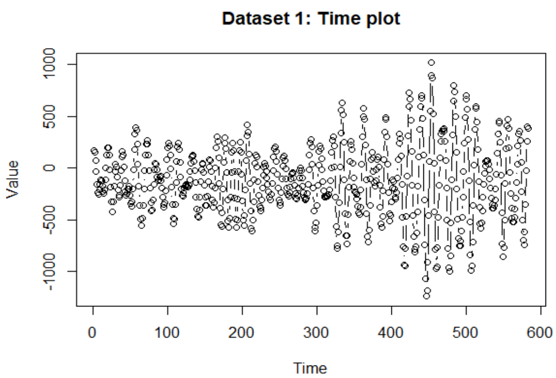
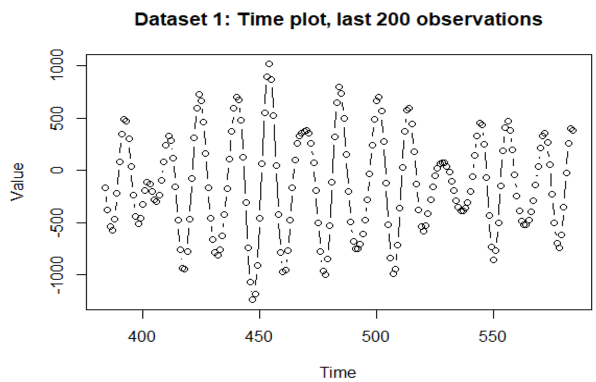
Technical Appendix – data set 1

Summary of the time series

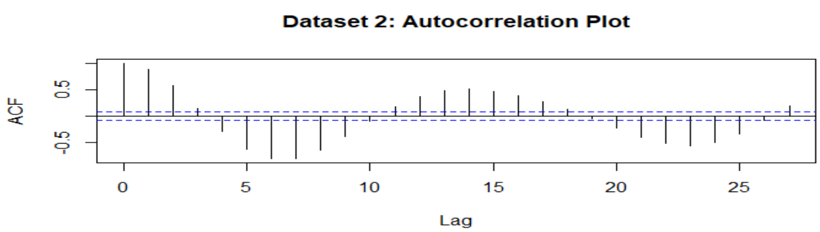
Min.	*1st Qu.*	*Median*	*Mean*	*3rd Qu.*	*Max.*
-1240.78	-337.71	-147.49	-126.62	98.05	1016.42

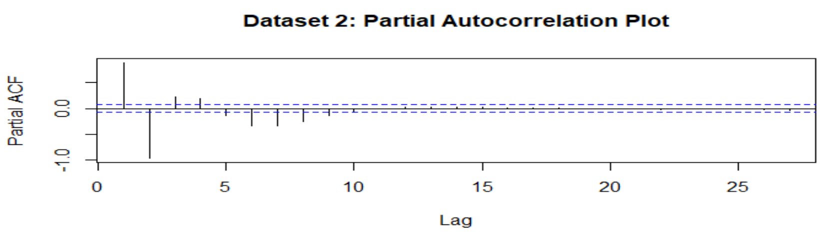
Time series plots





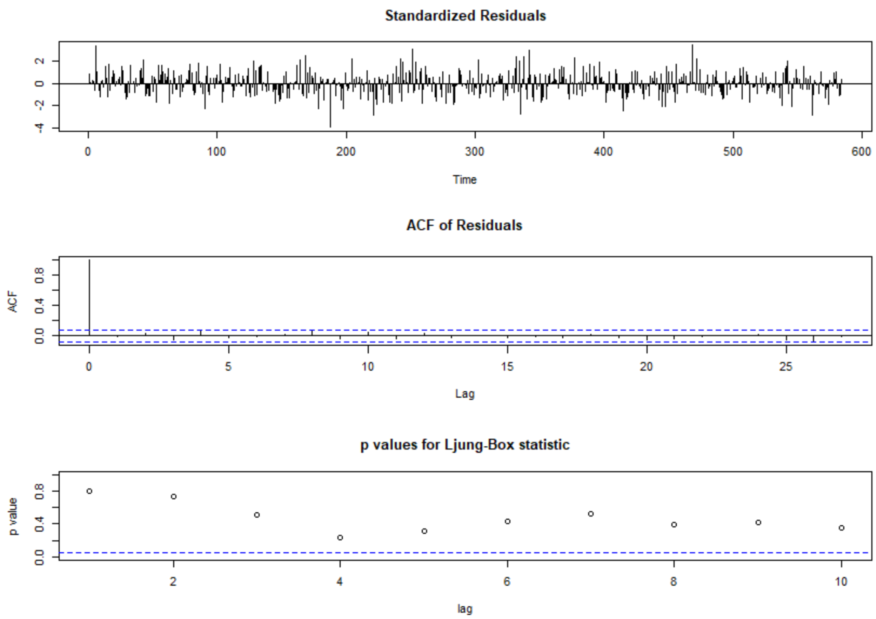
The timeseries plot data set indicates that data is stationary. There is no obvious trend in data. The Augmented Dickey-Fuller test confirmed that a good starting point for modeling this time series is an ARMA model.



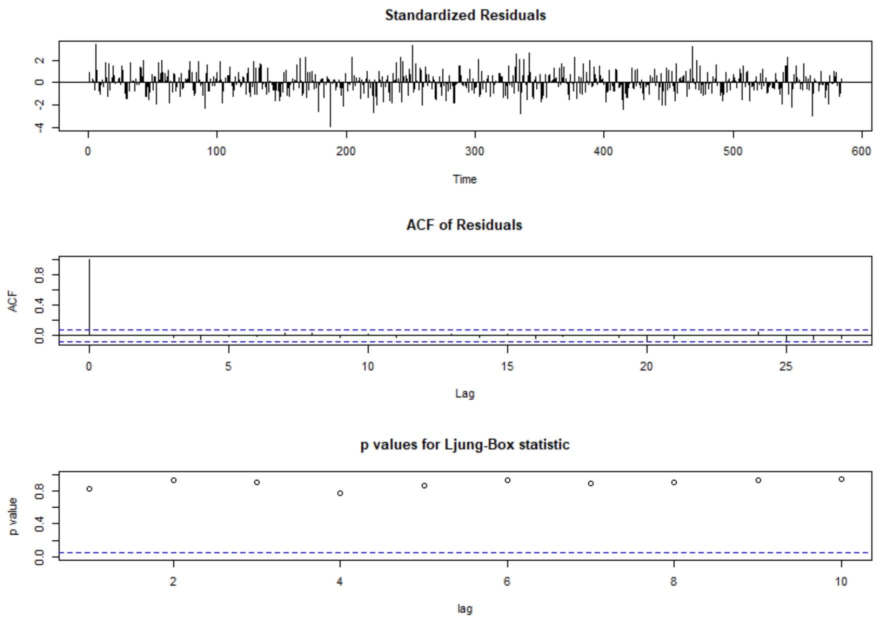


There is a sinusoidal decay in **ACF** plot, that implies that our model has an AR component of order 2 or higher. Also, it might be masking some peaks that would indicate the order of the MA component. The **PACF** plot has first two obvious spikes at lags 1 and 2 that indicate MA(2) after which we also have an sinusoidal decay.

Therefore, the smallest model according to ACF and PACF plots that we should try to fit should be an ARMA (2,2).
The auto.arima() R function recommended the ARMA (5,0) model and AIC 10x10 matrix gave us that the smallest AIC value has model ARMA(6,1).
Judging by the significance of the last coefficient all models that had p<5 and/or q<3 had all last coefficients significant. The largest model with significant last coefficient was the ARMA(7,2).

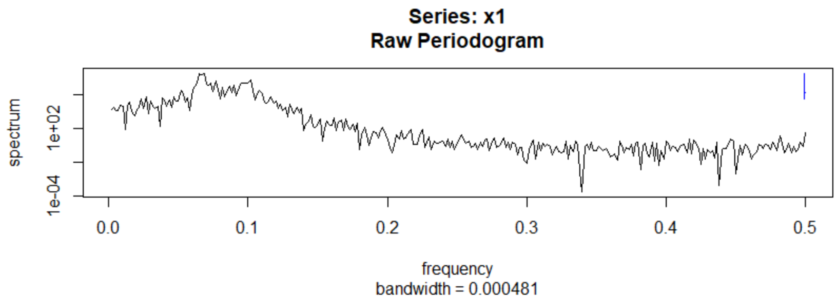


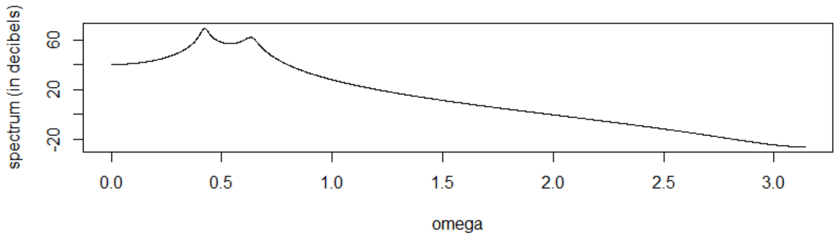
1 - Residual Plot of the model ARMA(4,3)



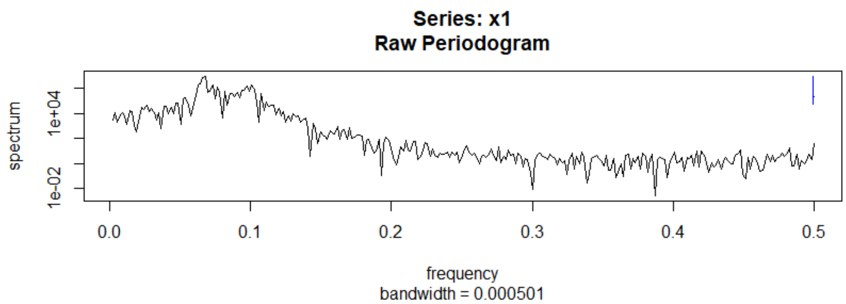
2 - Residual Plot of the model ARMA(7,2)

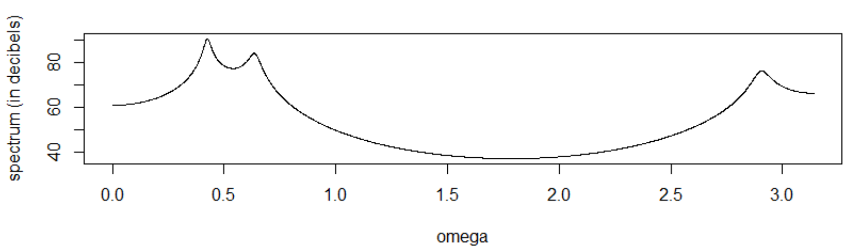
The closer examination of the residuals indicated that the smallest model that had random residuals with nonsignificant ACF values is ARMA(4,3). Therefore, models ARMA(7,2) and ARMA(4,3) are two candidate models.





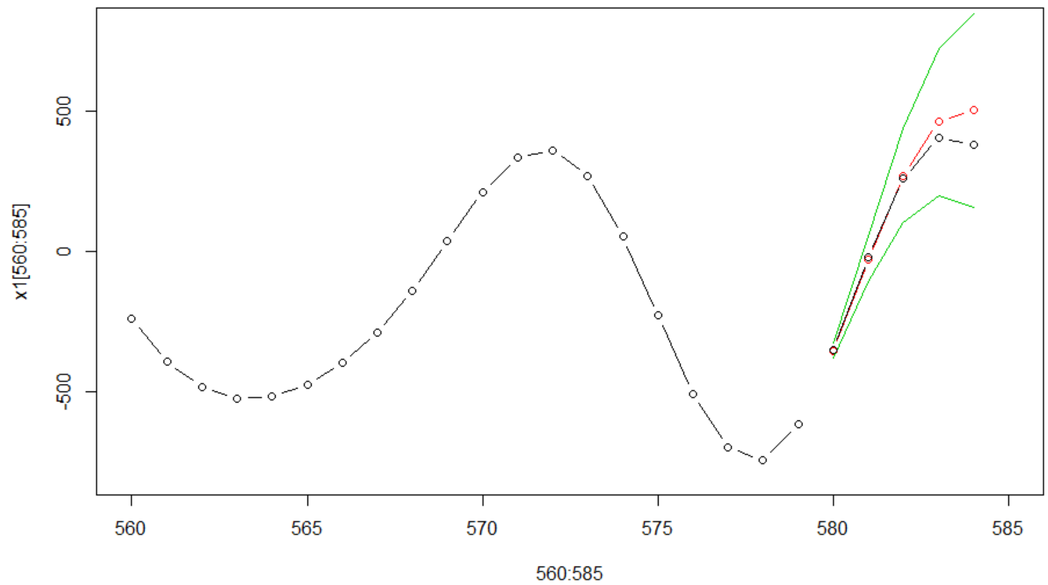
3 - Model ARMA(4,3)



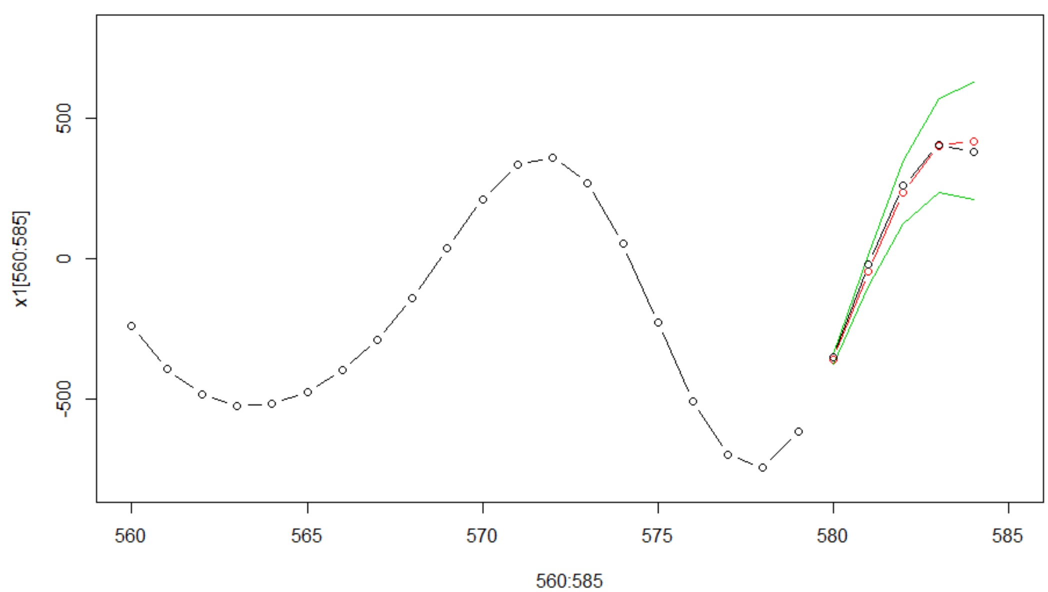


4 - ARIMA(7,2)

Periodgram compared to spectrogram shows that ARMA(4,3) is better fit.



5 - Prediction interval for model ARMA(4,3)



6 - prediction interval for model ARMA(7,2)

According to in sample predictions, two models are close.

Executive report – data set 1

By the thrall analysis of the data from DataSet 1 we concluded that the model appropriate to model the data is an model produces observations that depend on the previous 4 observations and 3 past error terms, or ARMA(4,3) with coefficients:

ar1	*ar2*	*ar3*	*ar4*	*ma1*	*ma2*	*ma3*	*intercept*
3.36	-4.73	3.22	-0.92	-0.63	0.38	-0.21	-127.66

We are 95% confident that next nine observations after the last observed one are going to be between values of the lower and upper boundary:

Upper boundry	*Point prediction*	*Lower boundry*
188.07	207.77	227.47
-96.30	-38.92	18.46
-382.73	-271.34	-159.95
-584.32	-416.64	-248.96
-651.36	-442.03	-232.71
-588.35	-361.89	-135.43
-453.61	-226.74	0.14
-338.39	-99.70	138.98
-313.07	-30.50	252.06

And that the overall look of the time series of the DataSet1is going to look like: