MATH1318 TIME SERIES

Semester 1, 2020

Forecasting International Arrivals in New Zealand

Project By

Millie Woo (s3806940)

lanka Wijeratne (s3808496)

ishal Beniwal (s3759790)

afeed Sultaan (s3763175)

nup Sakpal (s3801788)

Table of Content

1. Introduction 4
2. Summary Statistics 4
3. Data Analysis on the original time series data 4
4. Model Specification of Seasonal part: 6
Step 1: Seasonal differencing 6
Step 2: Ordinal differencing 7
Step 3: Specify the seasonal order SMA (1) 8
5. Model Specification of Ordinal part: 9
Step 4: Check the EACF for ordinal orders 9
Step 5: Model Diagnostics with Coefficient Analysis 10
Step 6: Model selection 12
6. Forecasting: Predicting for next 10 months 12
7. Conclusion 13
8. Reference 15
[R1] 15
9. Appendix 15
[A1] 15
[A2] 16
[A3] 16
[A4] 17
[A5] 18
[A6] 19
[A7] 19
[A8] 21

Table of Tables

Table 1-Summay statistics of International Arrivals in NZ between 2000 and 2012 4
Table 2 - EACF table for ordinal model specification 9
Table 3 - Coefficient Test result on SARIMA (0,1,1) X (0,1,1)12 11
Table 4 - Coefficient Test result on SARIMA (0,1,2) X (0,1,1)12 11
Table 5 - Coefficient Test result on SARIMA (1,1,2) X (0,1,1)12 12
Table 6 - AIC and BIC scores of all the models in ascending order 12

Table 7- Confidence Intervals for forecasting International Arrival (using SARIMA(0,1,1)x(0,1,1)12) for Feb 2012 to Nov 2012 13

Table of Figures

Figure 1- Time series plot of International Arrivals in NZ from 2000 to 2012 5
Figure 2- Scatter plot of International Arrivals in NZ with first time (previous month) lag ... 5
Figure 3- Q-Q plot of International Arrivals in NZ 5
Figure 4 - PACF Plot of International Arrivals in NZ 6
Figure 5 - ACF Plot of International Arrivals in NZ 6
Figure 6 - Plots of Residual Analysis for SARIMA (0,0,0) X (0,1,0)12 7
Figure 7 - Plots of Residual Analysis for SARIMA (0,1,0) X (0,1,0)12 7
Figure 8 - Plots of Residual Analysis for SARIMA (0,1,0) X (0,1,1)12 9
Figure 9 - Plots of Residual Analysis for SARIMA (0,1,1) X (0,1,1)12 and SARIMA (0,1,2) X (0,1,1)12 10
Figure 10 - Forecasts of International Arrivals in NZ using SARIMA (0,1,1) X (0,1,1)12 for next 10 months 13

1. Introduction

We have chosen to use the "New Zealand Air Passengers" dataset which contains data of passenger arrivals and departures in New Zealand from the year 2000 to 2012. (refer to Reference [R1] for the original source of the dataset). Our main objective of this project is to forecast the arrivals in New Zealand for the next 10 months i.e. from Feb 2012 to Nov 2012 by implementing SARIMA modeling.

2. Summary Statistics

From the descriptive statistics of the data set, we were able to analyze key features like the means of passenger arrivals was about 343 thousand passengers. The minimum number of passengers arrived is about 200 thousand which was in the month of June 2006 and the maximum number of passengers arrived was about 499 thousand in January 2012 with passengers arriving in New Zealand over a period of 145 months.

Refer to Appendix [A1] and [A2] for coding.

Table 1-Summay statistics of International Arrivals in NZ between 2000 and 2012

	Mean	Median	Std.Dev	Q1	Q3	IQR	Min	Max	N.Valid
nzArrivals	343309.8	346434	66558.46	295485	390574	95089	200423	499839	145

3. Data Analysis on the original time series data

From the time series plot in Figure 1, we can see that there is a positive increasing trend from the year 2000 to 2012. The mean of the series is increasing over the years. The regular peaks in the month of January and drops in the month towards May suggest there is seasonality in the trend. The slowly decreasing wave-like pattern in the ACF (Figure 4) suggests that there is non- stationarity present in the time series. From the QQ plot shown in Figure 3, we can observe that the data is normally distributed i.e. plot captures almost all the data points.

Refer to Appendix [A3] for coding.

By the end of the data analysis process on the original time series data, we can conclude the following properties in the time series that we are dealing with:

- **Trend**: An obvious upward trend can be seen from the time series plot.
- **Cyclicality or Seasonality**: An obvious seasonal pattern in the graph.
- **Fluctuation**: There are consistent fluctuations in the time series data.
- **Intervention point**: There is no obvious intervention point that cause any sudden pattern changes.
- The nature of the arrivals in NZ cannot be determined by a function of time, thus we would use the **residual approach** to model this seasonal stochastic data.

- **Changing Variance**: There is no significant changes in variance along the line

Figure 1- Time series plot of International Arrivals in NZ from 2000 to 2012

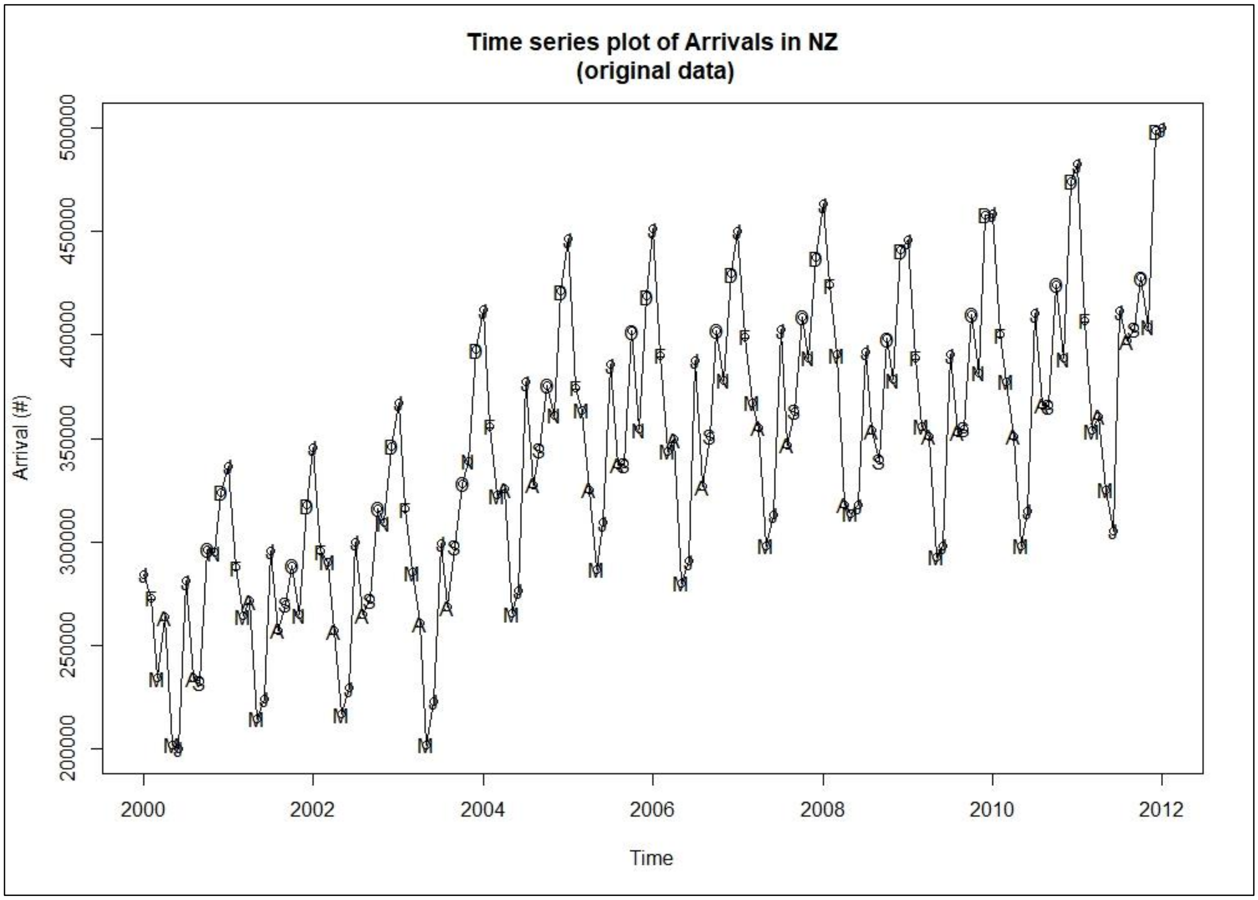


Figure 2- Scatter plot of International Arrivals in NZ with first time (previous month) lag

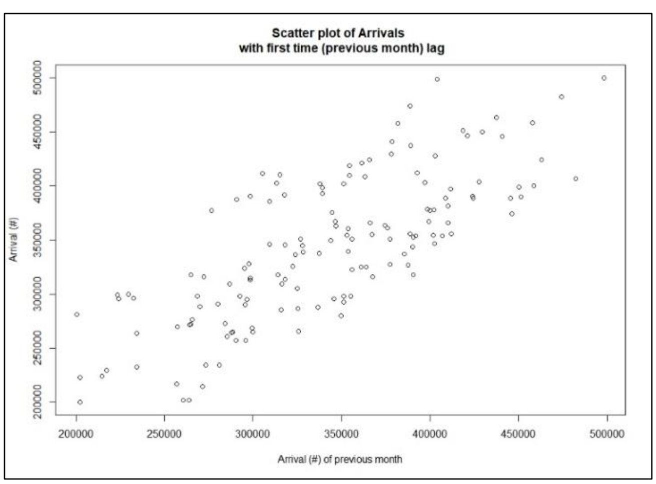


Figure 3- Q-Q plot of International Arrivals in NZ

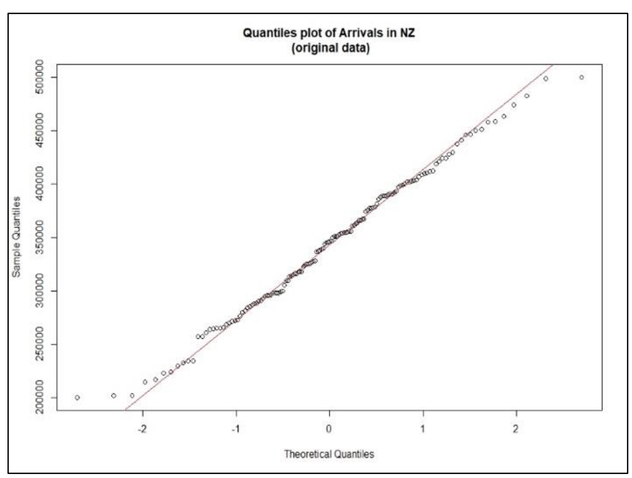


Figure 5 - ACF Plot of International Arrivals in NZ

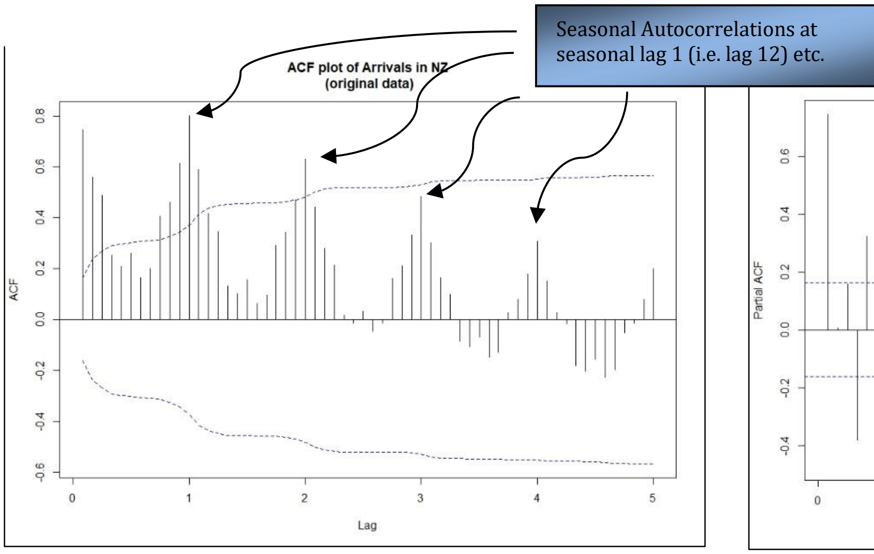
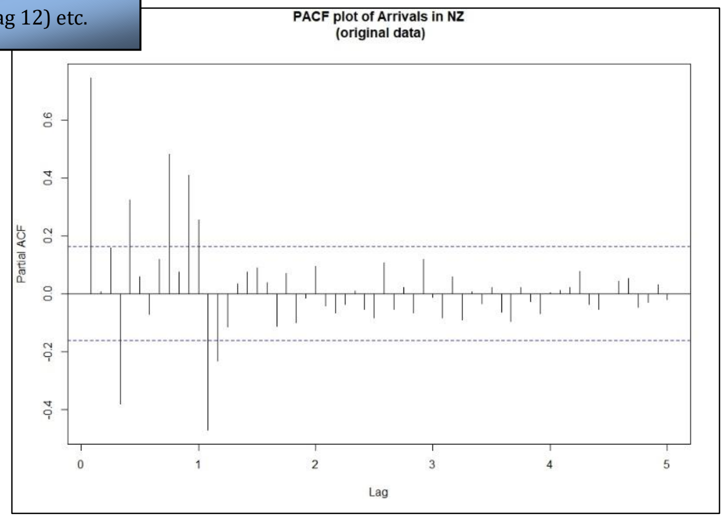


Figure 4 - PACF Plot of International Arrivals in NZ



4. Model Specification of Seasonal part:

To deal with the existence of seasonal autocorrelation, we took the residual approach of model specification by fitting a plain model and we examined the time series, ACF, PACF plots of the residual.

Step 1: Seasonal differencing

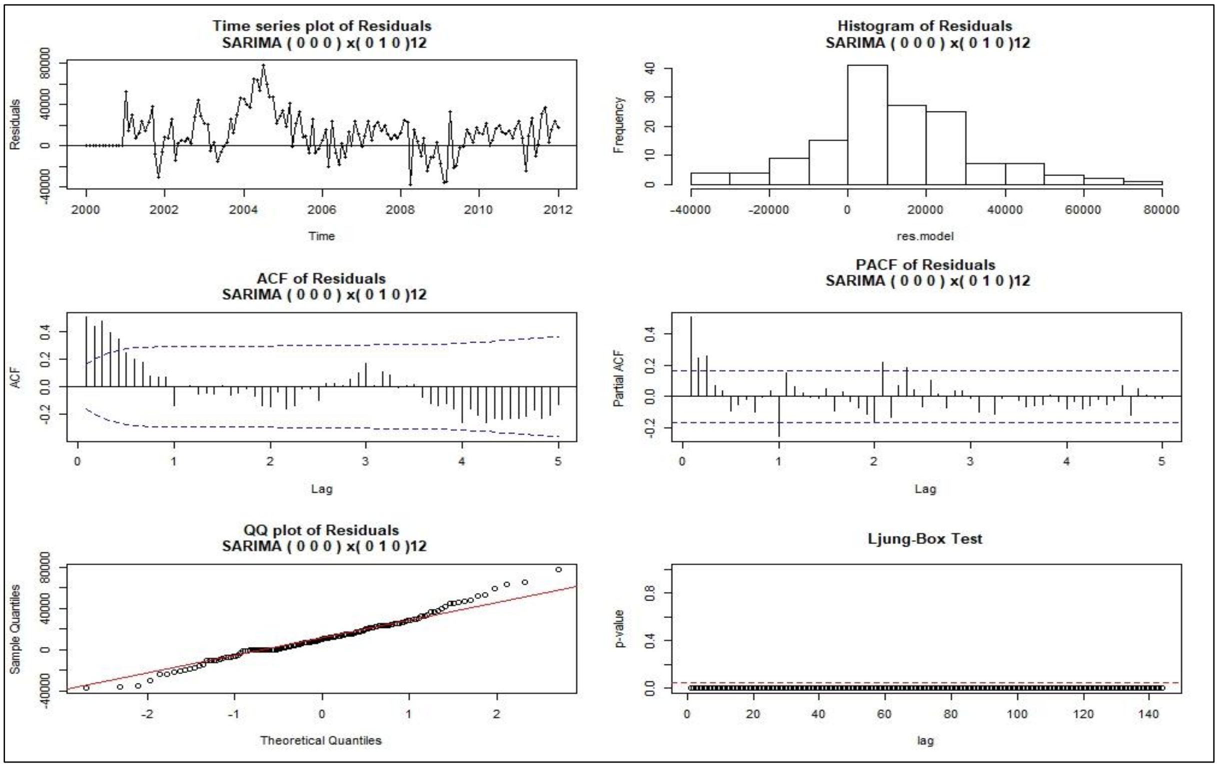
We first applied the seasonal differencing to the seasonal part of the model (D =1) and then we examined the autocorrelation structure of the residuals using time series, ACF, PACF plots of the residuals. From Figure 6, we can observe:

- Although the general upward trend is resolved, however, the time series plot of residual still has a sudden change in variance during the year 2004. The rest of the series seems randomly distributed around the zero-mean showing no sign of seasonality.

- The seasonal autocorrelation is absent now from, for example, seasonal lag 1 (lag 12) but the ACF plot still shows a steady decaying pattern which suggests the non-stationarity in the trend.

Refer to Appendix [A4] for coding.

Figure 6 - Plots of Residual Analysis for SARIMA (0,0,0) X (0,1,0)12



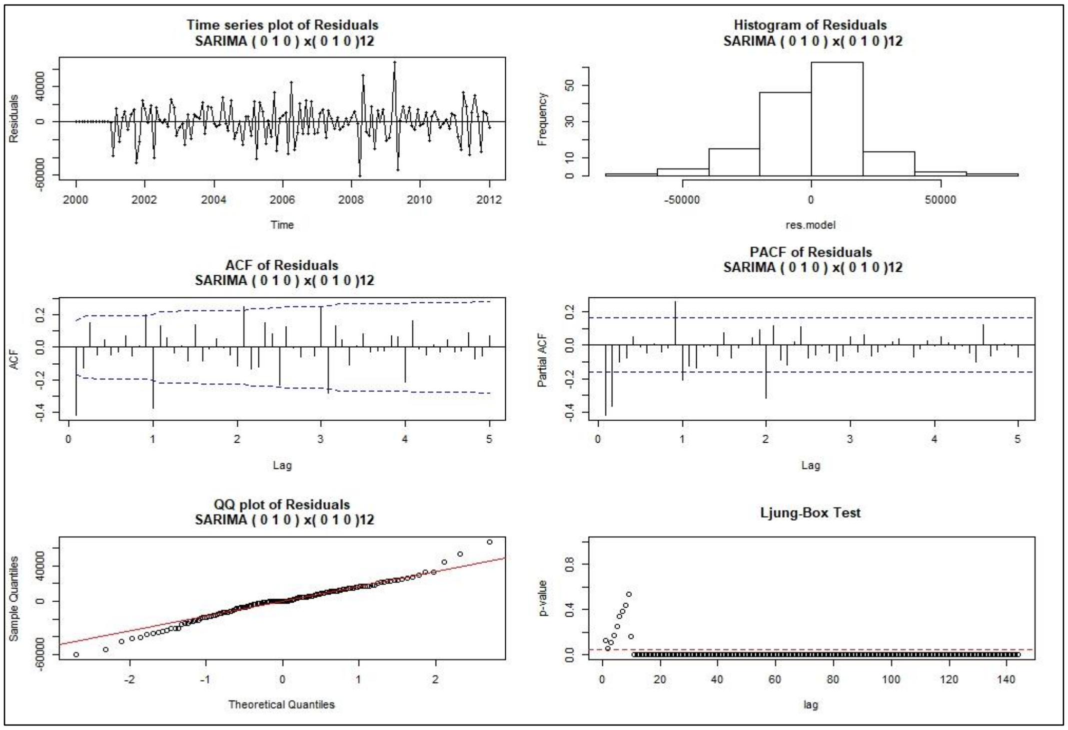
Step 2: Ordinal differencing

To tackle the problem of non-stationarity in the series, we applied ordinal differencing to get rid of the ordinary trend. Figure 7 shows the results obtained after applying first ordinal differencing where we can observe:

- The ACF of residuals have one significant autocorrelation at seasonal lag 1 (lag 12 in the series, so we can consider the SMA (1) model for the seasonal part, i.e. Q =1 and see if we can get rid of the effect of the seasonal component in the residuals.

Refer to Appendix [A5] for coding.

Figure 7 - Plots of Residual Analysis for SARIMA (0,1,0) X (0,1,0)12

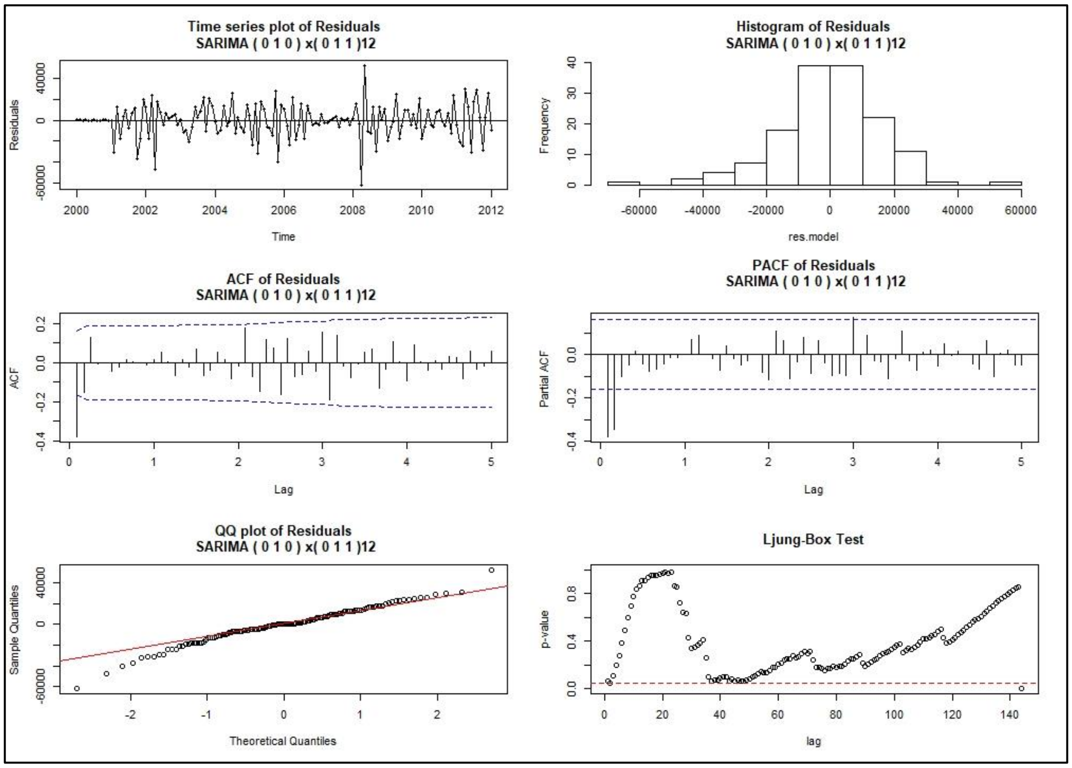


Step 3: Specify the seasonal order SMA (1)

We added the SMA (1) component to the model and checked the TS, ACF and PACF plots again. From the ACF and PACF plots shown in Figure 8, we can observe that there is no significant autocorrelation left at any of the seasonal lags which completes the process of model specification.

We can conclude that for the order Q =1 (SARIMA (0,1,0) X $(0,1,1)_{12}$), we get white noise residuals and therefore, we will move forward to ordinal model fitting and diagnostic checking.

Figure 8 - Plots of Residual Analysis for SARIMA (0,1,0) X (0,1,1)12



5. Model Specification of Ordinal part:

Step 4: Check the EACF for ordinal orders

Refer to Appendix [A6] for coding.

We used EACF on the residuals of the last stage to check the information about AR (p) and MA (q) components left in the residuals. From the top-left vertex of the EACF table (Table 2), our candidates for ARMA part came out as ARMA (0,1), ARMA (0,2) and ARMA (1,2).

Table 2 - EACF table for ordinal model specification

```
## AR/MA
##   0 1 2 3 4 5
## 0 x o o o o o
## 1 x x o o o o
## 2 x o o o o o
## 3 x o o o o o
## 4 x o o o o o
## 5 x o x o o o
```

Hence, the tentative models were specified as:

- *SARIMA(0,1,1)x(0,1,1)*$_{12}$ *by EACF*

- *SARIMA(0,1,2)x(0,1,1)$_{12}$ by EACF*
- *SARIMA(1,1,2)x(0,1,1)$_{12}$ by EACF*

Step 5: Model Diagnostics with Coefficient Analysis

We conducted the Coefficient Analysis to analyze the tentative models and choose the best model for forecasting. Models were first checked for the normality of the residuals in addition to being white noise and then, was checked whether the model was adequate. For this, models were compared with over-fitted models.

Figure 9 - Plots of Residual Analysis for SARIMA (0,1,1) X (0,1,1)12 and SARIMA (0,1,2) X (0,1,1)12

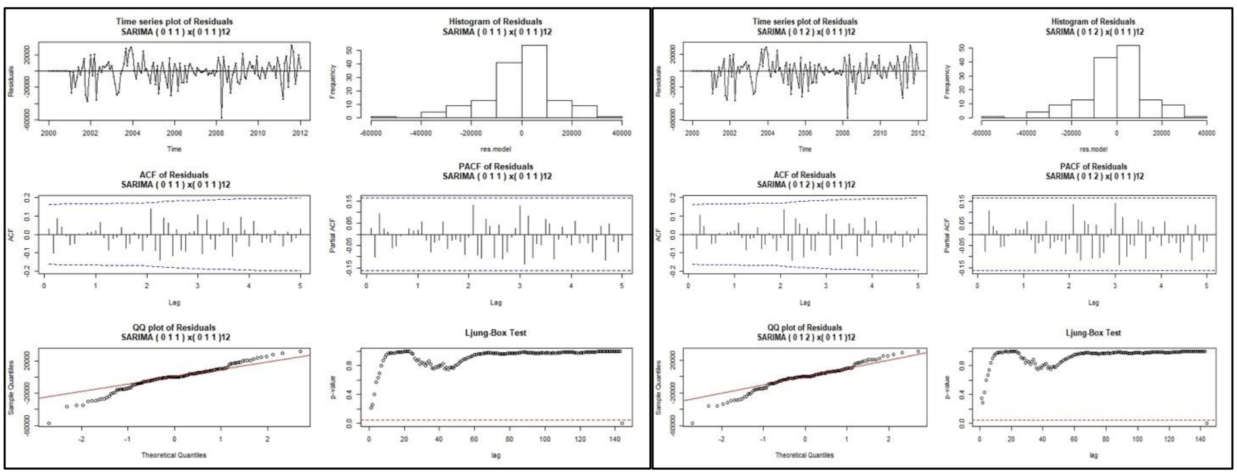
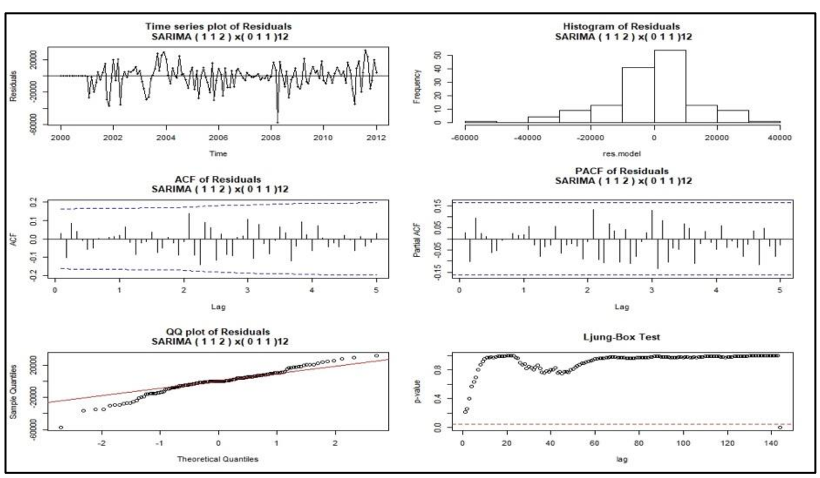


Figure 10 - Plots of Residual Analysis for SARIMA (1,1,2) X (0,1,1)12



From Figures 9 and 10, we can observe that for all the 3 candidate models, the time series plot suggests that there is no more seasonality presence in the residuals and most importantly, the histogram plot suggests that the residuals are normally distributed, tailing at both ends from the Centre. In addition, the QQ plot also demonstrates the residuals are normally distributed as they captured by the QQ line. Ljung-Box test confirms that there are no signs of autocorrelation left in the series, which brings us to a realization that all three model's residuals are normal in addition to being white noise and can be further analyzed to know the adequate model.

Refer to Appendix [A7] for coding.

From Table 3, the Coefficient Test result on SARIMA (0,1,1) X $(0,1,1)_{12}$ showed that both the coefficient (MA (1) & SMA (1)) were statistically significant at 5% level of significance.

Table 3 - Coefficient Test result on SARIMA (0,1,1) X $(0,1,1)_{12}$

```
##
## z test of coefficients:
##
##       Estimate Std. Error z value  Pr(>|z|)
## ma1  -0.621765   0.072064 -8.6280 < 2.2e-16 ***
## sma1 -0.633371   0.091654 -6.9105 4.831e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Reject $\theta_1 = 0$

Reject $\Theta_1 = 0$

From Table 4, the Coefficient Test result on SARIMA (0,1,2) X $(0,1,1)_{12}$ showed that MA (2) component was not significant, where MA (2) coefficient could be seen as an additional MA component added into our first model to check for overfitting.

Table 4 - Coefficient Test result on SARIMA (0,1,2) X $(0,1,1)_{12}$

```
##
## z test of coefficients:
##
##       Estimate Std. Error z value  Pr(>|z|)
## ma1  -0.587811   0.098442 -5.9711 2.356e-09 ***
## ma2  -0.050727   0.104077 -0.4874     0.626
## sma1 -0.639408   0.091376 -6.9975 2.605e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fail to Reject $\theta_2 = 0$
⇨ Overfitting

From Table 5, Coefficient Test result on SARIMA (1,1,2) X $(0,1,1)_{12}$ showed that both AR (1) and MA (1) component was not statistically significant at 5% level of significance.

Table 5 - Coefficient Test result on SARIMA (1,1,2) X (0,1,1)12

```
##
## z test of coefficients:
##
##      Estimate Std. Error z value  Pr(>|z|)
## ar1  -0.68758    0.38224 -1.7988   0.07204
## ma1   0.12273    0.36280  0.3383   0.73514
## ma2  -0.50350    0.20813 -2.4191   0.01556 *
## sma1 -0.63188    0.09093 -6.9491 3.675e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fail to Reject $\varphi_1 = 0$ ⇨ Overfitting

Fail to Reject $\theta_1 = 0$ ⇨ Overfitting

Step 6: Model selection

We calculated the AIC and BIC score of all the models to model comparison and from Table 6, found SARIMA(0,1,1)x(0,1,1)$_{12}$ with the lowest both AIC and BIC score.

Table 6 - AIC and BIC scores of all the models in ascending order

```
#Sort the AIC and BIC score
sort.score(sc.AIC, score = "aic")

##                     df      AIC
## m3_011.nzArrivals    3 2919.742
## m3_012.nzArrivals    4 2921.506
## m3_112.nzArrivals    5 2922.338

sort.score(sc.BIC, score = "bic")

##                     df      BIC
## m3_011.nzArrivals    3 2928.391
## m3_012.nzArrivals    4 2933.037
## m3_112.nzArrivals    5 2936.752
```

Hence, based on the significance we found for all the coefficients, as well as the lowest AIC and BICs score, we concluded that SARIMA(0,1,1)x(0,1,1)$_{12}$ is the final model to perform forecasting. Moreover, it was also observed that both SARIMA(0,1,2)x(0,1,1)$_{12}$ and SARIMA(1,1,2)x(0,1,1)$_{12}$ will overfit the time series data based on their insignificant results in coefficient analysis and AIC score.

6. Forecasting: Predicting for next 10 months

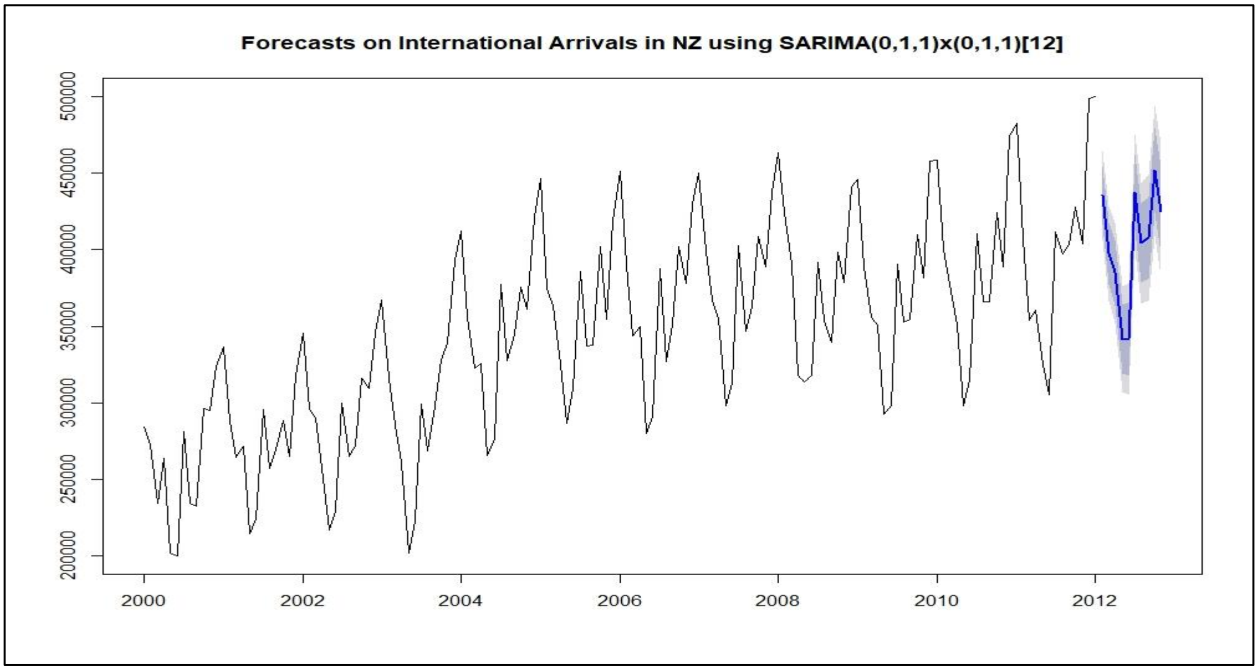
Using the final SARIMA model, forecasting was performed to predict the trend (Arrival of the passenger to New Zealand) for the next 10 months i.e. From Feb 2012 to Nov 2012. The results can be seen from the TS plot in Figure 10 and confidence intervals for forecasting International Arrival in Table 7 below. Using the TS plot, we can suggest that the trend is predicted to follow the same seasonal trend for the next 10 months. The downwards in May is evident in the trend, although the drop in the trend is not as obvious as previous years.

Refer to Appendix [A8] for coding.

Table 7- Confidence Intervals for forecasting International Arrival (using SARIMA(0,1,1)x(0,1,1)12) for Feb 2012 to Nov 2012

##	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
## Feb 2012	435419.3	416489.1	454349.5	406468.0	464370.5
## Mar 2012	397697.3	377458.3	417936.3	366744.4	428650.2
## Apr 2012	385013.0	363544.8	406481.2	352180.2	417845.8
## May 2012	341553.3	318922.6	364184.1	306942.6	376164.1
## Jun 2012	341690.0	317953.6	365426.4	305388.3	377991.7
## Jul 2012	437356.2	412563.4	462149.1	399438.9	475273.6
## Aug 2012	404275.8	378469.8	430081.8	364809.0	443742.7
## Sep 2012	407826.4	381045.5	434607.2	366868.5	448784.2
## Oct 2012	451685.9	423964.4	479407.4	409289.5	494082.3
## Nov 2012	425127.3	396496.1	453758.6	381339.6	468915.0

Figure 10 - Forecasts of International Arrivals in NZ using SARIMA (0,1,1) X (0,1,1)12 for next 10 months



7. Conclusion

After analyzing the NZ international dataset, we have the following findings:

- There is a steady upward trend in this seasonal time series. It has a repetitive pattern every 12 months, and it also has a high correlation with its first-time lag (which is the previous month, r=0.763).
- Data Transformation does not apply to this time series for modeling as the data points are already normally distributed.
- As the nature of international arrivals cannot be predicted by any function of time, we model the time series as a seasonal stochastic trend with the residual approach.

- The possible SARIMA candidate models are SARIMA(0,1,1)x$(0,1,1)_{12}$, SARIMA(0,1,2)x$(0,1,1)_{12}$ and SARIMA(1,1,2)x$(0,1,1)_{12}$
- All coefficients in SARIMA(0,1,1)x$(0,1,1)_{12}$ are significant, the other two models are found to be overfitting the arrival dataset from the coefficent test.
- Residuals of SARIMA(0,1,1)x$(0,1,1)_{12}$ are within 95% confidence to be uncorrelated. As this is a long time series (with 145 data samples), based on the central limit theorem, the residuals would converge to normal distribution.
- With the lowest AIC and BIC score, SARIMA(0,1,1)x$(0,1,1)_{12}$ is the best fit model in our analysis.
- Forecast for the next 10 months extend the seasonal pattern, which would jump after reaching the lowest estimation in May, and then keep increasing until November (which is the last month of predictions), with a slight drop in August and September.

8. Reference

[R1]

Census at school New Zealand. https://new.censusatschool.org.nz/resource/time-series-data-sets-2013/ Accessed on 5-06-2020.

9. Appendix

[A1]

```
#The following packages are needed in this assignment:

library(TSA)

library(forecast)

library(tseries)

library(knitr)

library(fUnitRoots)

library(lmtest)

library(FitAR)

library(summarytools)

#Read in the dataset
nzPassenger <- read.csv("D:/RMIT Master of Analytics/semester 2/MATH1318 - Ti
me Series/Project/NZAirPassenger.csv", header = TRUE)
head(nzPassenger)

##      DATE Arrivals Departures
## 1 2000M01   284361     288701
## 2 2000M02   273092     252533
## 3 2000M03   234368     286140
## 4 2000M04   263813     290177
## 5 2000M05   202172     235108
## 6 2000M06   200423     222173

#covert to a timeseries object.
nzArrivals <- ts(as.vector(nzPassenger$Arrivals), start=2000, end=2012, frequ
ency=12)
class(nzArrivals)

## [1] "ts"
```

[A2]

```
# Put the summary statistics in table format
kable(descr(nzArrivals, stats = c("mean", "med", "sd", "Q1", "Q3","IQR", "min
", "max", "n.valid"), transpose = TRUE), caption = "Summary statistics of Int
ernational Arrivals in NZ between 2000 and 2012")
```

[A3]

```
#Define a function plot.all, which would plot the following graphs of the pas
s-in time series data:
# 1. Generate the time series Plot
# 2. Scatter plot of the data with its first time lag, also show this correla
tion index
# 2. Normality via QQ-plot and Shapiro test
# 3. Generate the ACF and PACF plot
plot.all <- function(ts_data, ts_plot_title, scatter_plot_title, qq_plot_titl
e, acf_title, pacf_title, isDiff=TRUE){

        #Time series plot

        plot(ts_data, type='o', xlab = 'Time', ylab='Arrival (#)', main = ts_
plot_title)
        points(y=ts_data,x=as.vector(time(ts_data)), pch=as.vector(season(ts_
data)))

        if(isDiff == FALSE){

                #Scatter Plot and Check correlation of 1st lagging
                plot(y=ts_data,x=zlag(ts_data),ylab='Arrival (#)', xlab='Arri
val (#) of previous month' , main = scatter_plot_title)

                y = ts_data
                x = zlag(ts_data)        # Generate first lag of the series
                index = 2:length(x)
                print('Correlation Index:')
                print(cor(y[index],x[index]))

                #QQ Plot and check Normality
                qqnorm(ts_data, main=qq_plot_title)
                qqline(ts_data, col = 2)
                print(shapiro.test(ts_data))
        }
        #ACF and PACF plot
```

```
        #par(mfrow=c(1,2))
        acf(ts_data, xaxp=c(0,10,10), lag.max=60, ci.type='ma', main=acf_titl
e)
        pacf(ts_data, xaxp=c(0,10,10), lag.max=60, main=pacf_title)
        par(mfrow=c(1,1))


}




#Plot all the associated graphs for the Original Time series Data
plot.all(nzArrivals, 'Time series plot of Arrivals in NZ\n (original data)',
'Scatter plot of Arrivals \n with first time (previous month)lag', 'Quantiles
plot of Arrivals in NZ\n (original data)', 'ACF plot of Arrivals in NZ\n (ori
ginal data)', 'PACF plot of Arrivals in NZ\n (original data)', isDiff = FALSE
)
```

[A4]

```
#Define a function residual.analysis which would perform the following plots
of the residuals of
# the pass-in arima model
# 1. time series plot of the residuals
# 2, Histogram of the residuals
# 3. ACF plot of the residuals
# 4. PACF plot of the residuals
# 5. Q-Q plot of the residuals
# 6. Ljung-Box plot of the residuals
# 7. Ljung-Box test of the residuals
#
#this function is originated from the residual.analysis function developed by
Yong Kai, Wong
# I just add in the  Ljung-Box test and customized for SARIMA model

residual.analysis <- function(model, p, d, q, P, D, Q){
    res.model = residuals(model)
    par(mfrow=c(3,2))
    arimaOrderStr <- paste("SARIMA (", p, d, q, ") x(", P, D, Q, ")12")
    plot(res.model,type='o',ylab='Residuals', main=paste("Time series plot of
Residuals\n", arimaOrderStr))
    abline(h=0)
    hist(res.model,main=paste("Histogram of Residuals\n", arimaOrderStr))
    acf(res.model, xaxp=c(0,10,10), lag.max=60, ci.type='ma', main=paste("ACF
of Residuals\n",arimaOrderStr))
    pacf(res.model,xaxp=c(0,10,10), lag.max=60, main=paste("PACF of Residuals
```

```
\n", arimaOrderStr))
    qqnorm(res.model,main=paste("QQ plot of Residuals\n", arimaOrderStr))
    qqline(res.model, col = 2)
    print("==================================================================
===")
    cat("Model:",arimaOrderStr)
    print(shapiro.test(res.model))
    print(signif(acf(res.model,plot=F)$acf[1:6],2))
    print(Box.test(res.model, lag = 6, type = "Ljung-Box", fitdf = 0))
    k=0
    LBQPlot(res.model, lag.max = length(model$residuals)-1 , StartLag = k + 1
, k = 0, SquaredQ = FALSE)

    par(mfrow=c(1,1))

}

#                                         p,d,q                            P,D,Q
m1.nzArrivals = arima(nzArrivals,order=c(0,0,0),seasonal=list(order=c(0,1,0),
period=12))
residual.analysis(m1.nzArrivals, 0, 0, 0, 0, 1,0)

## [1] "=====================================================================
"
## Model: SARIMA ( 0 0 0 ) x( 0 1 0 )12
##  Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.98016, p-value = 0.03381
##
## [1] 0.51 0.44 0.48 0.39 0.35 0.25
##
##  Box-Ljung test
##
## data:  res.model
## X-squared = 152.48, df = 6, p-value < 2.2e-16
```

[A5]

```
#                                         p,d,q                            P,D,Q
m2.nzArrivals = arima(nzArrivals,order=c(0,1,0),seasonal=list(order=c(0,1,0),
period=12))
residual.analysis(m2.nzArrivals, 0, 1, 0, 0, 1, 0)

## [1] "=====================================================================
"
## Model: SARIMA ( 0 1 0 ) x( 0 1 0 )12
```

```
##  Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.97476, p-value = 0.008839
##
## [1] -0.420 -0.130  0.150 -0.047  0.045 -0.046
##
##  Box-Ljung test
##
## data:  res.model
## X-squared = 32.465, df = 6, p-value = 1.329e-05
```

[A6]

```
m3.nzArrivals = arima(nzArrivals,order=c(0,1,0),seasonal=list(order=c(0,1,1),
period=12))
residual.analysis(m3.nzArrivals, 0, 1, 0, 0, 1, 1)
```

```
## [1] "=============================================================================
"
## Model: SARIMA ( 0 1 0 ) x( 0 1 1 )12
##  Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.9733, p-value = 0.006217
##
## [1] -0.3800 -0.1500  0.1300 -0.0064 -0.0027 -0.0430
##
##  Box-Ljung test
##
## data:  res.model
## X-squared = 27.768, df = 6, p-value = 0.0001039
```

```
res.m3=residuals(m3.nzArrivals)
eacf(res.m3, ar.max = 5, ma.max =5)
```

[A7]

```
#Define a function sort.score which sort the AIC or BIC scores in ascending o
rder
sort.score <- function(x, score = c("bic", "aic")){
    if (score == "aic"){
        x[with(x, order(AIC)),]
    } else if (score == "bic") {
        x[with(x, order(BIC)),]
    } else {
        warning('score = "x" only accepts valid arguments ("aic","bic")')
    }
}
```

```
# Run coeftest and residual analysis for SARIMA(0,1,1)x(0,1,1)_12
m3_011.nzArrivals = arima(nzArrivals,order=c(0,1,1),seasonal=list(order=c(0,1
,1), period=12))
coeftest(m3_011.nzArrivals)

residual.analysis(m3_011.nzArrivals, 0, 1, 1, 0, 1, 1)  #non-normal,  but lar
ge sample: ok, not correlated

## [1] "=====================================================================
"
## Model: SARIMA ( 0 1 1 ) x( 0 1 1 )12
##  Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.95604, p-value = 0.0001429
##
## [1]  0.0290 -0.1000  0.0870  0.0420 -0.0069 -0.0580
##
##  Box-Ljung test
##
## data:  res.model
## X-squared = 3.5984, df = 6, p-value = 0.7308
```

```
# Run coeftest and residual analysis for SARIMA(0,1,2)x(0,1,1)_12

m3_012.nzArrivals = arima(nzArrivals,order=c(0,1,2),seasonal=list(order=c(0,1
,1), period=12))
coeftest(m3_012.nzArrivals)

residual.analysis(m3_012.nzArrivals, 0, 1, 2, 0, 1, 1) #non-normal but large
sample: ok, not correlated

## [1] "=====================================================================
"
## Model: SARIMA ( 0 1 2 ) x( 0 1 1 )12
##  Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.95355, p-value = 8.748e-05
##
## [1]  0.00160 -0.07700  0.10000  0.04400 -0.00036 -0.05100
##
##  Box-Ljung test
##
## data:  res.model
## X-squared = 3.2042, df = 6, p-value = 0.7828
```

```
# Run coeftest and residual analysis for SARIMA(1,1,2)x(0,1,1)_12

m3_112.nzArrivals = arima(nzArrivals,order=c(1,1,2),seasonal=list(order=c(0,1
,1), period=12))
coeftest(m3_112.nzArrivals)

residual.analysis(m3_011.nzArrivals, 1, 1, 2, 0, 1, 1) #non-normal but large
sample: ok, not correlated

## [1] "=============================================================================
"
## Model: SARIMA ( 1 1 2 ) x( 0 1 1 )12
##  Shapiro-Wilk normality test
##
## data:  res.model
## W = 0.95604, p-value = 0.0001429
##
## [1]  0.0290 -0.1000  0.0870  0.0420 -0.0069 -0.0580
##
##  Box-Ljung test
##
## data:  res.model
## X-squared = 3.5984, df = 6, p-value = 0.7308




#Generate the AIC and BIC scores for all 3 models
sc.AIC=AIC(m3_011.nzArrivals, m3_012.nzArrivals, m3_112.nzArrivals)
sc.BIC=BIC(m3_011.nzArrivals, m3_012.nzArrivals, m3_112.nzArrivals)

#Sort the AIC and BIC score
sort.score(sc.AIC, score = "aic")

##                   df      AIC
## m3_011.nzArrivals  3 2919.742
## m3_012.nzArrivals  4 2921.506
## m3_112.nzArrivals  5 2922.338

sort.score(sc.BIC, score = "bic")

##                   df      BIC
## m3_011.nzArrivals  3 2928.391
## m3_012.nzArrivals  4 2933.037
## m3_112.nzArrivals  5 2936.752
```

[A8]

```
m1.nzArrival = Arima(nzArrivals,order=c(0,1,1),seasonal=list(order=c(0,1,1),
period=12))
```

```
future = forecast(m1.nzArrival, h = 10)
future

plot(future)
```