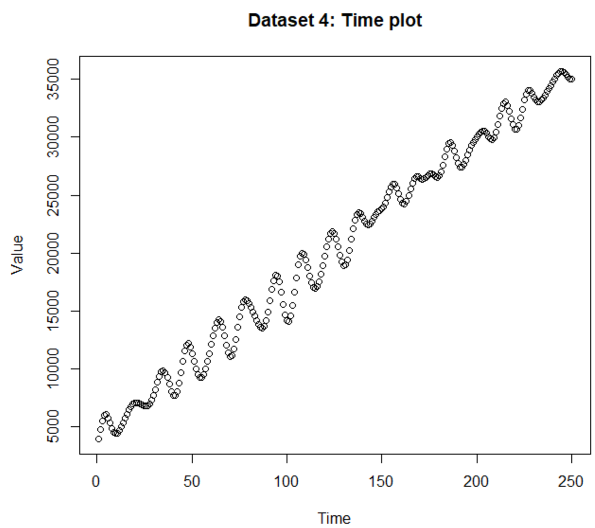
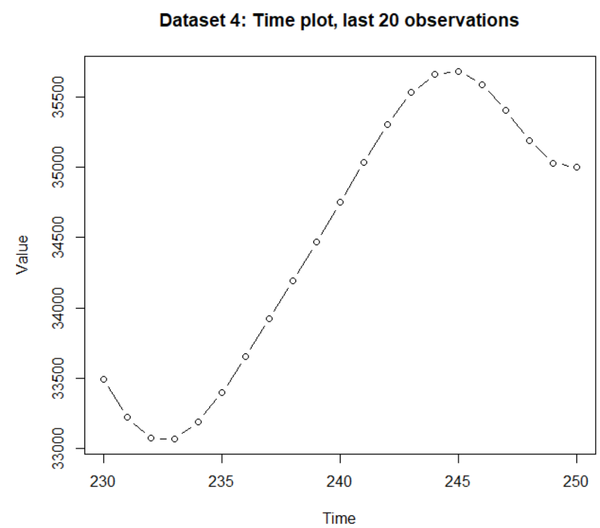
Technical Appendix – data set 4

Summary of the time series, 250 observations,

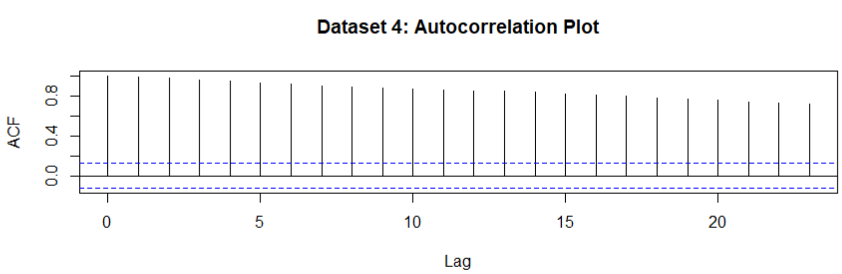
Min.	*1st Qu.*	*Median*	*Mean*	*3rd Qu.*	*Max.*
3967.20	12078.20	20117.70	20211.91	28187.85	35683.60

Time series plots

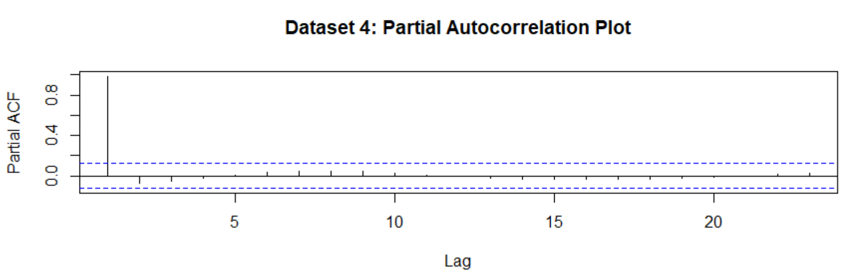




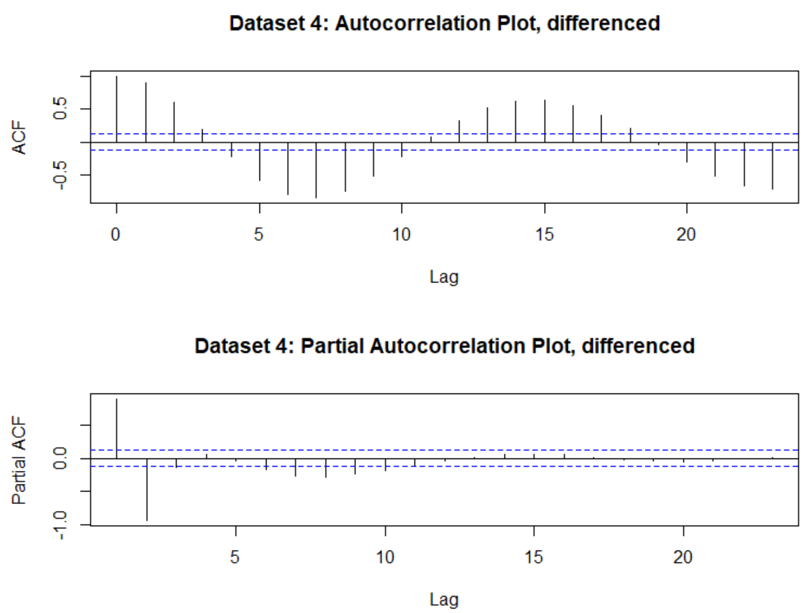
The timeseries plot data set indicates that data is not stationary. There is an obvious trend in data. The Augmented Dickey-Fuller test confirmed that a good starting point for modeling this time series is an ARMA model on the differenced data.



We can see gradual decay in the **ACF** plot which indicates that our data is not stationary.



The auto.arima() R function also recommends the ARI MA(3,1,5) model, which s eems plausible according to plots.

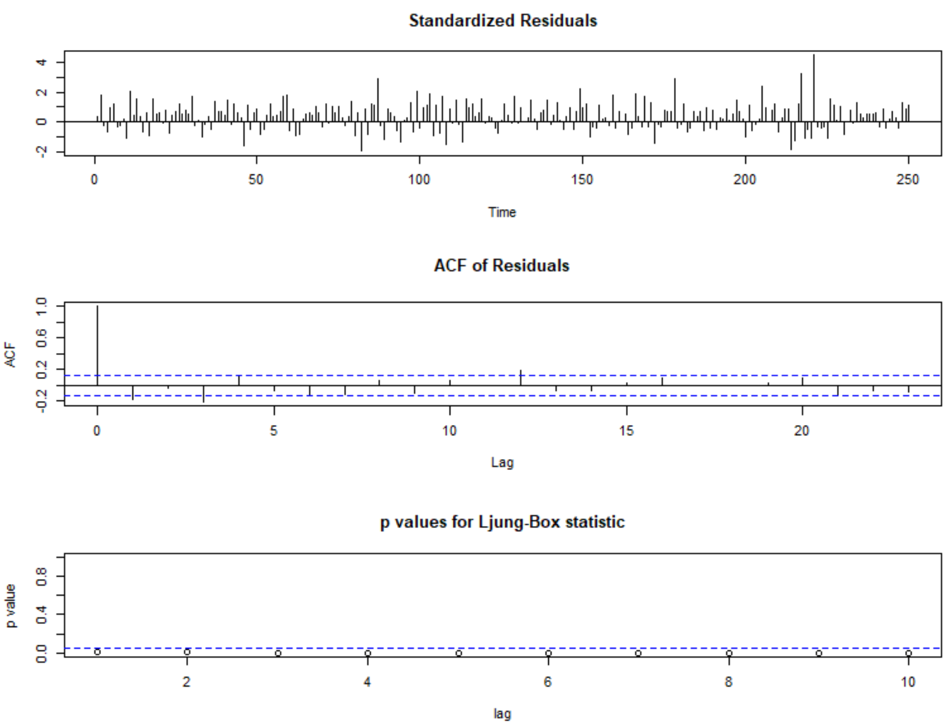


The ACF and PACF plots of differenced data show us that here is a sinusoidal decay in **ACF** plot, that implies that our model has an AR component of order 2 or higher. Also, it might be masking some peaks that would indicate the order of the MA component. The **PACF** plot has first two obvious spikes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

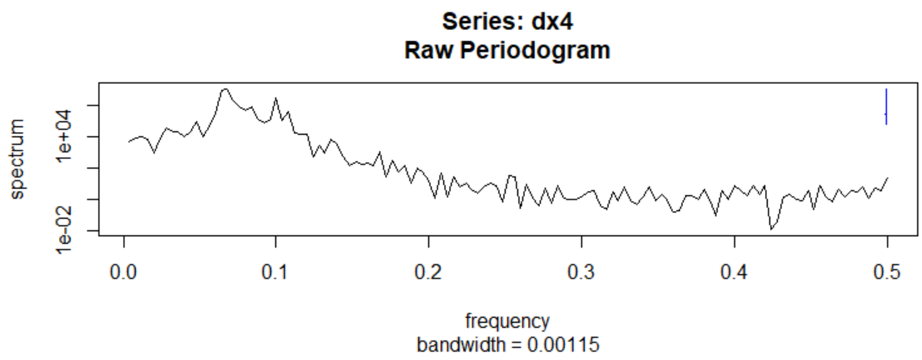
For the differenced data, auto.arima() R function recommends the ARIMA(3,0,5) model, which seems consistent with the previous suggestion.

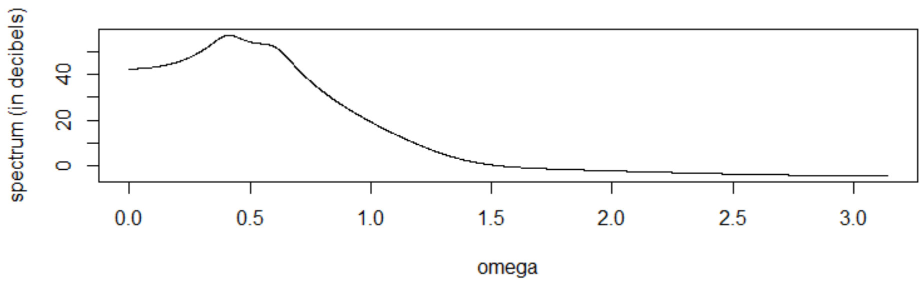
The closer examination of the residuals indicated that the smallest model that had random residuals with nonsignificant ACF values is ARIMA(7,1,3) is the first model that had both random residuals and high p-values. Unfortunately, this model doesn't have MA components significant. All models up to ARIMA(10,1,10) have some kind of issue. Some of them have relatively random residuals, but nonsignificant parameters and vice versa.

The only model that remotely matches spectrogram and periodgam is ARIMA(4,1,3). It has all coefficients significant, random residuals, but small p-values for the Ljung-Box statistics.



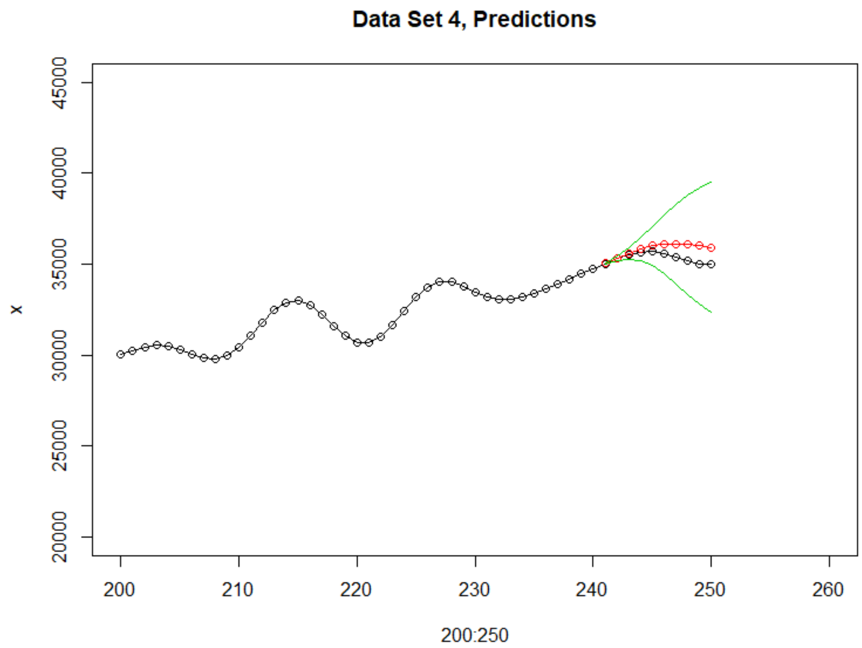
1 - ARIMA (4,1,3)





2 - ARIMA(4,1,3)

If we look at the in sample predictions, and predictions for model ARIMA(4,1,3) look decent. Since the data has obvious sharp trend, it is not wrong to assume that future values will continue with similar trend.



3 - Prediction interval for model ARMA(4,1,3)

Executive report – data set 4

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 4 observations and 3 past error terms, or ARIMA(4,1,3) on once differenced data 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*
35138.54	35162.67	35186.80
35422.25	35524.72	35627.20
35762.07	36033.87	36305.67
36029.61	36586.78	37143.95
36100.40	37058.14	38015.88
35896.40	37337.67	38778.93
35412.13	37362.65	39313.18
34715.91	37135.44	39554.97
33927.30	36720.95	39514.59

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