Predictive model report

Firstly we use pandas and python for finding the revenues per day (in file **revenues_python**) and save the result in **Revenues.csv**. As we plot the Revenues data with time we can see that there is an outlier for the revenue of the last day. We also notice that the time series isn't stationary (Figure 1). For that reason we plot in Figue 2 the correlogram and the partial correlogram (ACF and PACF).

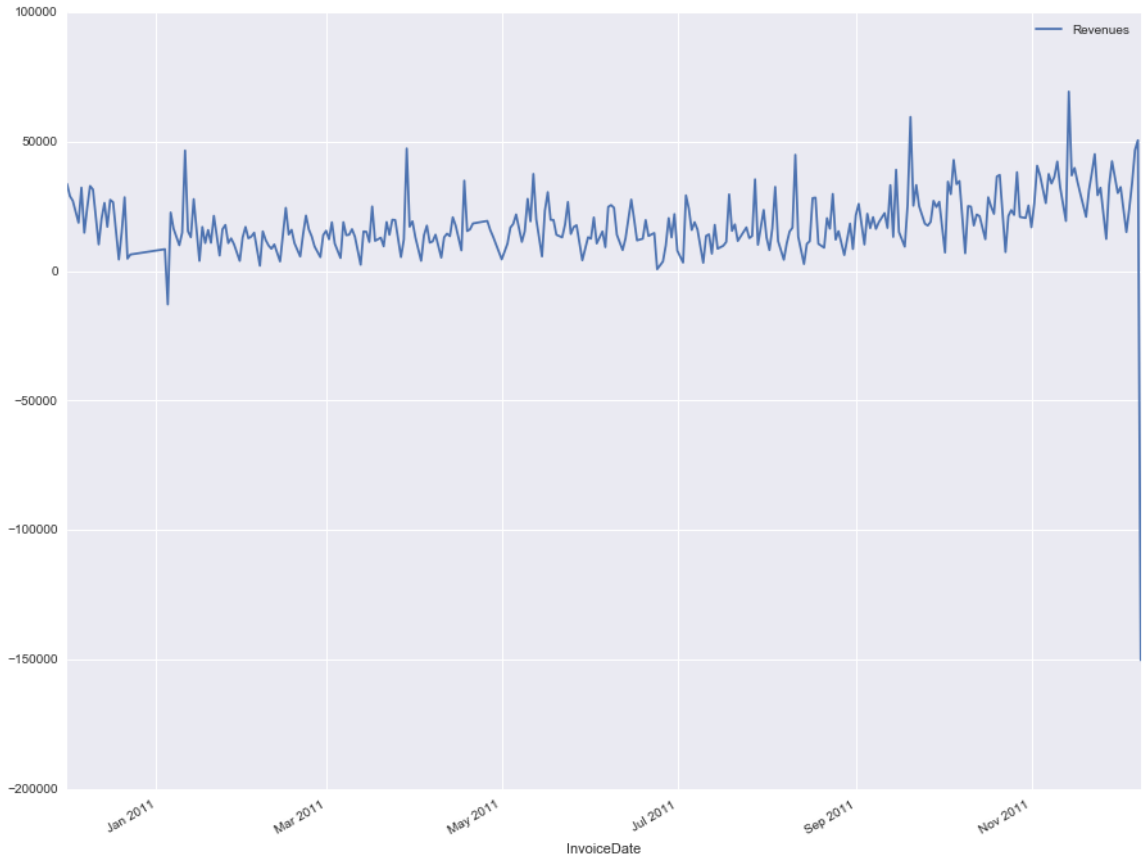

Figure 1 Revenues from days 1 to 305

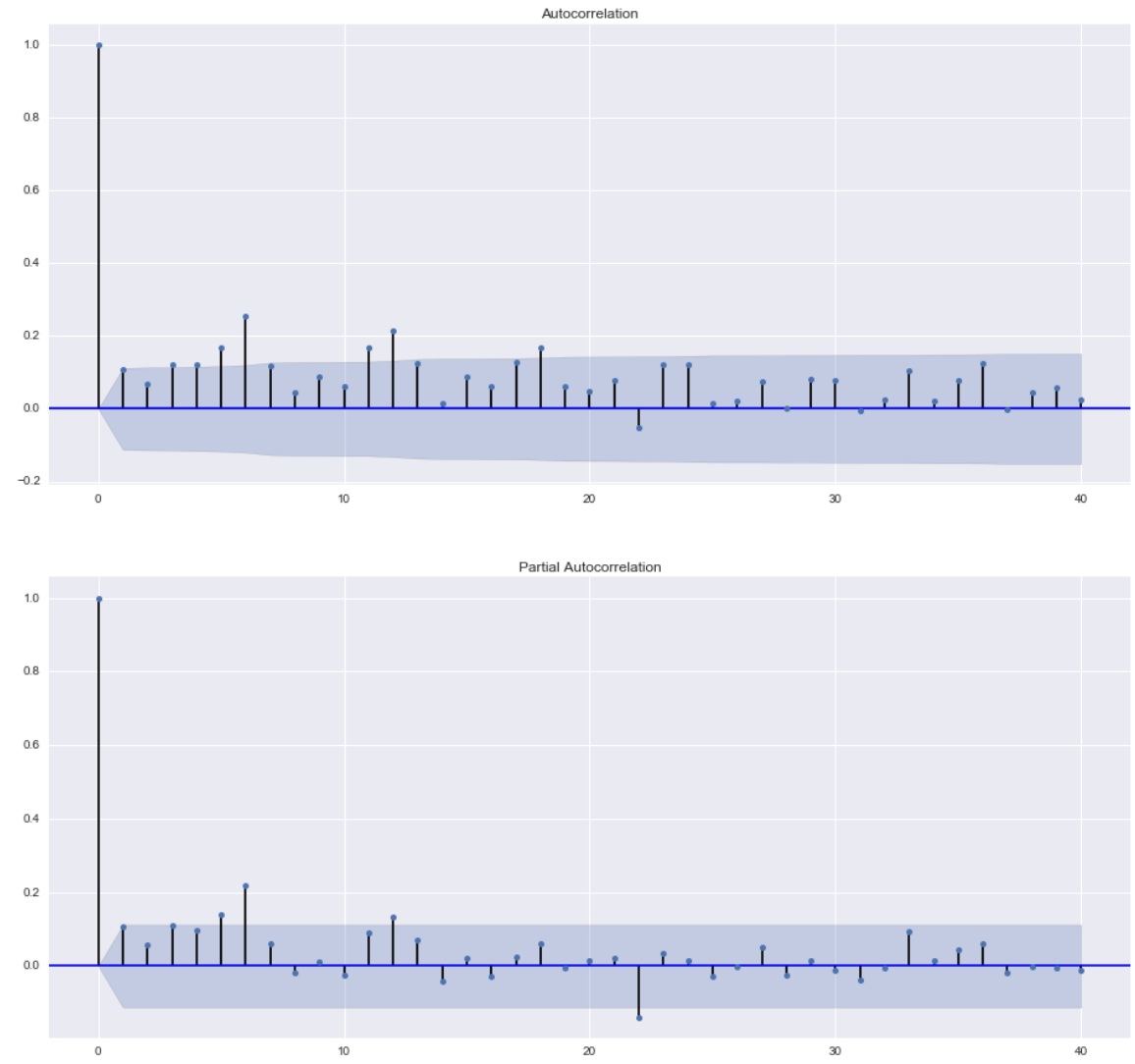

Figure 2 ACF and PACF of the Revenues series

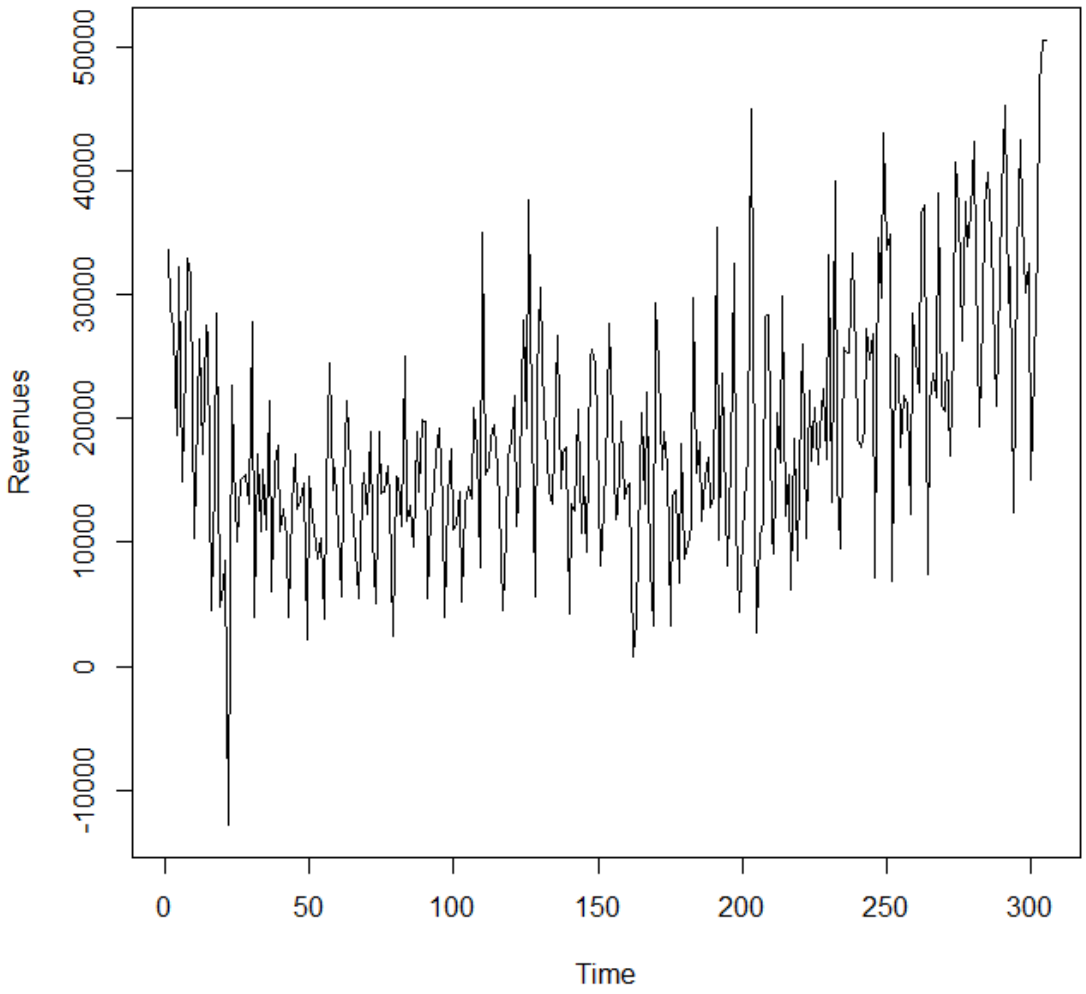

Figure 3 Revenues from days 1 to 305 after the replacement of outlier

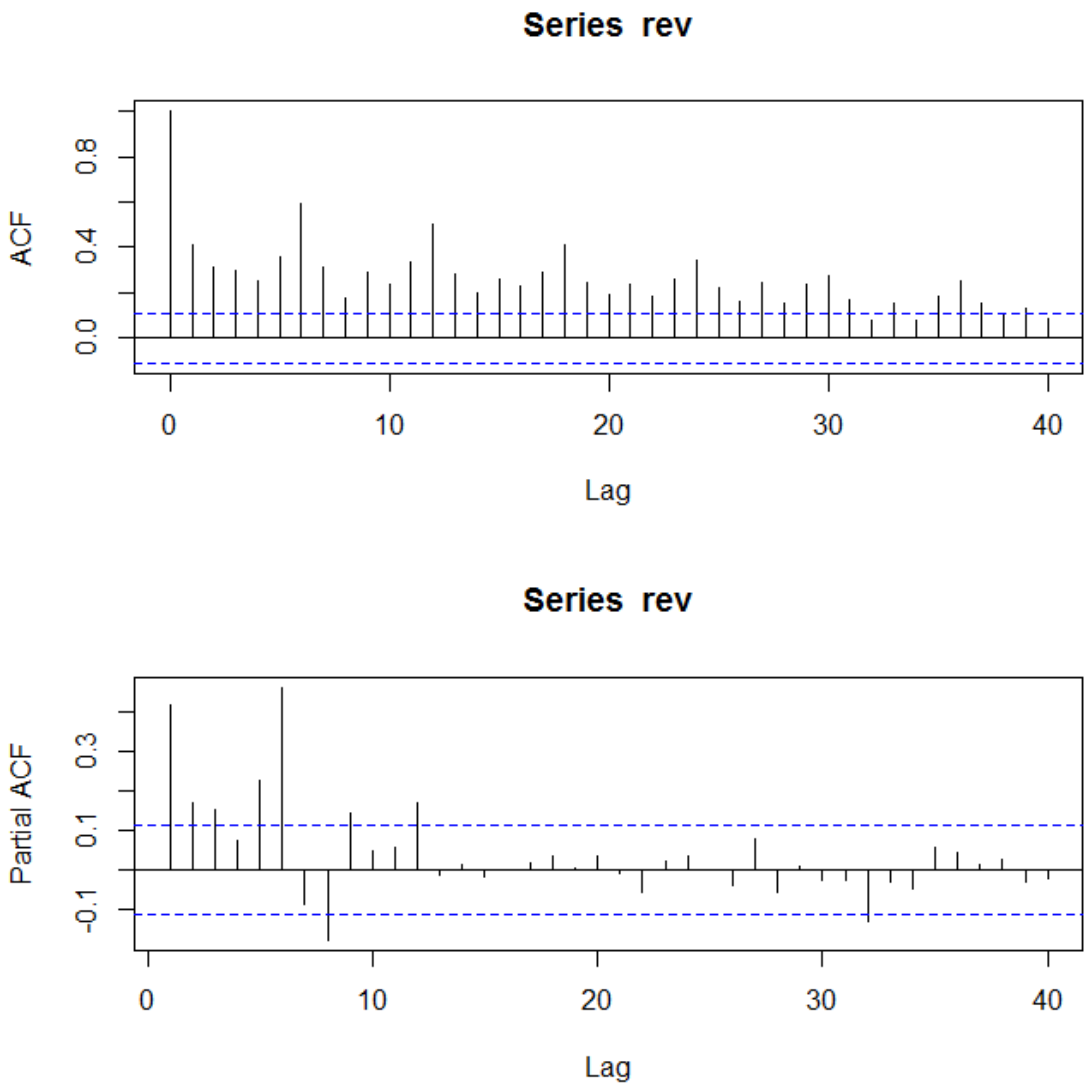

Figure 4 ACF and PACF of the Revenues series after the replacement of outlier

Subsequently, we use the R file (**revenues_R**) where we open the **Revenues.csv** and we use the tclean() for replacing the outlier. Then we plot again the Revenues series and the ACF and PACF plots in Figures 3 and 4 respectively.

Then we will difference the data for eliminating the trend. For this purpose we will take the first difference:

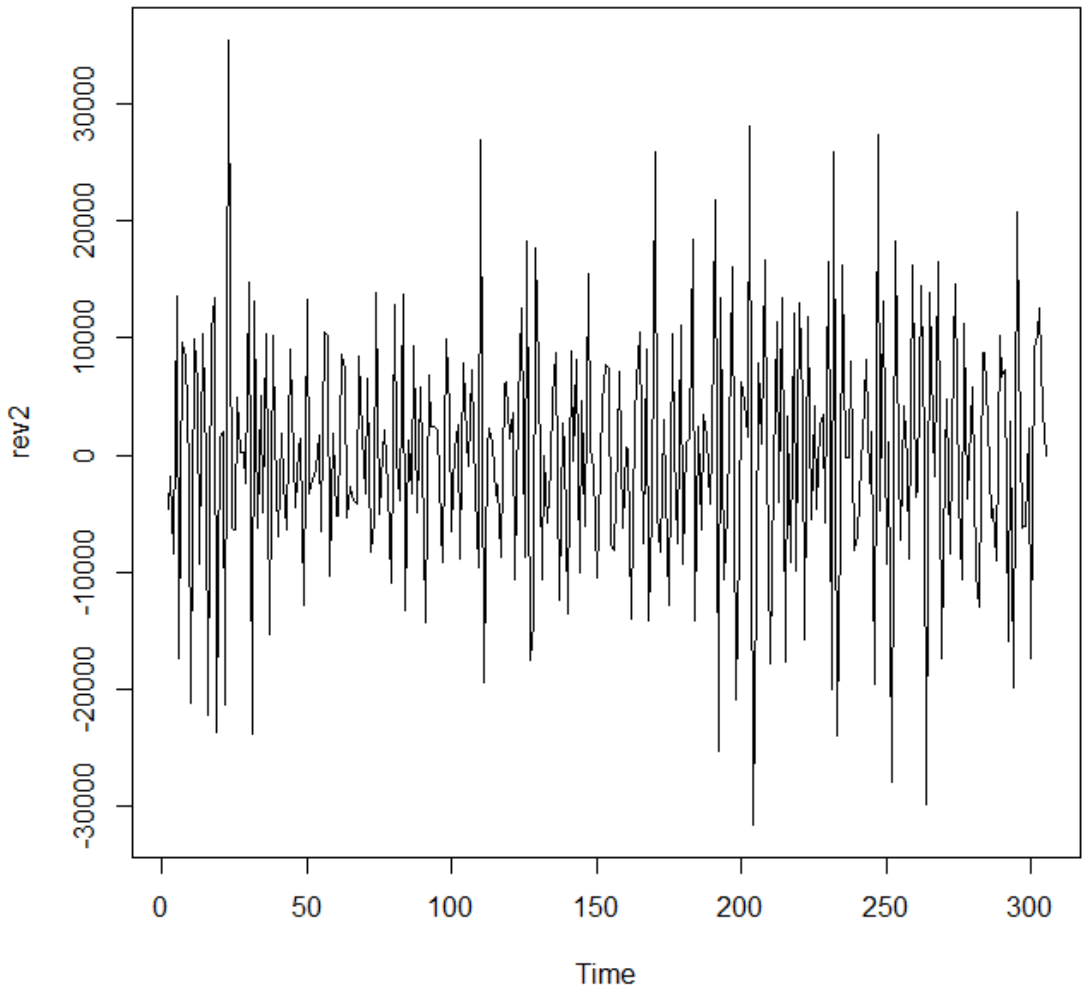

Figure 5 The first differenced Revenues series

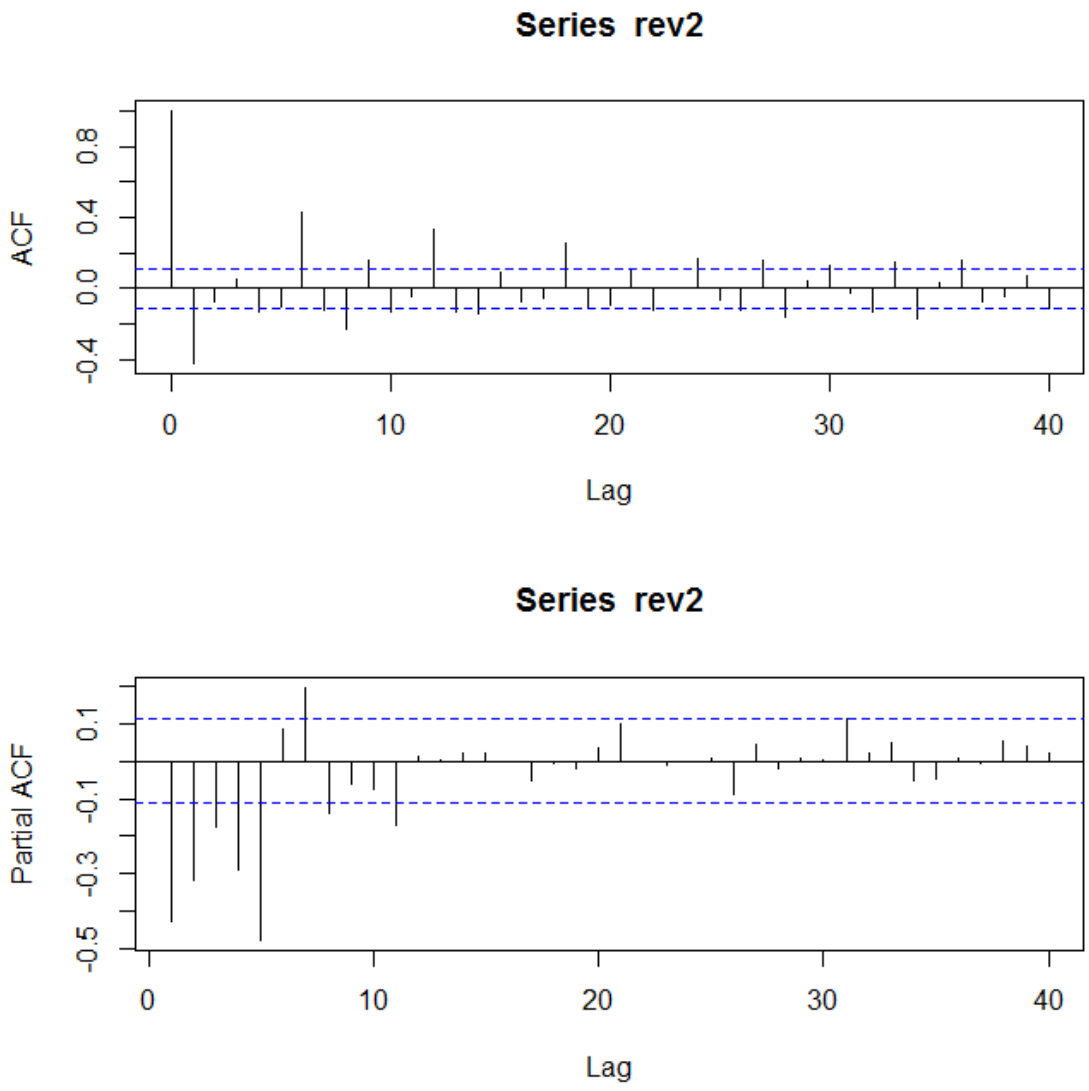

Figure 6 ACF and PACF of the first differenced Revenues

From the plot in Figure 6 we can note that we have peaks at 6, 12, 18 suggested a seasonal model at lag 6. The plots suggest either a seasonal moving average of order Q = 0 and a seasonal autoregression of possible order P = 1, or of order Q=0 and P=2. Inspecting the ACF and PACF at the within season lags it appears that both the ACF and PACF are tailing off. This result indicates that we should consider fitting a model with both p = 1 and q = 1. An alternative choice is p = 1 and q=0.

Fitting the three models suggested by the observations and computing the AIC for each, we obtained:

1. $ARIMA(1,1,1)X(2,0,0)_6$, , AIC = 6270.49

2. $ARIMA(1,1,0)X(1,0,0)_6$, , AIC = 6343.29
3. $ARIMA(1,1,1)X(1,0,0)_6$, , AIC = 6281.64

We prefer the $ARIMA(1,1,1)X(2,0,0)_6$ model because it the has the smallest AIC.

Figure 7. shows the diagnostics for this model, leading to the conclusion that the model is adequate.

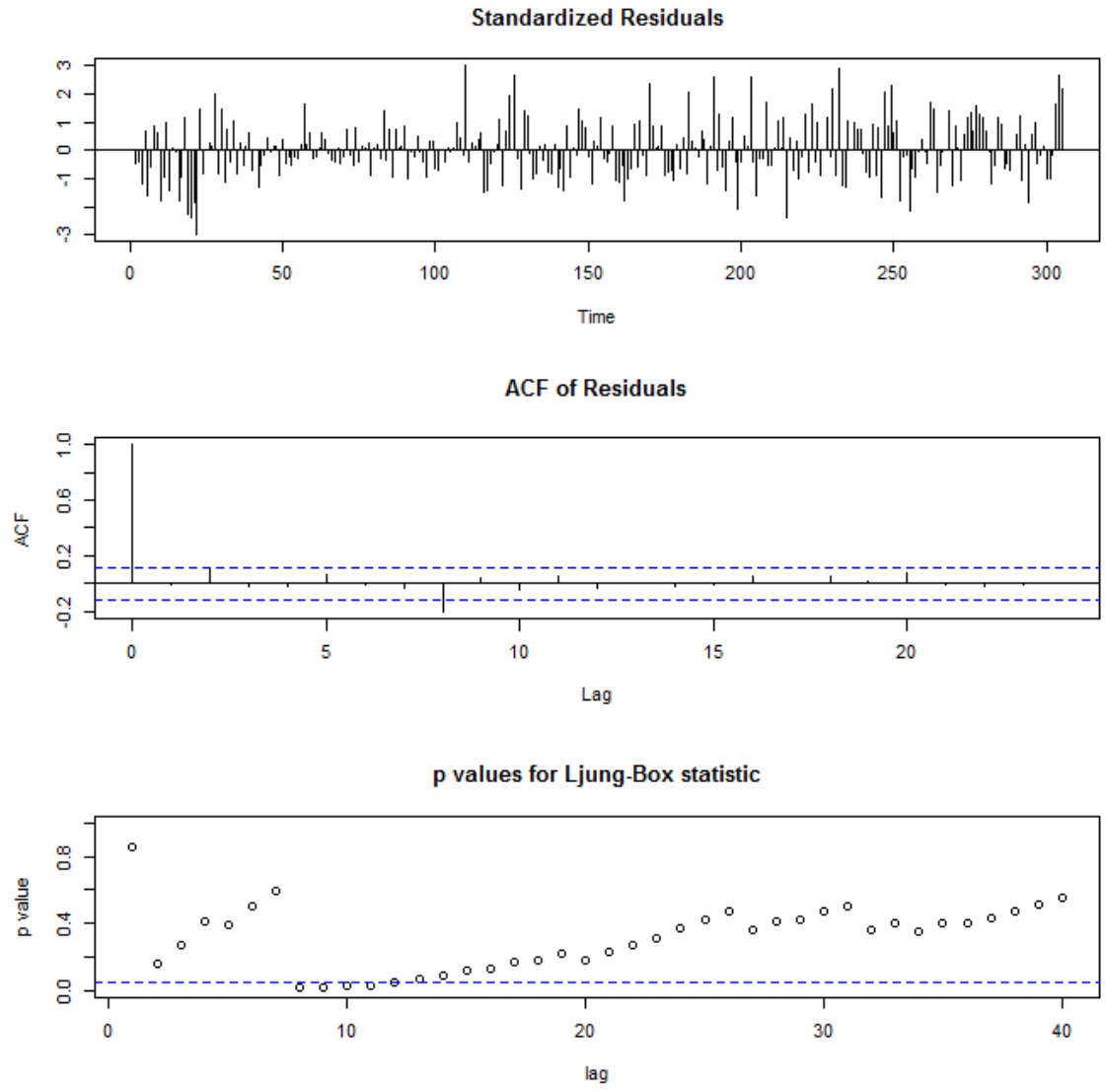

Figure 7 Diagnostics for the $ARIMA(1,1,1)X(2,0,0)_6$ fit on the Revenues data

Finally in Figure 8 we present a forecasting for the final model for the next 15 days.

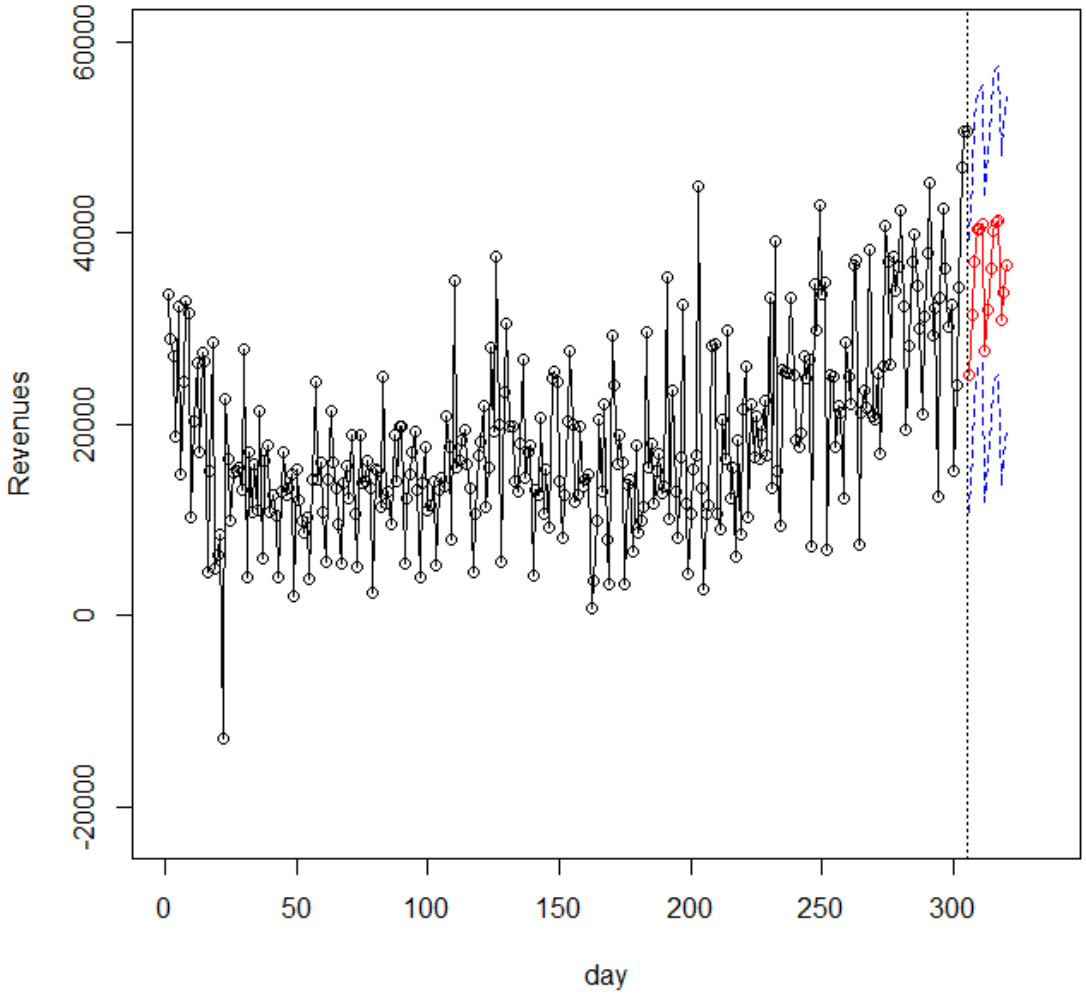

Figure 8 Forecasts and limits for Revenues index. The vertical dotted line separates the data from the predictions

The predicted revenue for the next day order (306th day) is 25080 with the upper limit = 39349 and the lower limit = 10812.