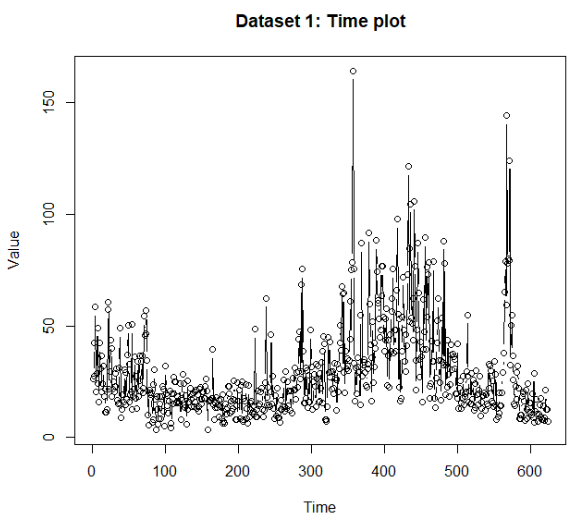
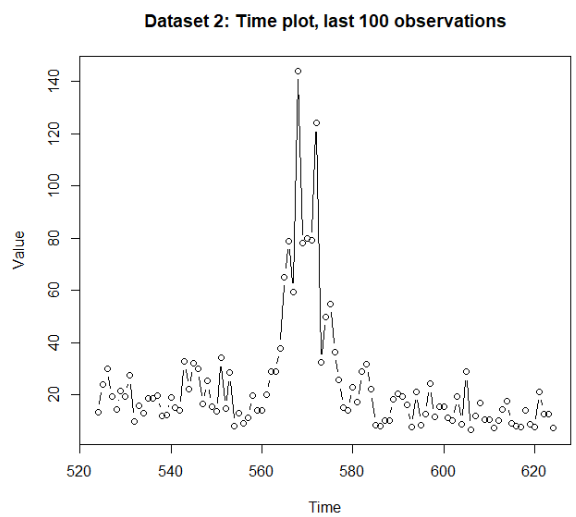
Technical Appendix – data set 5

Summary of the time series, 624 observations

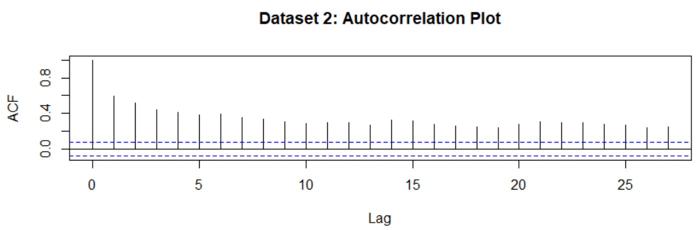
Min.	*1st Qu.*	*Median*	*Mean*	*3rd Qu.*	*Max.*
3.48	15.04	21.40	28.09	34.26	164

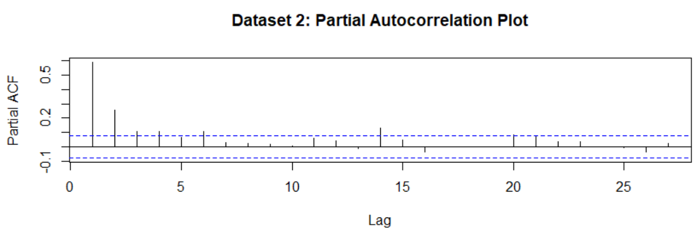
Time series plots



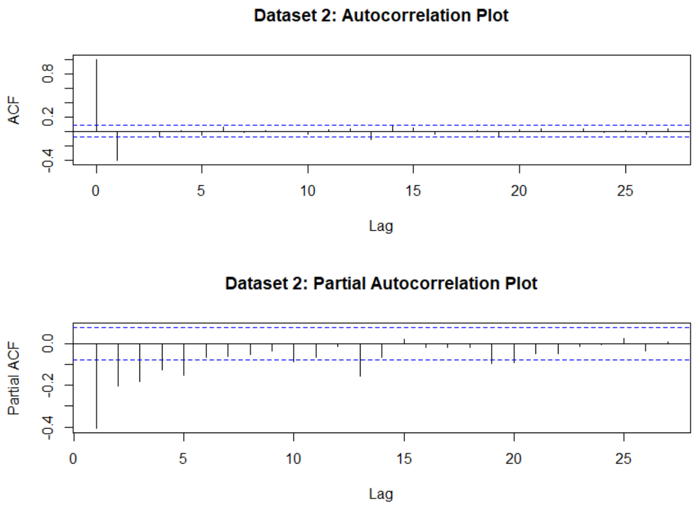


The timeseries plot data set indicates that the stationarity of the data is questionable. There are some unusually high values, but they might be due to the large variance of the noise terms. That might be remedied by the introduction of the log transformation. The Augmented Dickey-Fuller test confirmed that a good starting point for modeling this time series is an ARMA model.





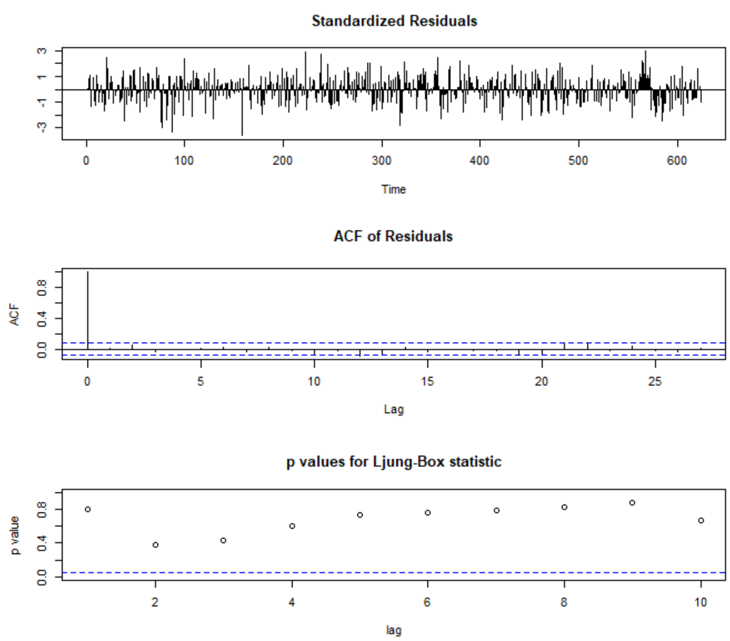
There is a slow gradual decay in **ACF** plot, that implies that our data is not stationary. It might be masking some peaks that would indicate the order of the MA component. The **PACF** plot has first two obvious spikes which indicate MA(2) after which we also have an exponential decay. Therefore, according to the plots possible candidate models are ARMA(0,2), ARMA(1,2), ARMA(2,2), ARMA(3,2)…
Therefore, it is necessary to examine other methods for model selection, and to explore both differenced and the 'raw' data. The auto.arima() R function recommended the ARMA (1,2) which looks plausible.



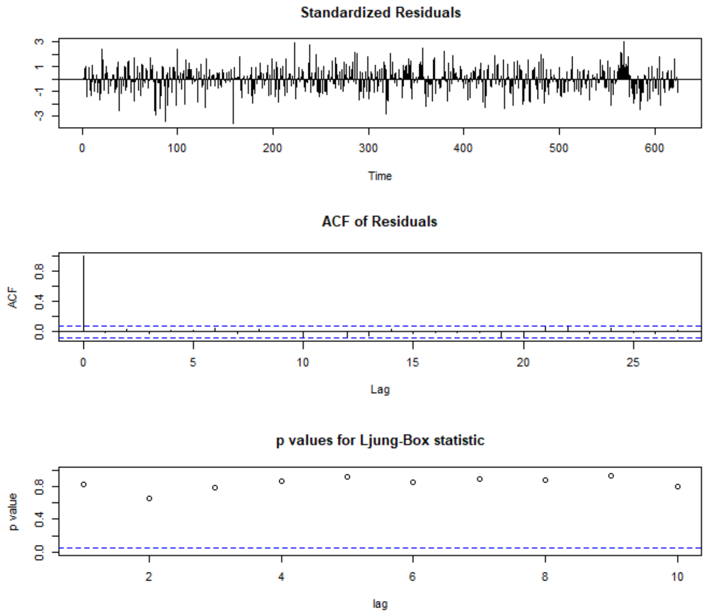
The differenced data shows us different set of ACF and PACF plots. We can notice that there is one significant value in the ACF plot and first four in PACF. That indicates possible models ARIMA(1,1,1), ARIMA(2,1,1), ARIMA(3,1,1) and ARIMA(4,1,1).

Log transformation will give different appearance to our data, but it will not influence the ACF and PACF plots.

Judging by the significance of the last coefficient only ARIMA(0,1,2), ARIMA(1,1,1), ARIMA(1,1,2), ARIMA(1,1,3) have last coefficients significant. Out of them, ARIMA(0,1,2) and ARIMA(1,1,1) have random residuals.



1 - Residual Plot of the model ARIMA(0,1,2)



2 - Residual Plot of the model ARIMA(1,1,1)

The closer examination of the in sample prediction accuracy 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.

	ME	RMSE	MAE	MPE	MAPE
Test set	0.2271763	4.132758	3.422288	-9.356793	30.50123

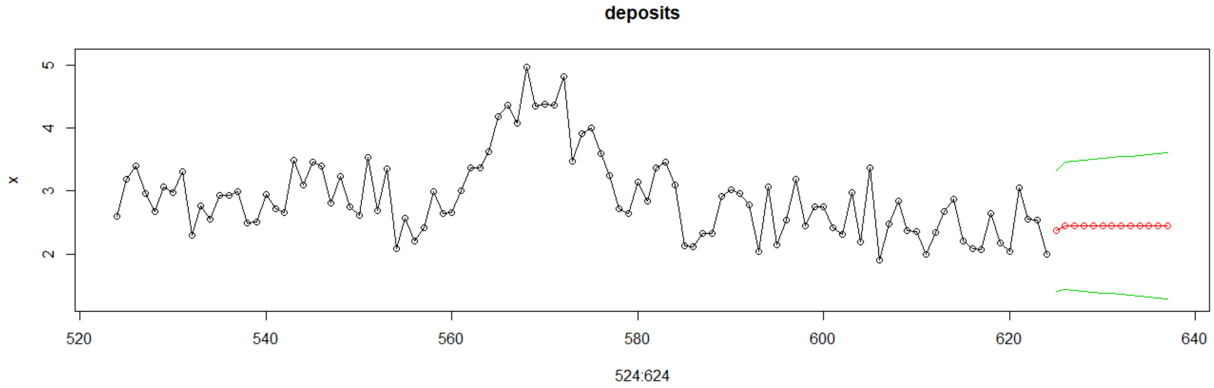
3 – Prediction accuracy of model ARIMA(0,1,2)

	ME	RMSE	MAE	MPE	MAPE
Test set	-0.177578	4.151697	3.455176	-13.28734	31.95136

4 - Prediction accuracy of model ARIMA(1,1,1)

Periodgram compared to spectrogram also shows that ARIMA(0,1,2) is better fit.

In sample predictions of the model ARIMA(1,1,1)



Executive report – data set 5

By the thrall analysis of the data from the Deposits data set we concluded that the model appropriate to model the data is an ARMA(1,1,1) model, produces observations that depend on the previous observation and 1 past error term, and it is differenced onetime, with coefficients:

ar1	*ma1*
0.23	-0.88

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*
3.86	10.03	26.09
3.92	10.77	29.62
3.91	10.95	30.63
3.88	10.99	31.11
3.85	11.00	31.46
3.81	11.00	31.78
3.77	11.00	32.10
3.73	11.00	32.42
3.70	11.00	32.74

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