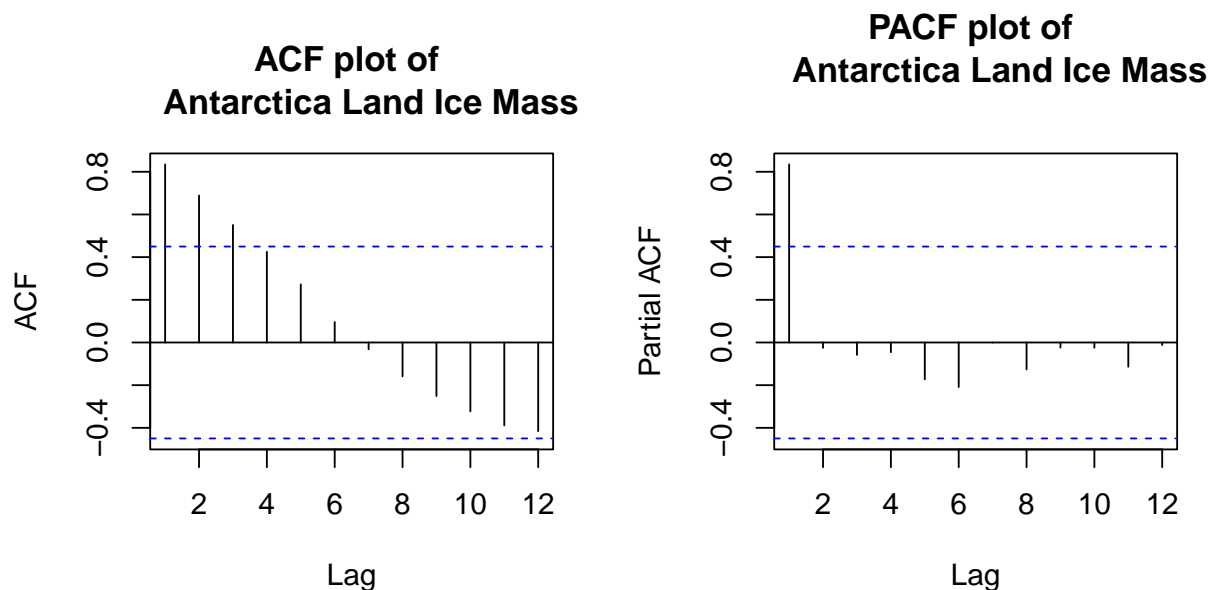


Figure 2: ACF and PACF plots

- From the ACF plot we can see a slowly decaying pattern which indicates the series we are dealing with is non-stationary

- The PACF plot shows us a large first lag followed by lower lags at succeeding lags which also confirms the series is non stationary.
- We will perform the Augmented Dickey-Fuller test now for testing stationarity.

```
adf.test(ice_mass_ts, alternative=c("stationary"))
```

```
##
## Augmented Dickey-Fuller Test
##
## data: ice_mass_ts
## Dickey-Fuller = -2.7141, Lag order = 2, p-value = 0.3004
## alternative hypothesis: stationary
```

- The p-value is very high ( $>0.05$  significance level) indicating we cannot reject the null hypothesis that the series is non-stationary. So, all the tests and plot are pointing to the fact that the series is non-stationary.
- We will also see the normality of the data by using the Q-Q plot and Shapiro-Wilk normality test.

```
qqnorm(ice_mass_ts)
qqline(ice_mass_ts, col = 2)
```

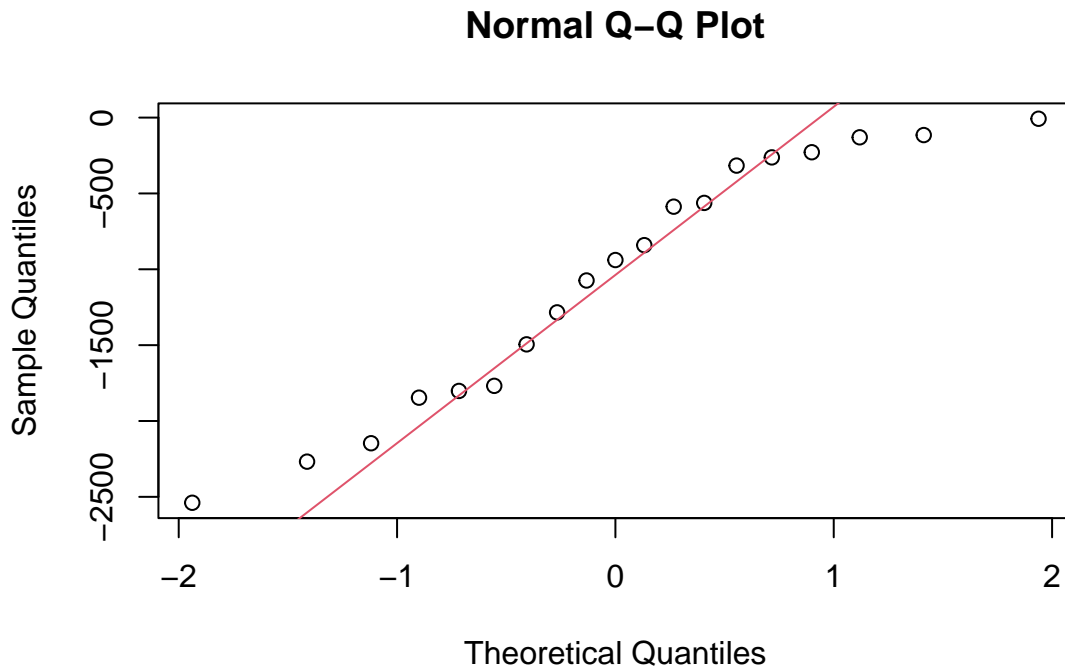


Figure 3: Q-Q Plot

```
shapiro.test(ice_mass_ts)
```

```
##
```

```
## Shapiro-Wilk normality test
##
## data: ice_mass_ts
## W = 0.92616, p-value = 0.1471
```

- The Q-Q plot shows that majority of the points are coming nearer to the normal line and Shapiro test confirms the normality of the data.

## Stationary Conversion

- As we know, we cannot proceed with non-stationary series we have to convert the series to be stationary. First differencing will be done for converting the series to stationary.
- If there are any transformations to be done, it has to be done before differencing. But, as we are having a series with no changing variance and as normality of the data is confirmed we don't need to perform any transformation to the series.

```
# Applying differencing to the data set
ice_mass_ts_diff <- diff(ice_mass_ts, differences = 1)
# Plotting the differenced series
plot(ice_mass_ts_diff, ylab = "First difference of Antarctic land ice mass", xlab = "Year",
     main = "First difference of Antarctic land ice mass", type="o")
```

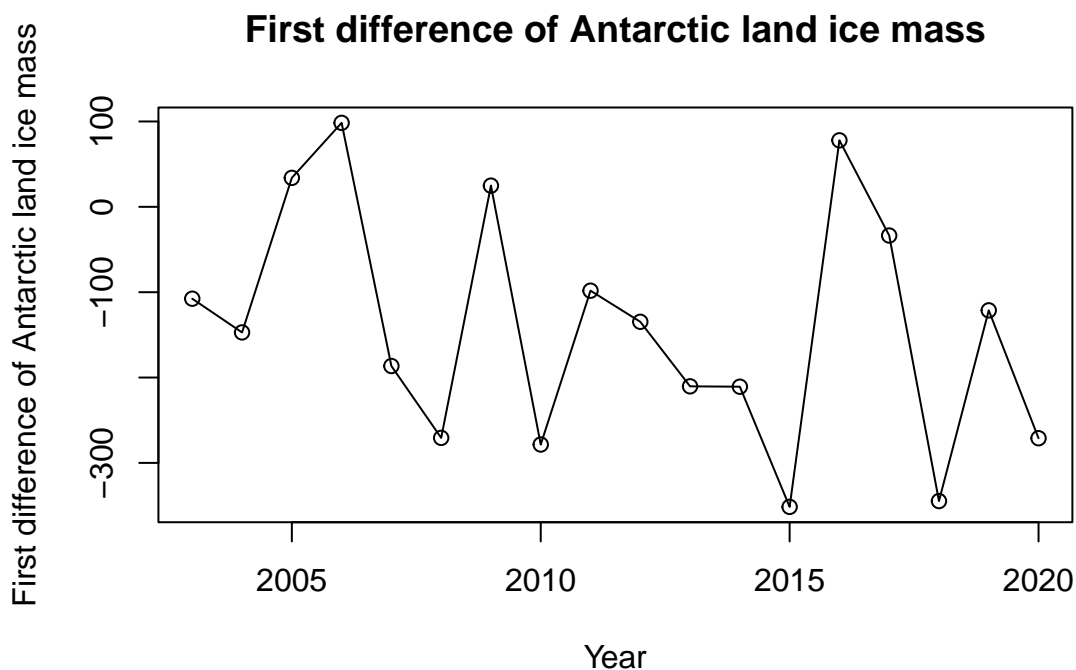


Figure 4: First difference of Antarctic land ice mass

We can still see that there is a small downward trend for the first differenced series. We can confirm whether the series is stationary using the ACF-PACF and the unit root tests.

```
plot_acf_pacf("ACF plot of
Antarctica Land Ice Mass", "PACF plot of
Antarctica Land Ice Mass")
```

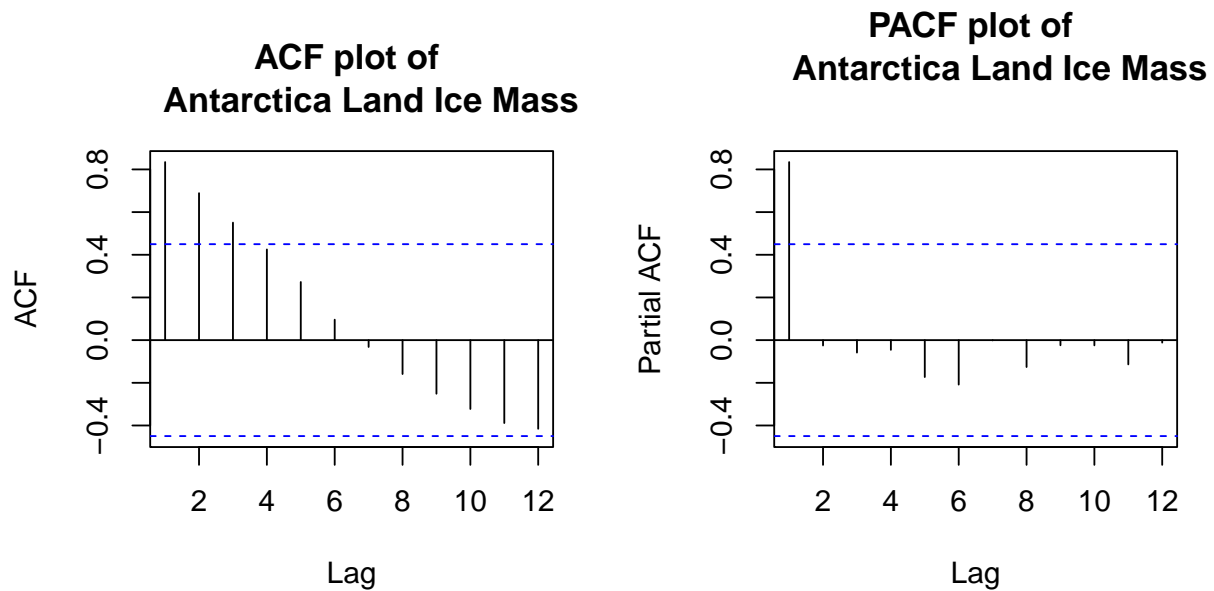


Figure 5: ACF and PACF plots

There is no slowly decaying pattern in ACF plot (Figure 5) and no very large first lag in the PACF plot (Figure 5). This indicates that we are in the right path to stationarity. We have to confirm it with unit root tests. We will perform the Dickey-Fuller test now.

```
adf.test(ice_mass_ts_diff, alternative=c("stationary"))
```

```
##
## Augmented Dickey-Fuller Test
##
## data: ice_mass_ts_diff
## Dickey-Fuller = -2.5664, Lag order = 2, p-value = 0.3566
## alternative hypothesis: stationary
```

The unit test is suggesting that the series is still non-stationary. We will do the other unit test to confirm this.

```
pp.test(ice_mass_ts_diff)
```

```
##
## Phillips-Perron Unit Root Test
##
## data: ice_mass_ts_diff
## Dickey-Fuller Z(alpha) = -17.99, Truncation lag parameter = 2, p-value
## = 0.04887
## alternative hypothesis: stationary
```

Here we can see that the p-value is very close to 0.05 and coming in the range of 0.03-0.1 and hence the test cannot be considered that reliable. We will also perform a KPSS test for testing.

```
kpss.test(ice_mass_ts_diff)
```

```
##
## KPSS Test for Level Stationarity
##
## data: ice_mass_ts_diff
## KPSS Level = 0.30558, Truncation lag parameter = 2, p-value = 0.1
```

For the KPSS test as the hypothesis is inverse, it is suggesting that the difference series is stationary. But, the p-value 0.1 is at the borderline. The value shown by ADF test is higher 0.3566. So, we cannot make a conclusion here. For the ADF test we can try changing the lag order and see the p-values for the neighboring lag orders.

```
adf.test(ice_mass_ts_diff, alternative=c("stationary"),k = 1)
```

```
##
## Augmented Dickey-Fuller Test
##
## data: ice_mass_ts_diff
## Dickey-Fuller = -4.2788, Lag order = 1, p-value = 0.01353
## alternative hypothesis: stationary
```

```
adf.test(ice_mass_ts_diff, alternative=c("stationary"),k = 3)
```

```
##
## Augmented Dickey-Fuller Test
##
## data: ice_mass_ts_diff
## Dickey-Fuller = -2.9425, Lag order = 3, p-value = 0.2133
## alternative hypothesis: stationary
```

Here, we can see that for the lag order 3 p-value suggests that series is non-stationary and lag order 1 suggests that series is stationary. So, again we have a problem. We can put our findings as below.

- ACF-PACF suggests that the series might be stationary.
- The difference plot is going slightly downwards and suggests that the series is still non-stationary
- ADF test suggests strongly that the difference series is non-stationary with one neighbour suggesting it can be stationary.
- pp test and kpss test suggests that the series is stationary. But, the p-values are at the borderline.

**Overall, we can see that this is pointing that the series is non-stationary and we have to consider the second differencing.** Before we proceed, we can do one more thing. We can apply the transformation and see if any new information can be obtained.

We will apply the BoxCox transformation for the series. But, as we mentioned before we have to convert the values to positive else the transformation will throw us an error.

As the minimum value is -2539.055, we will add the scalar value 3000 so that all the values are positive.

```
ice_mass_with_scalar <- ice_mass_ts + 3000
summary(ice_mass_with_scalar)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  460.9  1214.9  2061.0  1936.2  2710.6  2992.5
```

All the values are positive now. We can go ahead with the transformation.

```
BC = BoxCox.ar(ice_mass_with_scalar, main)
```

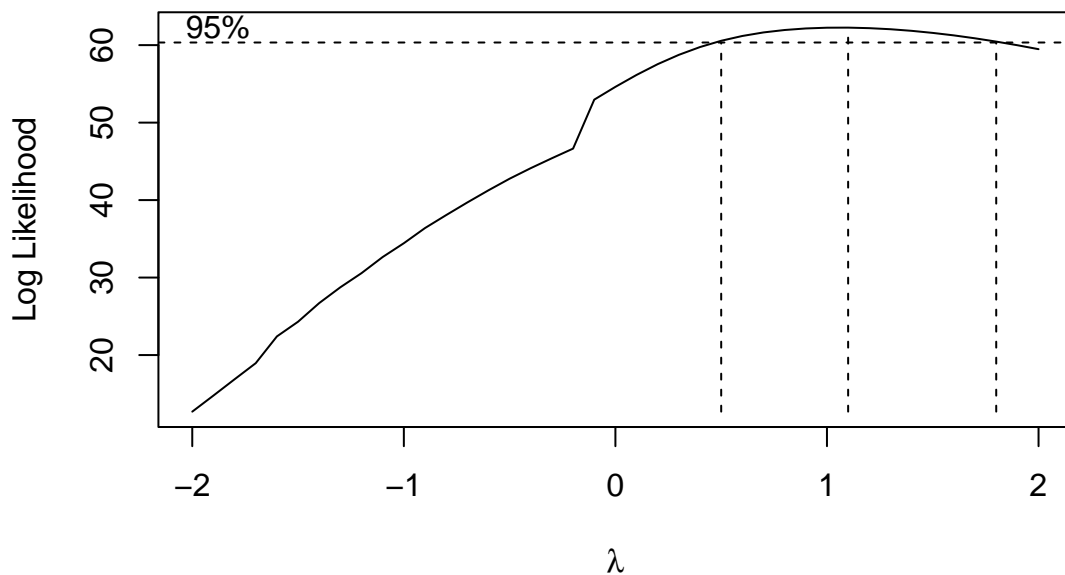


Figure 6: BoxCox Transformation Fitting

```
# Confidence interval limits
BC$ci
```

```
## [1] 0.5 1.8
```

```
# Calculating the lambda
lambda <- BC$lambda[which(max(BC$loglike) == BC$loglike)]
lambda
```

```
## [1] 1.1
```



```
ice_mass_transformed = (ice_mass_with_scalar^lambda-1)/lambda
# Plotting the transformed series
plot(ice_mass_transformed, ylab = "Transformed Antarctic land ice mass", xlab = "Year",
     main = "Transformed Antarctic land ice mass", type="o")
```

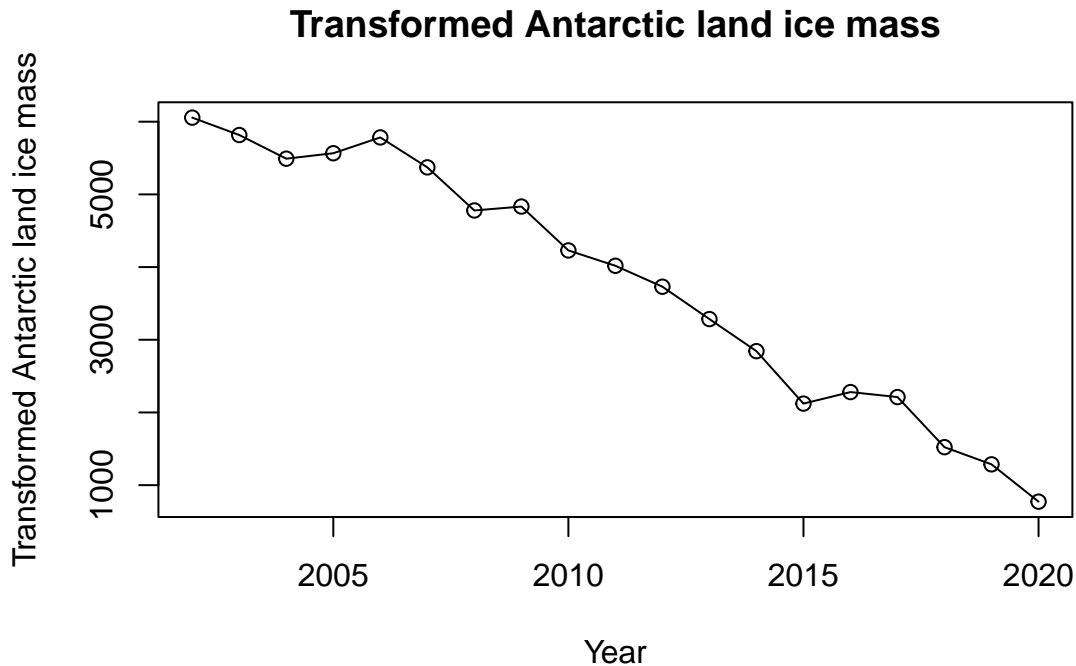


Figure 7: BoxCox Transformation Fitting

Series is remaining almost with the same variations (Figure 7). We will see once again if the transformation has done anything to the normality of the data.

```
qqnorm(ice_mass_transformed)
qqline(ice_mass_transformed, col = 2)
```

```
shapiro.test(ice_mass_transformed)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  ice_mass_transformed
## W = 0.92658, p-value = 0.1498
```

We can see there is no change in the normality of the data. Now, we will directly go with the first differencing and see the plot if the log transformed first difference plot is stationary

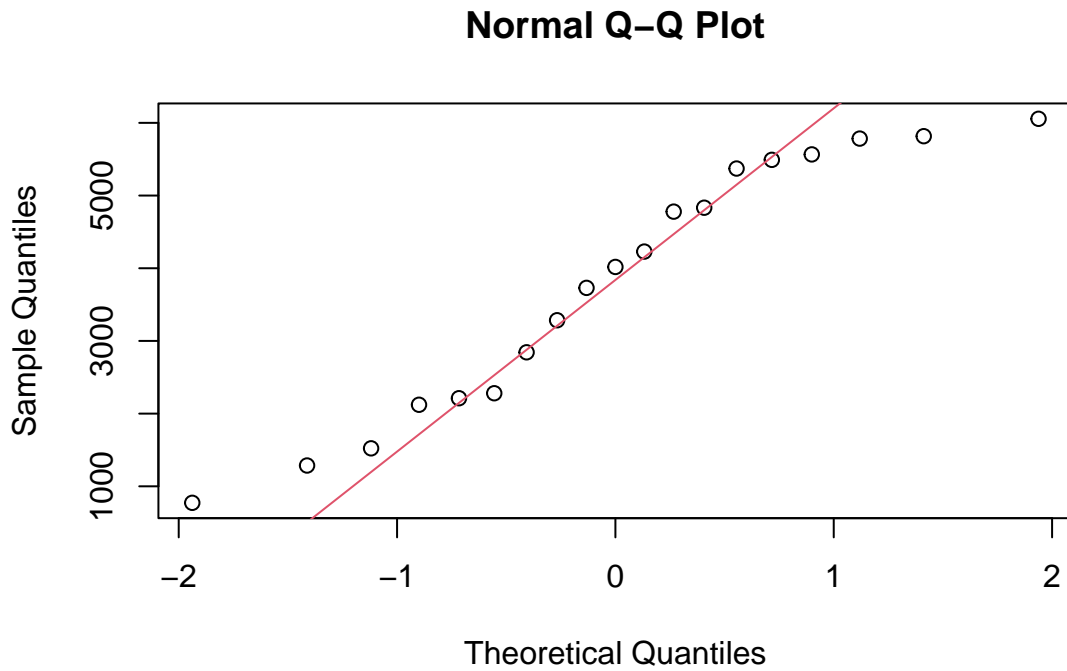


Figure 8: Q-Q Plot

```
ice_mass_transformed_diff <- diff(ice_mass_transformed)
plot(ice_mass_transformed_diff, ylab = "First difference of
    Transformed Antarctic land ice mass", xlab = "Year",
    main = "First difference of Transformed Antarctic land ice mass", type="o")
```

Now, we can see that there is slight downward trend and then picking up from 2015. We will directly conduct the ADF, pp and kpss test now.

```
adf.test(ice_mass_transformed_diff, alternative=c("stationary"))
```

```
##
## Augmented Dickey-Fuller Test
##
## data: ice_mass_transformed_diff
## Dickey-Fuller = -2.5047, Lag order = 2, p-value = 0.3801
## alternative hypothesis: stationary
```

```
pp.test(ice_mass_transformed_diff)
```

```
##
## Phillips-Perron Unit Root Test
##
## data: ice_mass_transformed_diff
## Dickey-Fuller Z(alpha) = -17.88, Truncation lag parameter = 2, p-value
```

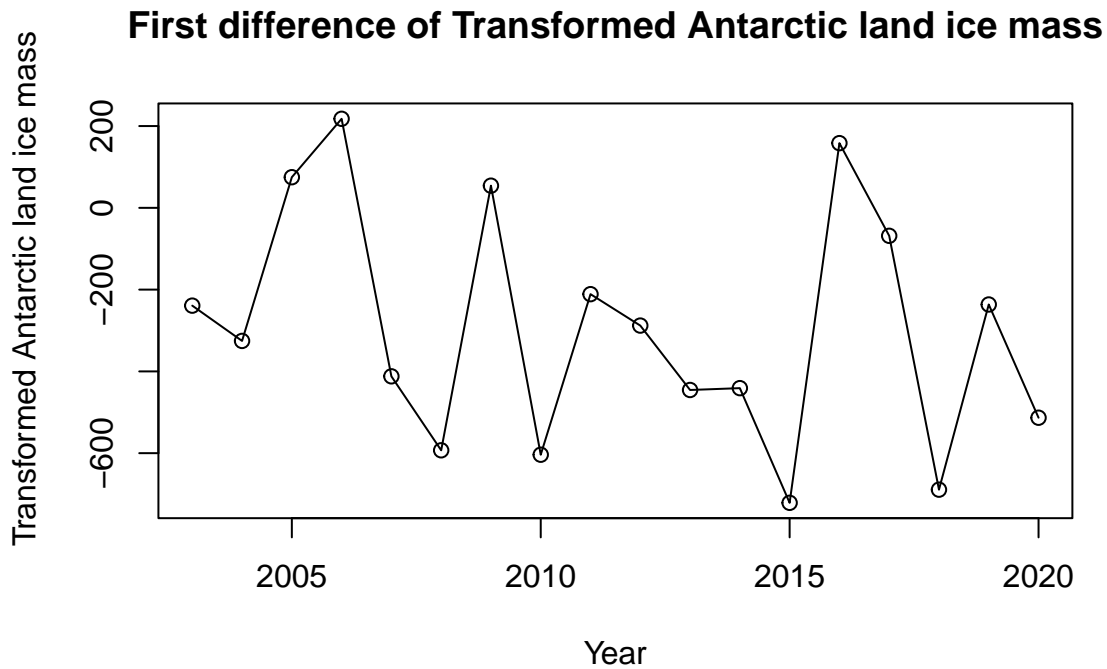


Figure 9: First difference of Transformed Antarctic land ice mass

```
## = 0.05044
## alternative hypothesis: stationary
```

```
kpss.test(ice_mass_transformed_diff)
```

```
##
## KPSS Test for Level Stationarity
##
## data: ice_mass_transformed_diff
## KPSS Level = 0.26015, Truncation lag parameter = 2, p-value = 0.1
```

- ADF test tells the series is non-stationary.
- PP test suggests now that the the series is non-stationary. But the value is borderline.
- KPSS test suggests that series is stationary. But the value is borderline.

Again, **if we weigh both stationary and non-stationary we can say that overall suggestion is the series is non-stationary.** Hence, we will go with the second differencing for the time series. As we have done the transformation we can go with the transformed series itself further since it doesn't matter whether we are going with the raw series or the transformed series.

```
ice_mass_transformed_diff2 <- diff(ice_mass_transformed, differences = 2)
plot(ice_mass_transformed_diff2,
     ylab = "Second difference of Transformed Antarctic land ice mass",
     xlab = "Year", main = "Second difference of Transformed Antarctic land ice mass",
     type="o")
```

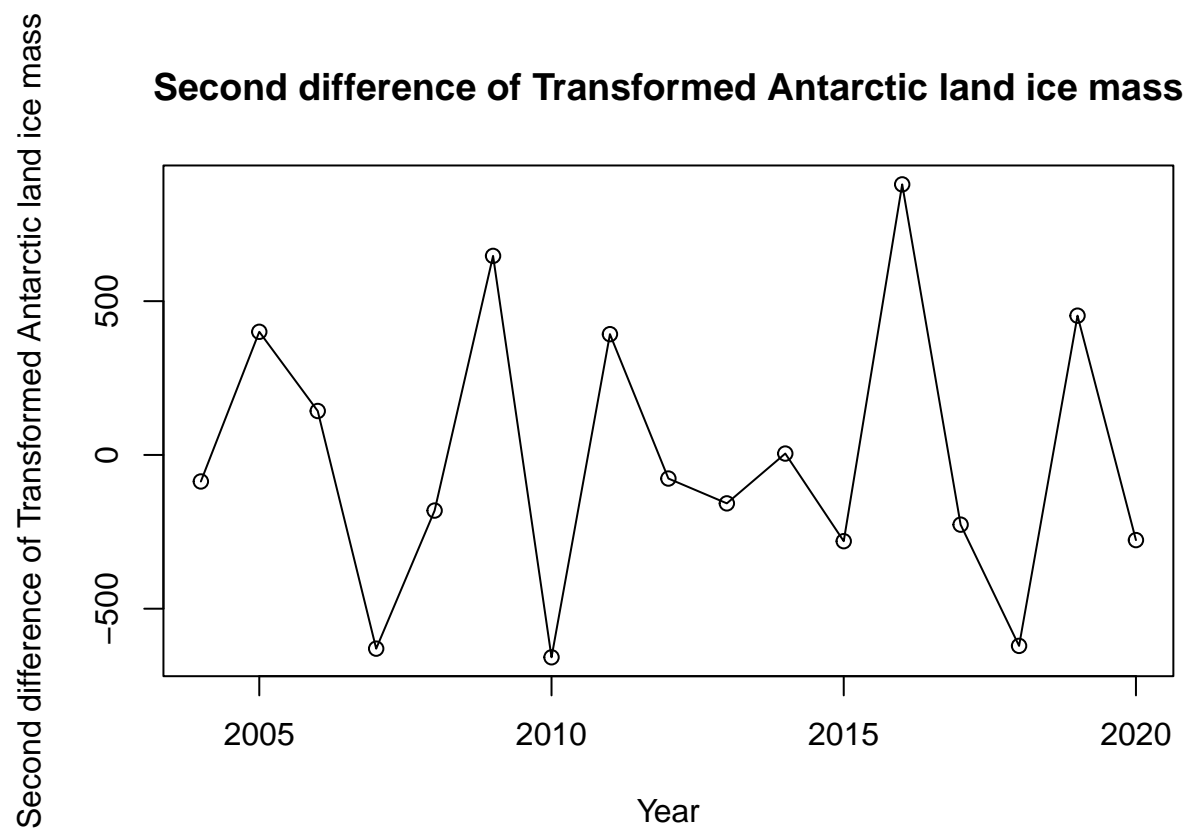


Figure 10: Second difference of Transformed Antarctic land ice mass

Now we could see that the series appears to be stationary (Figure 10). We will plot the ACF and PACF plots and do the ADF test to confirm the stationarity.

```
plot_acf_pacf("ACF plot of Antarctica Land Ice
Mass (Second Difference)", "PACF plot of Antarctica Land Ice
Mass (Second Difference)")
```

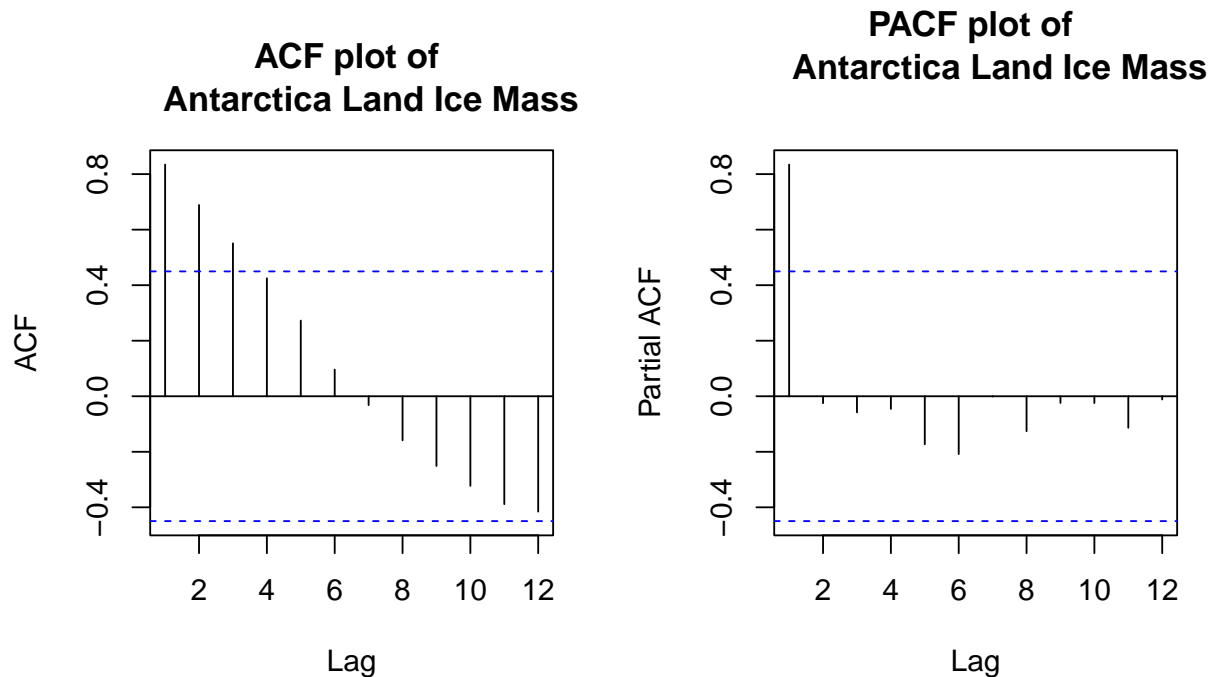


Figure 11: ACF and PACF plots

```
adf.test(ice_mass_transformed_diff2, alternative=c("stationary"))
```

```
##
## Augmented Dickey-Fuller Test
##
## data: ice_mass_transformed_diff2
## Dickey-Fuller = -3.5722, Lag order = 2, p-value = 0.05386
## alternative hypothesis: stationary
```

The ACF plot (Figure 11) is not having decaying pattern and PACF plot (Figure 11) doesn't have very high first lag. So, the series may be stationary. But the result of ADF is concerning for us since the value is above 0.05. We have to try the pp test now.

```
pp.test(ice_mass_transformed_diff2)
```

```
##
## Phillips-Perron Unit Root Test
##
## data: ice_mass_transformed_diff2
```

```
## Dickey-Fuller Z(alpha) = -19.719, Truncation lag parameter = 2, p-value
## = 0.02727
## alternative hypothesis: stationary
```

The pp test confirms that we are having a stationary series. We will perform the KPSS test also to confirm.

```
kpss.test(ice_mass_transformed_diff2)
```

```
##
## KPSS Test for Level Stationarity
##
## data: ice_mass_transformed_diff2
## KPSS Level = 0.11081, Truncation lag parameter = 2, p-value = 0.1
```

KPSS suggests the series is stationary. But, we are having the borderline value. Overall we have the following.

- ACF-PACF and series plot suggests the series is stationary
- ADF test suggests that series is non-stationary. Reason might be we are having very few data points. And the p-value is at borderline
- PP test clearly suggests series is stationary.
- KPSS test says it is stationary but value within borderline

Hence, now we can proceed considering the series as stationary.

## Model Specification

As we have the stationary series available now, we can start with the model specification now.

1. Models based on ACF-PACF plots

```
plot_acf_pacf("ACF plot of Antarctica
Land Ice Mass", "PACF plot of Antarctica
Land Ice Mass")
```

From the ACF plot (Figure 12),  $q = 0$  or  $1$  (Since the significance is less) From the PACF plot (Figure 12),  $p = 1$  or  $2$  (Since one lag is significant and the other lag is little significant) So, the models suggested by ACF-PACF plots are:

- ARIMA(1,2,0)
- ARIMA(1,2,1)
- ARIMA(2,2,0)
- ARIMA(2,2,1)

2. Models based one Extended auto correlation function (EACF)

```
eacf(ice_mass_transformed_diff2, ar.max = 3, ma.max = 3)
```

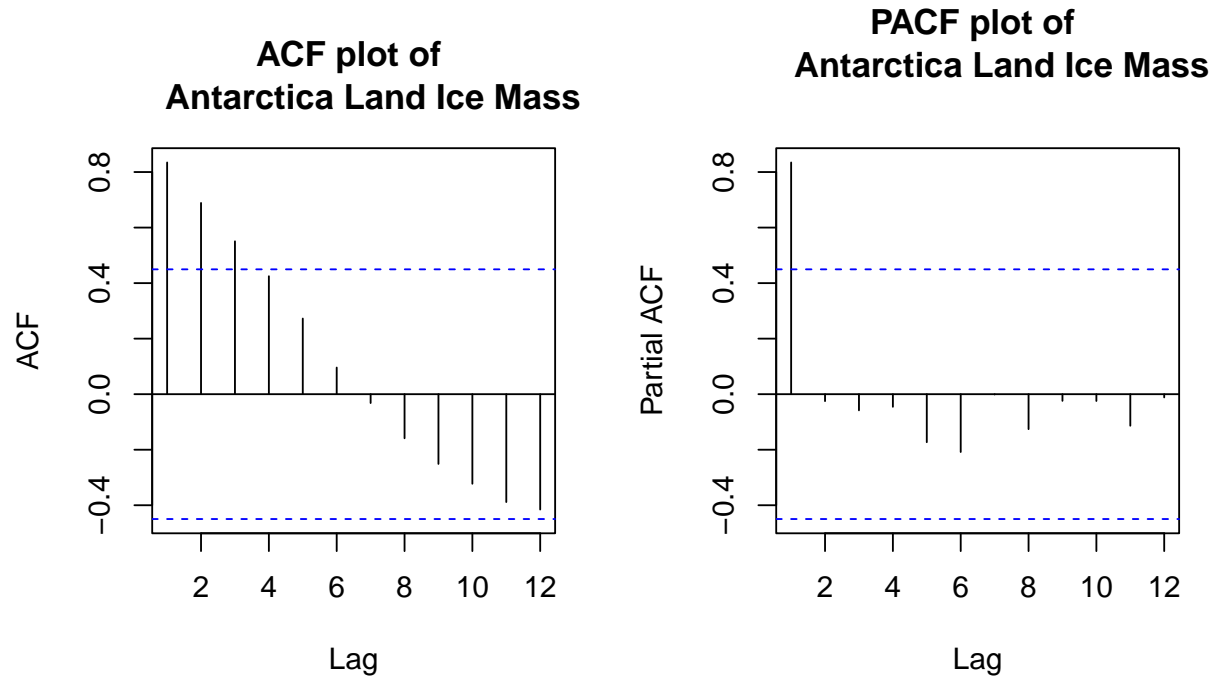


Figure 12: ACF and PACF plots

```
## AR/MA
##   0 1 2 3
## 0 o o o o
## 1 x x o o
## 2 o o o o
## 3 o o o o
```

As there are fewer data points, we can have a maximum level of AR and MA set as 3. The top left most 0 is at location (0,0). The neighbour model is (0,1). As we are concentrating on ARIMA models we are not considering (0,0) as the AR and MA degree is 0 there. Hence, from the EACF we can find that the suggested ARIMA model is:

- ARIMA(0,2,1)

### 3. Models based on Bayesian Information Criterion (BIC) table

```
bic_table = armasubsets(y=ice_mass_transformed_diff2 , nar=5 , nma=5, y.name='p',
                        ar.method='ols')
plot(bic_table)
```

Because of the fewer data points for the BIC table also we won't be getting any large models. From the above BIC table, the models suggested are:

- ARIMA(3,2,1)

We have a second best model with MA order of 4 and AR order 0. But we are not going to consider that since the shade is too low and from the y axis we can see that the value is about 1/3rd of the best model.

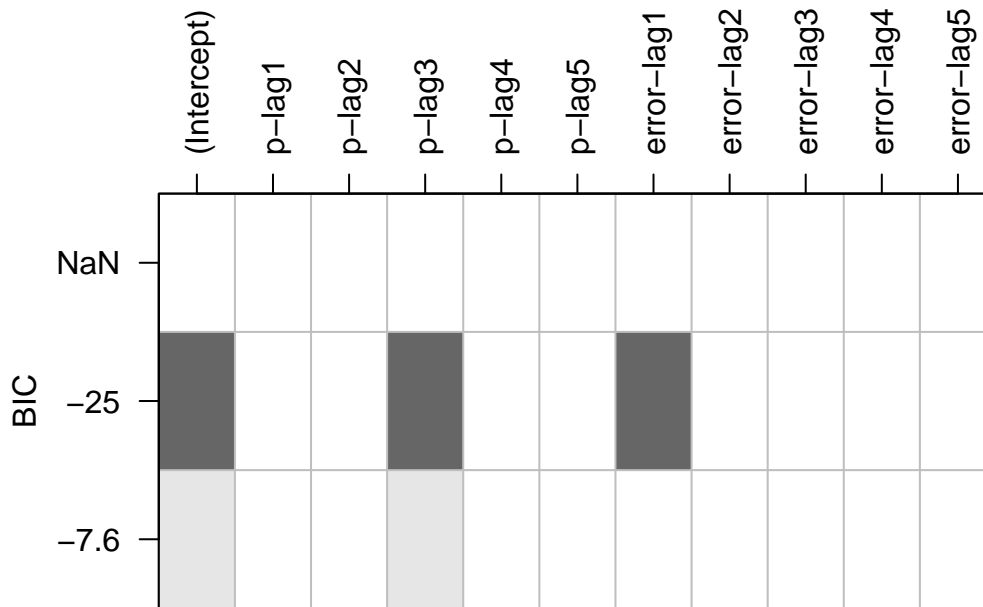


Figure 13: BIC Table

## Results

We have conducted a thorough analysis of the time series given and found out the suggested models from the different model specification tools as below:

- ARIMA(1,2,0)
- ARIMA(1,2,1)
- ARIMA(2,2,0)
- ARIMA(2,2,1)
- ARIMA(0,2,1)
- ARIMA(3,2,1)

## Conclusion

- Stationarity is one of the major checkpoints for model specification.
- We cannot rule out stationarity based on a single test or plot. Multiple tests has to be conducted for this as none of the tests available are perfect Also, we have to weigh stationarity and non-stationarity based on the results and the decision taken should be justifiable.
- It is the person's perspective and diagnostics as to which model we have to consider frn model fitting.
- The same data set given to a different individual can result in some extra or fewer ARIMA models. But most of the models will be matching.
- Over differencing can introduce unnecessary correlations and problems in parameter estimation. The principle of parsimony has to be followed always.



## References

- Haydar Demirhan, 2022, ‘Week 1 - 5 materials’, viewed 10th to 23rd March 2022, file:///C:/RMIT/Semester%201%202022/Time%20Series/Module%201/Module%201%20-%20Online%20Notes.html
- Jonathan D. Cryer, Kung-Sik Chan 2008, *Time Series Analysis With Applications in R*, Springer, USA