Laëtitia Lachat
Erasmus Student
N° : 17338341

FORECASTING

Report about iGraal

Part I : Visualisation tools

I have chosen to analyse a time series from Google Trends. Our time series shows how often the term "iGraal" is searched since 2004. iGraal is a famous cash back website in France. The cash back is a system where a client is using a cash back website to buy a product from an online shopping website. The online shopping website will remunerate the cash back website, and then the cash back website will give a percentage of this amount to the client. A lot of famous websites exist (like iGraal, Ebuyclub, Poulpeo, Alkamio).

The forecasting analysis of iGraal is interesting from a financial point of view. We can imagine it would be useful for a company, which recently has created an online shopping website, to understand the popularity of iGraal. It would allow the company to decide if they would like to conclude an agreement with iGraal or if it would be preferable to use another cash back website. The company could also choose to conclude a seasonal agreement in order to pay iGraal for advertising only during the periods whose visits on iGraal are maximum.

iGraal has been created in 2006, so we can delete observations from 2004 to 2006 (indeed, the results are null). Moreover, we can also delete the observation for January 2018 seeing as the information are incomplete for this month. As we can see on **Plot 1**, for the first two years (2007 and 2008) of iGraal, the search for the term "iGraal" is very fluctuating. It's likely because iGraal was very new. So we can deduce that observations from 2007 and 2008 are inconsistent, and we are not going to use them.

From **Plot 2**, we can identify different patterns with the time plot. It is suggesting a seasonal period of 12 months with a positive trend and noises. From the ACF plot, we can see strong peaks for k=12 and k=24. From PACF plot, we can see a strong and positive peak for k=12. It confirms the hypothesis of a periodicity equal to 12. Moreover, the peaks in ACF plot are all positive : there is a positive trend. In the PACF plot, we can also remark there is a negative peak for k=15.

Moreover, from **Plot 3**, which is a seasonal plot where the period is equal to 12 months, the plot seems to be coherent.

From **Plot 4** and **Plot 5**, which are respectively the additive and multiplicative decomposition of our time series, we can confirm the seasonal pattern of 12 months, with a general positive trend and noises.

There is a positive trend so we can exclude SES algorithm. Moreover, there is a seasonal pattern so we can exclude DES algorithm. In conclusion, it seems logical to use Seasonal Holt-Winters additive or Seasonal Holt-Winters multiplicative algorithms (presences of a trend, a seasonal pattern).

There is a seasonal pattern, so we would be using seasonal ARIMA Models. Moreover, the first step would be to remove the trend so there would be a non-null coefficient for the differentiating term of ARIMA algorithm.

Part II : Holt Winters Algorithms

As it was explained previously, we have a positive trend and a seasonal pattern. So we are going to apply a Seasonal Holt Winters.

We have chosen as a criteria to minimize the Sum of Square Errors : for the additive seasonal Holt Winters, the SSE is equal to 2269.229 and for the multiplicative seasonal Holt Winters, the SSE is equal to 2349.09. So we are going to apply the additive seasonal Holt Winters.

The computations of parameters for the SHW_+ are showing in the **R Output 1**. α and Υ are positive parameters, and β is null. Moreover, the parameter b is suggesting that the trend is positive but that the slope is very slow. Between November and January, the coefficients (s1, s11, s12) are positive and high. Therefore, we can deduce that the research "iGraal" is more strong during those months. However, between February and October, the parameters (from s2 to s10) are negative with a summit for August. Therefore, we can conclude that the search "iGraal" on Google is not usual for those months.

When we are computing predictions for 24 months with confidence intervals, we obtained the **Plot 6**. The predictions are drawn in blue. We can conclude that the search "iGraal" will continue to increase for the next two years, with important peaks between November-January (during the period of Christmas and Winter sales). However, we can see that the confidence interval of 80% in dark grey is important. Moreover, the confidence interval of 95% in pale grey is also high. In conclusion, the application of seasonal Holt Winters doesn't give us a very good result because there is a high confidence interval.

Sum up of The Holt-Winters Approach to Exponential Smoothing : 50 Years Old and Going Strong, P. Goodwin, Foresight 2010 :

Holt-Winters method is 50 years old and it is still currently used because of this simplicity. However, it has been adapted to deal with more problems.

The first issue was to identify outliers in order to limit their influences over the estimates of the level, trend, seasonal pattern and smoothing constants. A mechanism was created to discover outliers thanks to a threshold and then replace them by "cleaned" and moderate values, which are on the boundary of the threshold.

The second one was to handle data with multiple seasonal cycles, not with just one cycle. For each extra cycle, the solution was to add a smoothing equation and a smoothing constant.

The last one was to obtain prediction intervals, which are not too narrow and not misleading us. Using Bayesian framework will allow to represent smoothing constants and values with probability distributions. Then, those distributions are adapted with each new observation.

Part III : ARIMA models

From **Plot 2**, we can see there is a positive trend. So we test the ARIMA $(0,1,0)\ (0,0,0)_{12}$. The residuals of this algorithm give us the **Plot 7**. The trend was removed. However, we see strong peaks for k=12, k=24, k=36 for ACF with a slow decrease.

Therefore, it indicates a seasonal trend so we apply the ARIMA $(0,1,0)\ (0,1,0)_{12}$. The **Plot 8** shows us the residuals with this algorithm. We can see seasonal peaks for k=12, k=24 in the ACF Plot.

Therefore, we are going to apply the ARIMA $(0,1,0)\ (0,1,1)_{12}$. The residuals give us the **Plot 9**. We can see there is still one strong peak on the ACF plot and there is an exponential decrease on the PACF plot.

Therefore, we apply the ARIMA (0,1,1) $(0,1,1)_{12}$. The residuals give us the **Plot 10**. We don't distinguish something significant. Therefore, we deduce that the appropriate ARIMA model is the ARIMA (0,1,1) $(0,1,1)_{12}$.

When we are computing predictions for 24 months with confidence intervals for the ARIMA (0,1,1) $(0,1,1)_{12}$, we obtained the **Plot 11**. As for Holt-Winters algorithm, we can conclude that the search "iGraal" will continue to increase for the next two years, with important peaks between November-January (during the period of Christmas and Winter sales).

However, we can see that the confidence intervals are not better with this ARIMA model. We can suggest it's because of the year 2017 which seems to be a particular bad year for the search "iGraal". It seems to be an outlier which disturbs every algorithm used. However, it doesn't necessarily mean that iGraal is a cash back website in decline. Different reasons could explain this decrease : for example, the application iGraal could have more success than before.

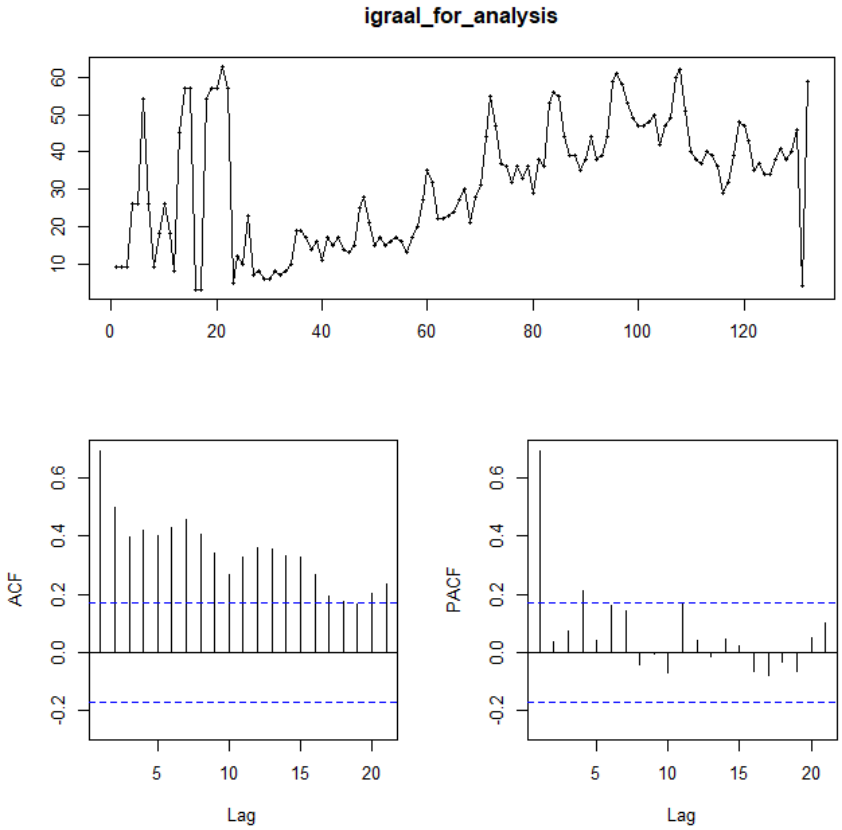
Warning about hypotheses used and limit of mathematical techniques for forecasting :

Hypotheses are very strict for the forecasting algorithms we have used. Indeed, we are assuming that continuity assumption holds. And this is a very powerful assumption we have made. We can easily imagine that the creation of a new service by iGraal will have repercussion on the search "iGraal" in the search engine Google. In this case, the continuity assumption would not hold. Therefore, it's difficult to interpret with certainty the forecasting reports.

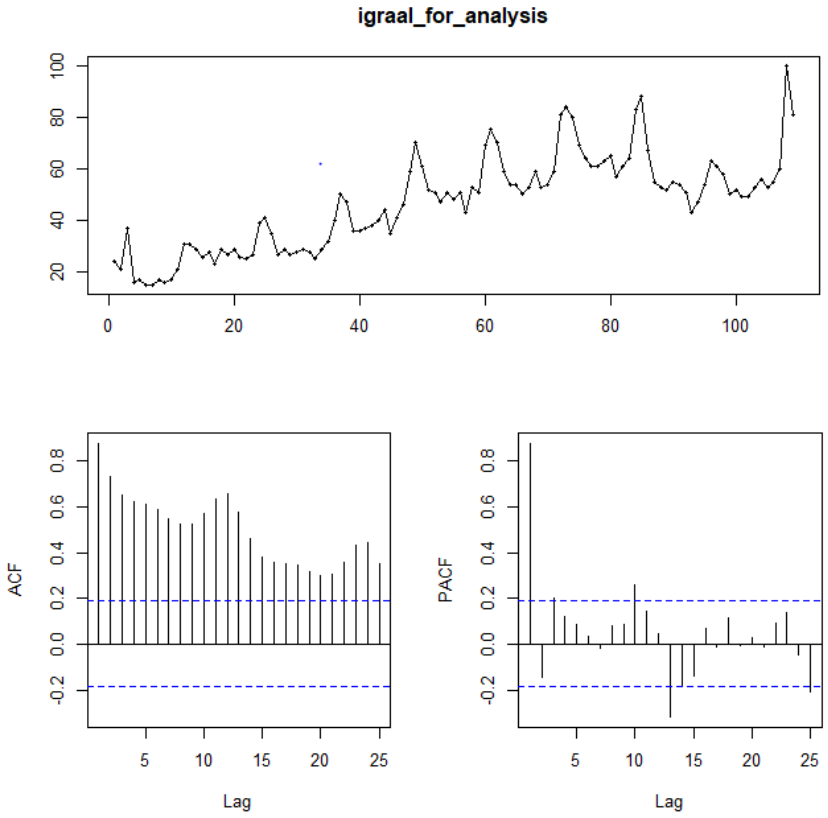
Moreover, we have to be careful about the interpretation of forecasting analysis. Indeed, during lectures, we have seen the example of "the decline of Facebook" based on the analysis of the search "Facebook" in Google. If the search "Facebook" in Google was decreasing, it did not mean that Facebook was losing influence. For example, the daily use of applications could explain this decrease.

APPENDIX

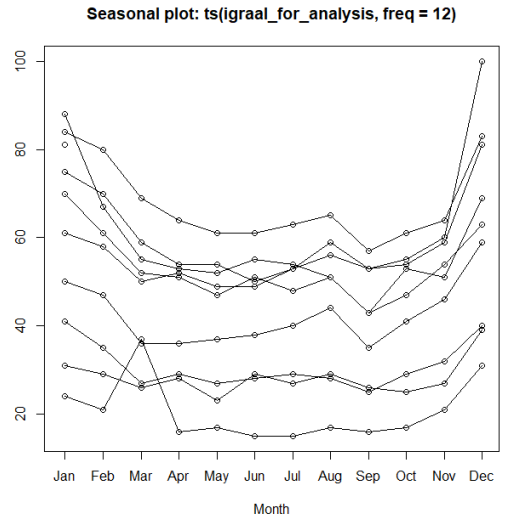
Plot 1



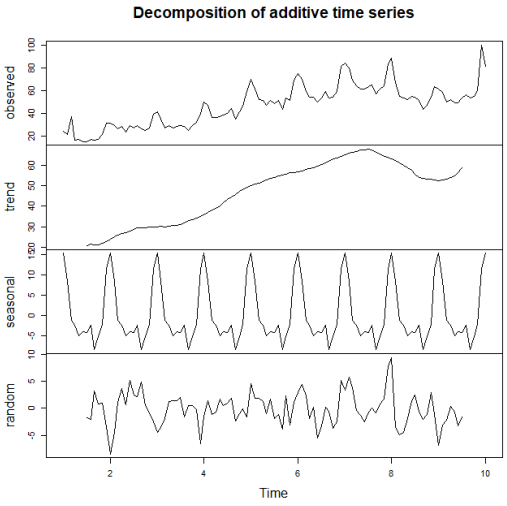
Plot 2



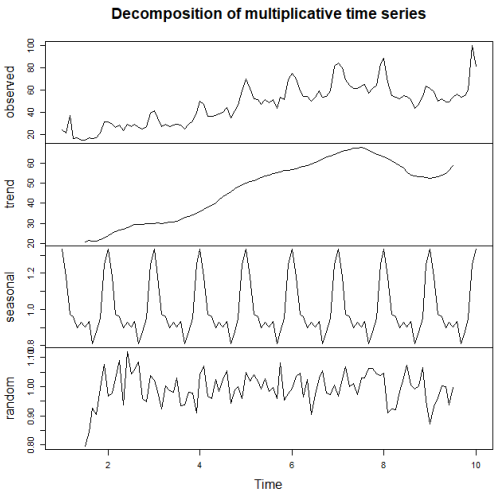
Plot 3



Plot 4



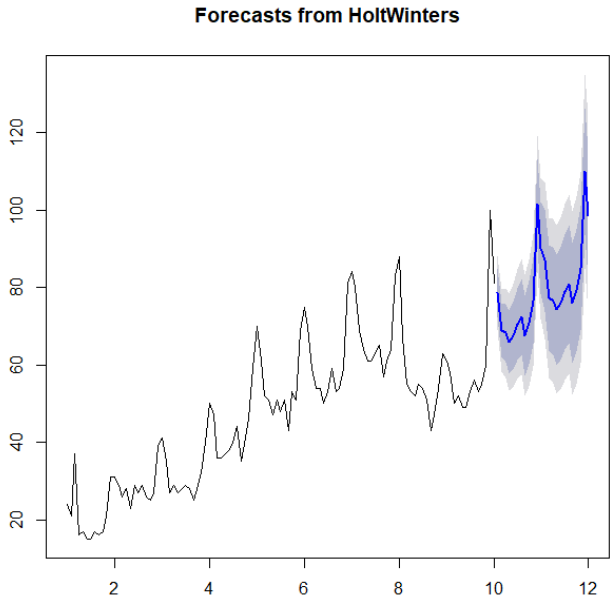
Plot 5



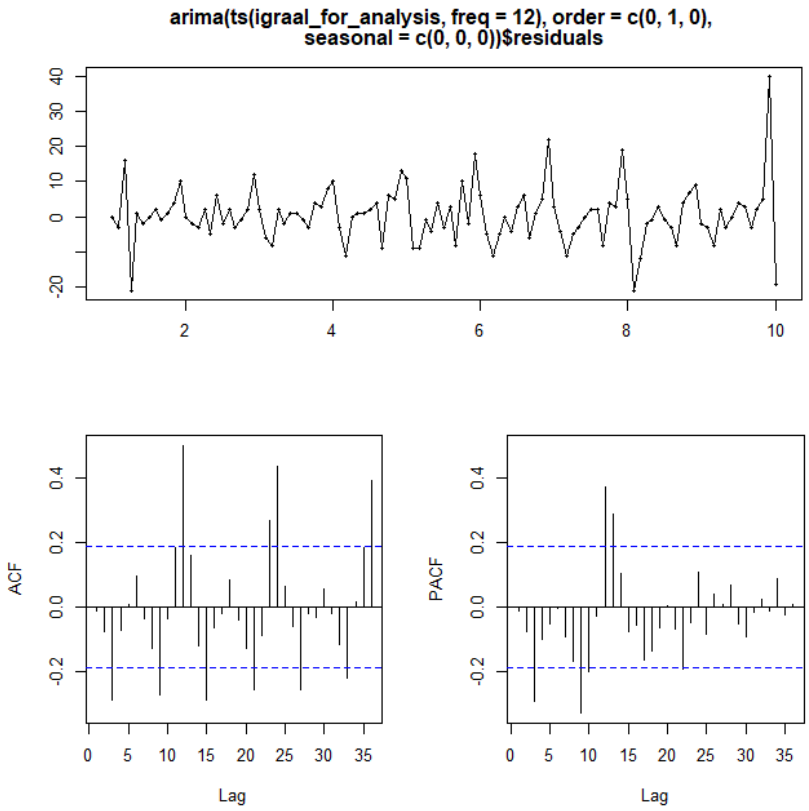
R Output 1

```
HoltWinters(x = ts(igraal_for_analysis, freq = 12), seasonal = c("additive"))

Smoothing parameters:
 alpha: 0.4939187
 beta : 0
 gamma: 0.816648

Coefficients:
          [,1]
a   71.4450444
b    0.7004662
s1   6.4576105
s2  -3.9513678
s3  -5.3326701
s4  -8.3976005
s5  -7.5159902
s6  -5.2298161
s7  -3.9293546
s8  -9.5126836
s9  -6.8821206
s10 -1.8673157
s11 22.4617034
s12 10.3905740
```

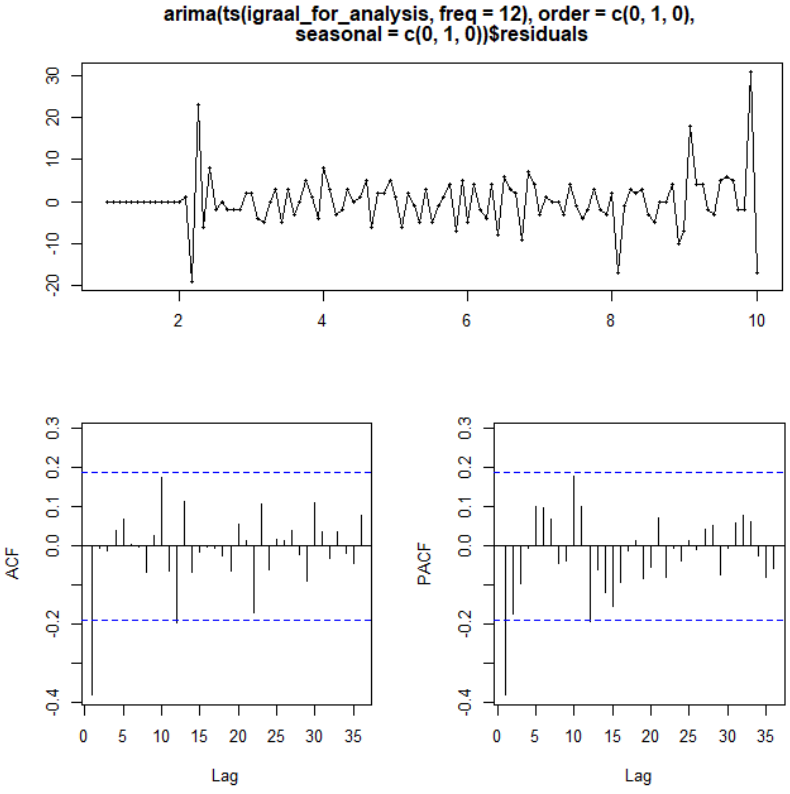
Plot 6



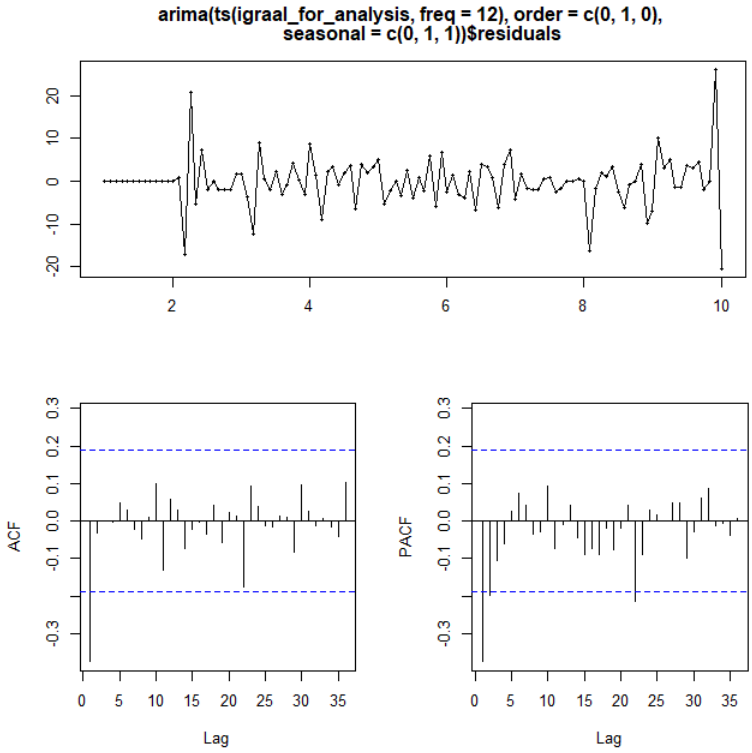
Plot 7


arima(ts(igraal_for_analysis, freq = 12), order = c(0, 1, 0),
seasonal = c(0, 0, 0))$residuals
ACF
PACF
Lag
Lag

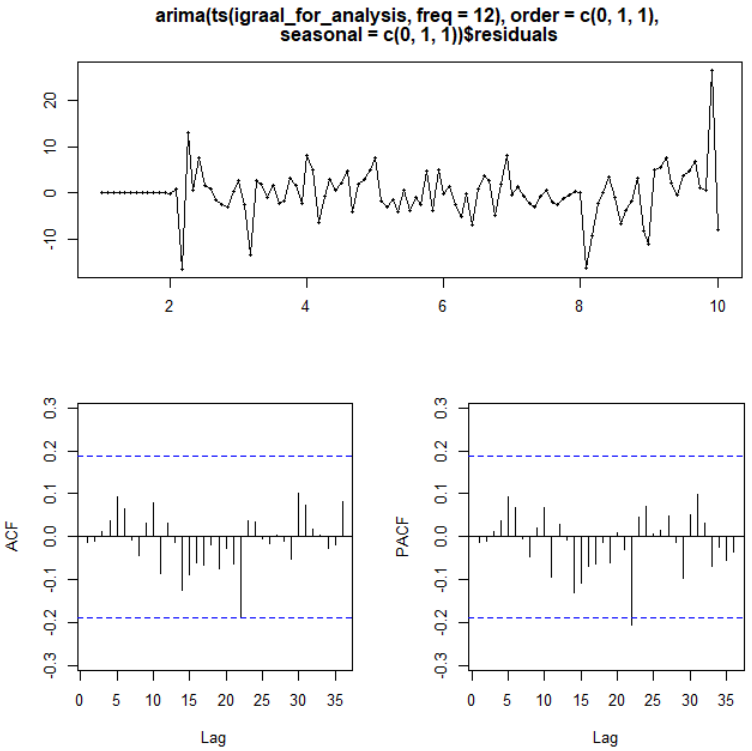
Plot 8


arima(ts(igraal_for_analysis, freq = 12), order = c(0, 1, 0),
seasonal = c(0, 1, 0))$residuals
ACF
PACF
Lag
Lag

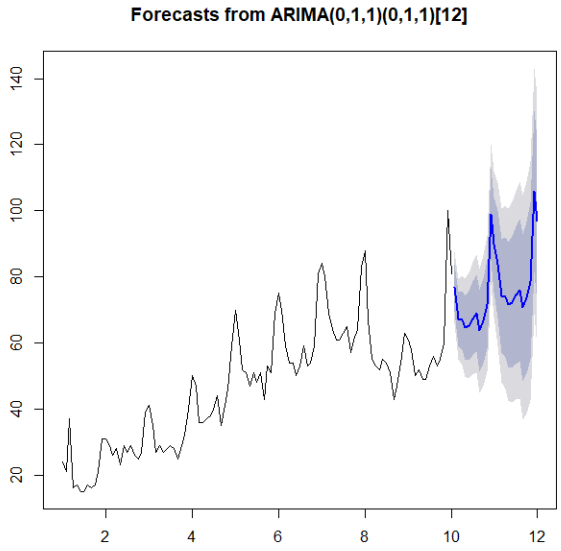
Plot 9



Plot 10



Plot 11


Forecasts from ARIMA(0,1,1)(0,1,1)[12]
20
40
60
80
100
120
140
2
4
6
8
10
12