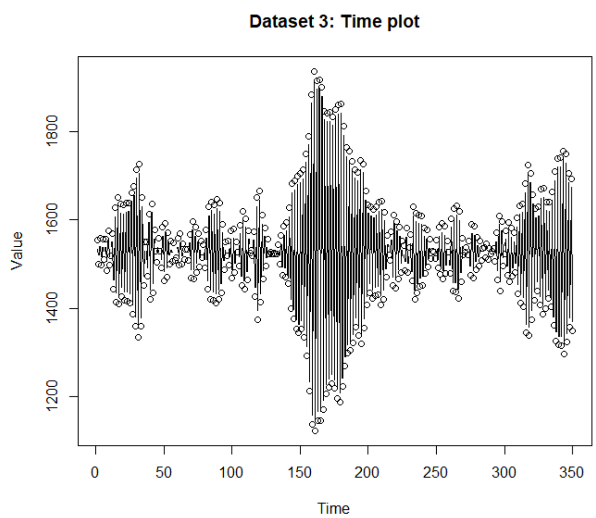
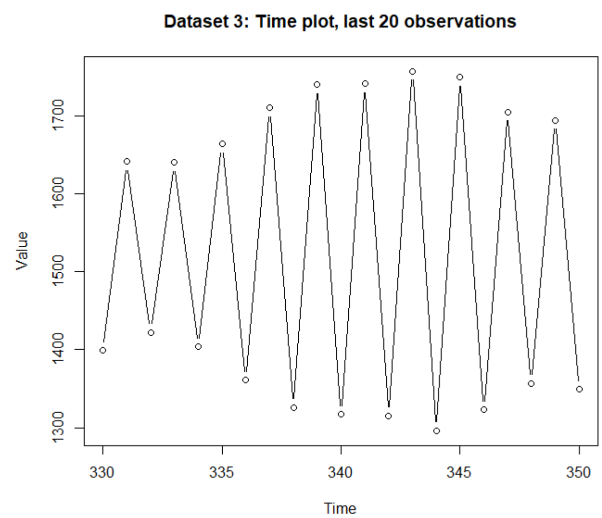
Technical Appendix – data set 3

Summary of the time series, 350 observations,

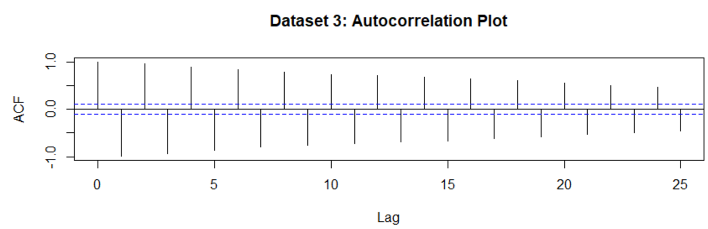
Min.	*1st Qu.*	*Median*	*Mean*	*3rd Qu.*	*Max.*
1122	1451.25	1528	1527.75	1603.75	1937

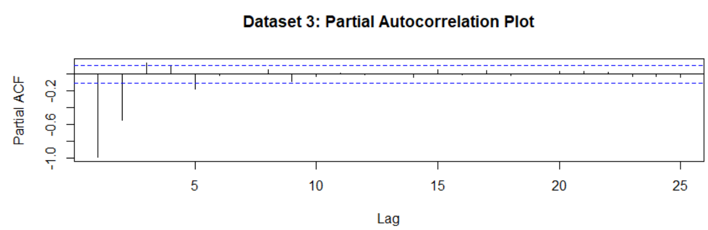
Time series plots





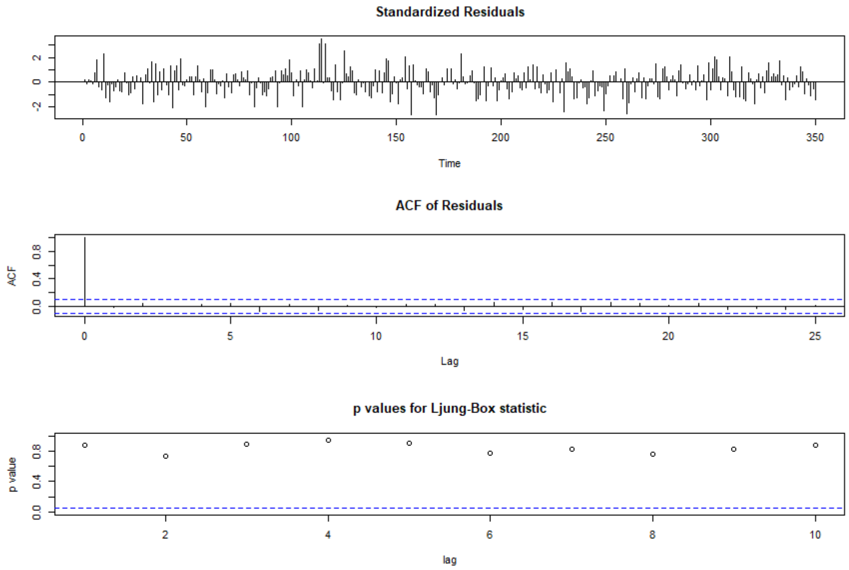
The timeseries plot data set indicates that data is stationary. There is no obvious trend in data. But we can see a clear period of 2, which indicates that AR component is of the order 1 or higher. The Augmented Dickey-Fuller test confirmed that a good starting point for modeling this time series is an ARMA model.



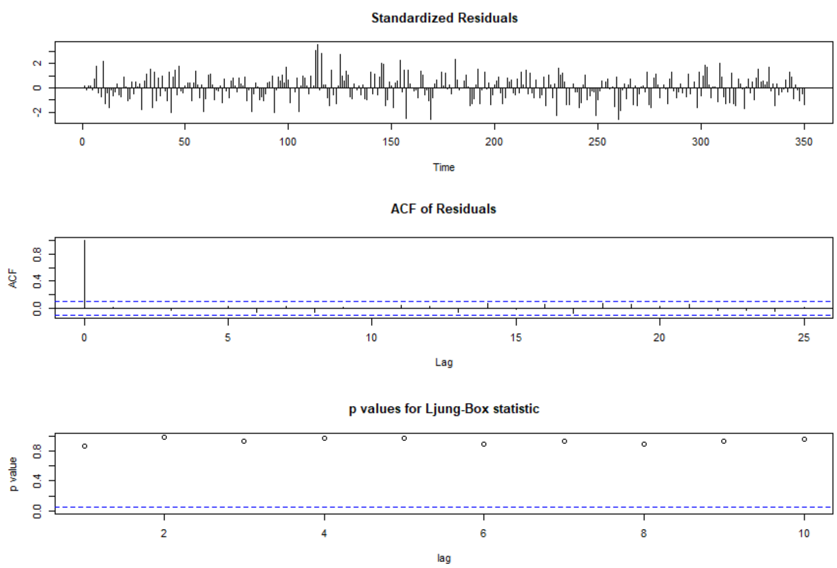


We can see clear sinusoidal decay in the **ACF** plot which indicates that the AR component is of the order 2 or higher. There is a possibility that it is masking some MA peaks, also that we have an AR process that has big negative coefficient(s). There is a first significant value at lag 1 and 2 at **PACF** plot, and then at lag 3, 4, and 5 which might be significant by chance. We can expect around 5% of false significant values.
The auto.arima() R function also recommends the ARMA (1,4) model, which seems plausible according to plots.

AIC 10x10 matrix gave us that the smallest AIC value has model ARMA(5,1), but that can be due a large number of predictors (models with large number of predictors have smaller AIC values). Some models did not converge in the default number of iterations of ML algorithm. Judging by the significance of the last coefficient model ARMA(2,4) has significant coefficient.

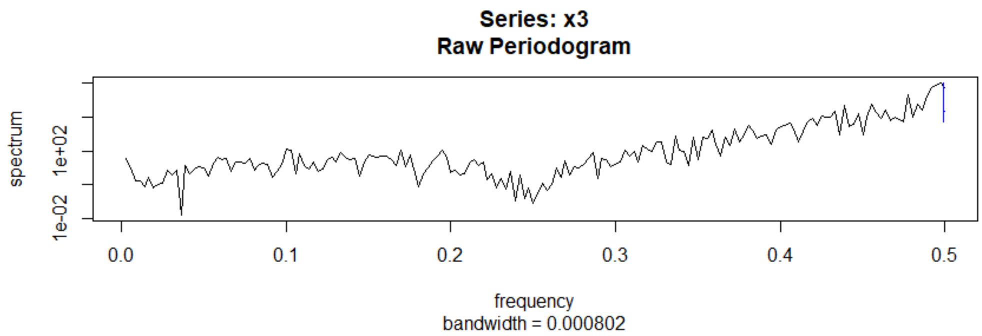


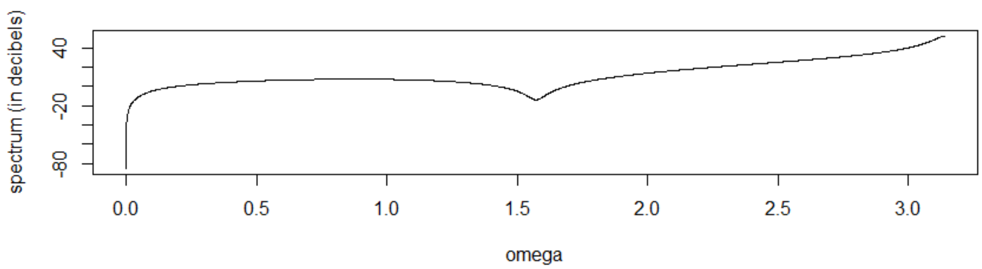
1 - Residual Plot of the model ARMA(2,3)



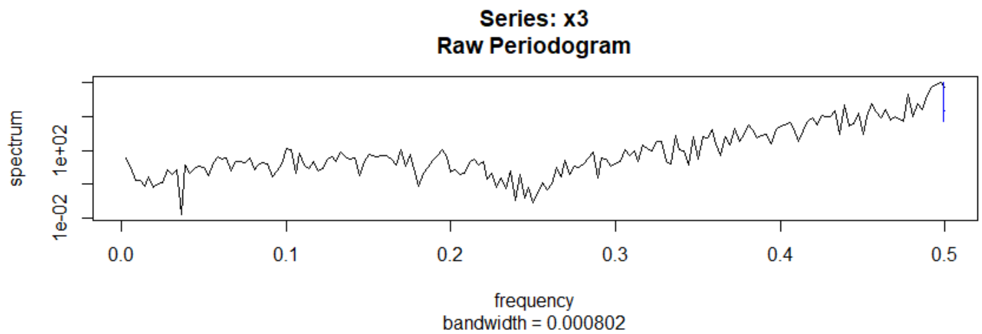
2 - Residual Plot of the model ARMA(2,4)

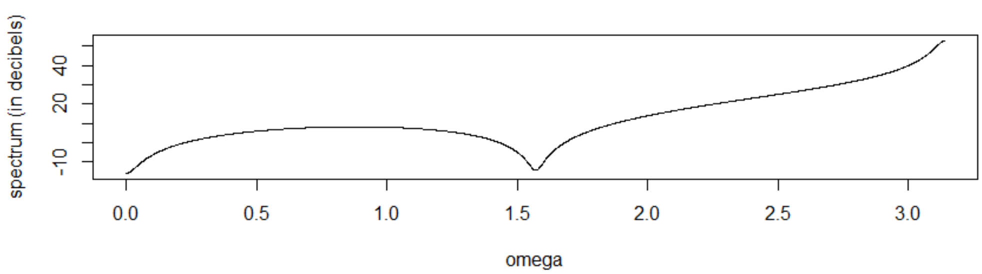
The closer examination of the residuals indicated that the smallest model that had random residuals with nonsignificant ACF values is ARMA(2,4). Curiously, model ARIMA(2,3) had an insignificant AR3 coefficient.





3 - ARIMA(2,4)

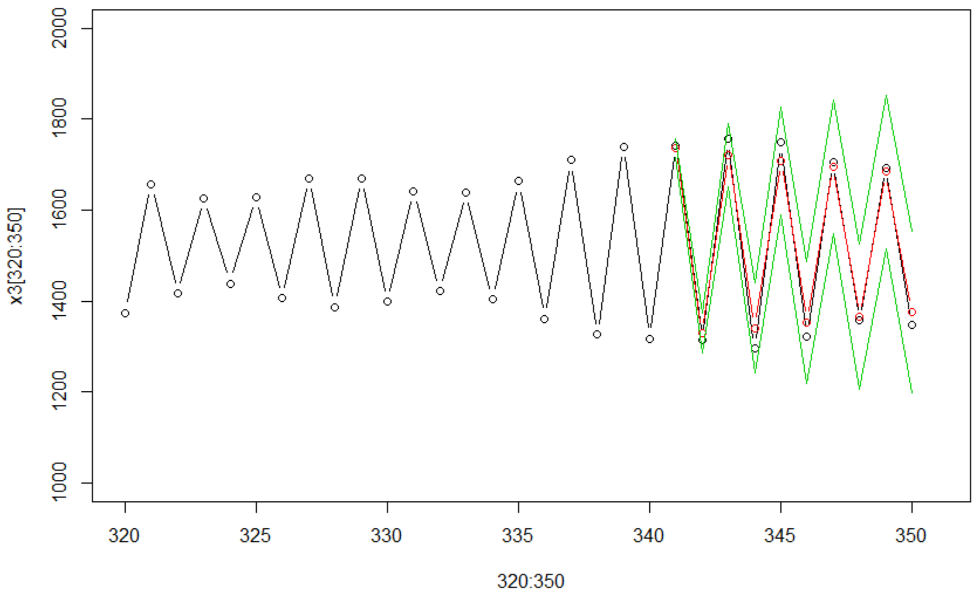




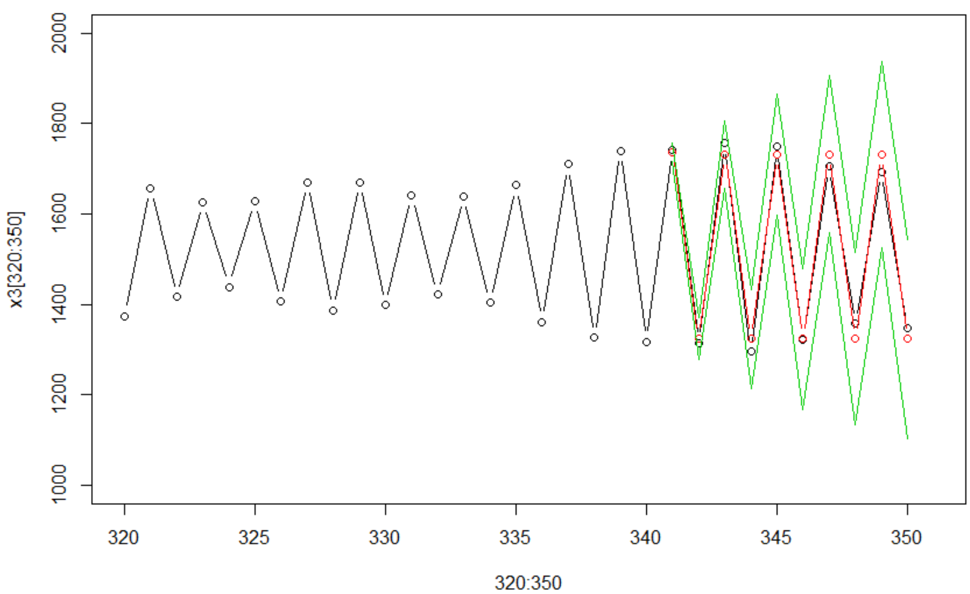
4 - ARMA(2,3)

Period gram compared to spectrogram shows that ARMA(2,3) is better fit.

If we look at the in sample predictions, we can notice that predictions of both models look nice. Since the model ARMA(2,4) has smaller squared prediction error (5709, compared to 7157), it can be our model for the data.



5 - Prediction interval for model ARMA(2,3)



6 - Prediction interval for model ARIMA(2,4)

Executive report – data set 3

By the thrall analysis of the data from DataSet 13we concluded that the model appropriate to model the data is an model produces observations that depend on the previous 2 observations and 4 past error terms, or ARMA(2,4) with coefficients:

ar1	*ar2*	*ma1*	*ma2*	*ma3*	*ma4*	*intercept*
-0.17	0.76	-1.68	1.63	-1.60	0.65	1527.99

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*
1697.27	1717.74	1738.22
1293.65	1336.73	1379.82
1654.74	1724.95	1795.17
1238.89	1338.86	1438.82
1590.11	1710.96	1831.81
1214.35	1352.06	1489.77
1546.47	1698.00	1849.54
1200.91	1364.38	1527.84
1512.26	1685.99	1859.72

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