Time Series Analysis
sf2943

Project Report

Yiming Fan
yimingf@kth.se

In collaboration with

Jie Deng
jied@kth.se

Problem 1: White noise.

1. Tests on different n and h

We have tried various $h = \{2, 4, 6, 8, 10\}$ and $n = \{500, 1000, 2000\}$ and all iid samples passed the *one-sided Kolmogorov-Smirnov* test and the *Ljung-Box* test. We have also tested sample autocorrelations $\hat{\rho}(h)$ for iid samples on Q-Q plot, and the Q-Q plot confirm that the $\hat{\rho}(h)$ for different h are approximately independent and have a Gaussian distribution with mean zero and variance $1/n$. Samples simulated from Low-order AR and MA process with small coefficients tend to be independently normal distributed.

2. Tests on ARMAx processes

Below shows the ACF of an **AR(1)** process. By testing **MA(1)** and **AR(1)** processes on various n we reached our conclusion:

1. *One-sided Kolmogorov-Smirnov* test is more sensitive than *Ljung-Box* test.

2. We computed *sample autocorrelation function* with lag of 40, and counted values that beyond the confidence bound. If more than 3 values out of bounds, we reject iid for data. Otherwise, we cannot deny. This method is not accurate sometimes depending on the sample data generation and sample size.

3. *The turning point test* is another test method we tried. Same as *sample autocorrelation function*, not as stable and sensitive as *One-sided Kolmogorov-Smirnov* test and *Ljung-Box* test.

4. Increasing the size of the data could increase the testing sensitivity.

5. **MA(1)** processes are harder than **AR(1)** processes for both tests to detect the deviations.

3. Some explanations on applied `MATLAB` functions

1. *autocorr(y, h)*: this `MATLAB` function takes sample data y and lag h as input, and plots the graph of sample ACF with confidence bounds.

2. *kstest(x)*: takes data x for statistical test of null hypothesis that the data is a standard normal distribution, return decision of 1 as test rejects the null hypothesis, or 0 as fails to reject.

3. *arima*: this function creates a model in process of AR, MA, ARMA etc with input of specification of certain processes with corresponding values of coefficients, and white noise of certain variance. In our case, we created an AR(1) process with WN(1, 0.01).

4. *simulate(mdl, n)*: takes the model object mdl and size n as input, this function will give a sample data generated from the model with specified size n.

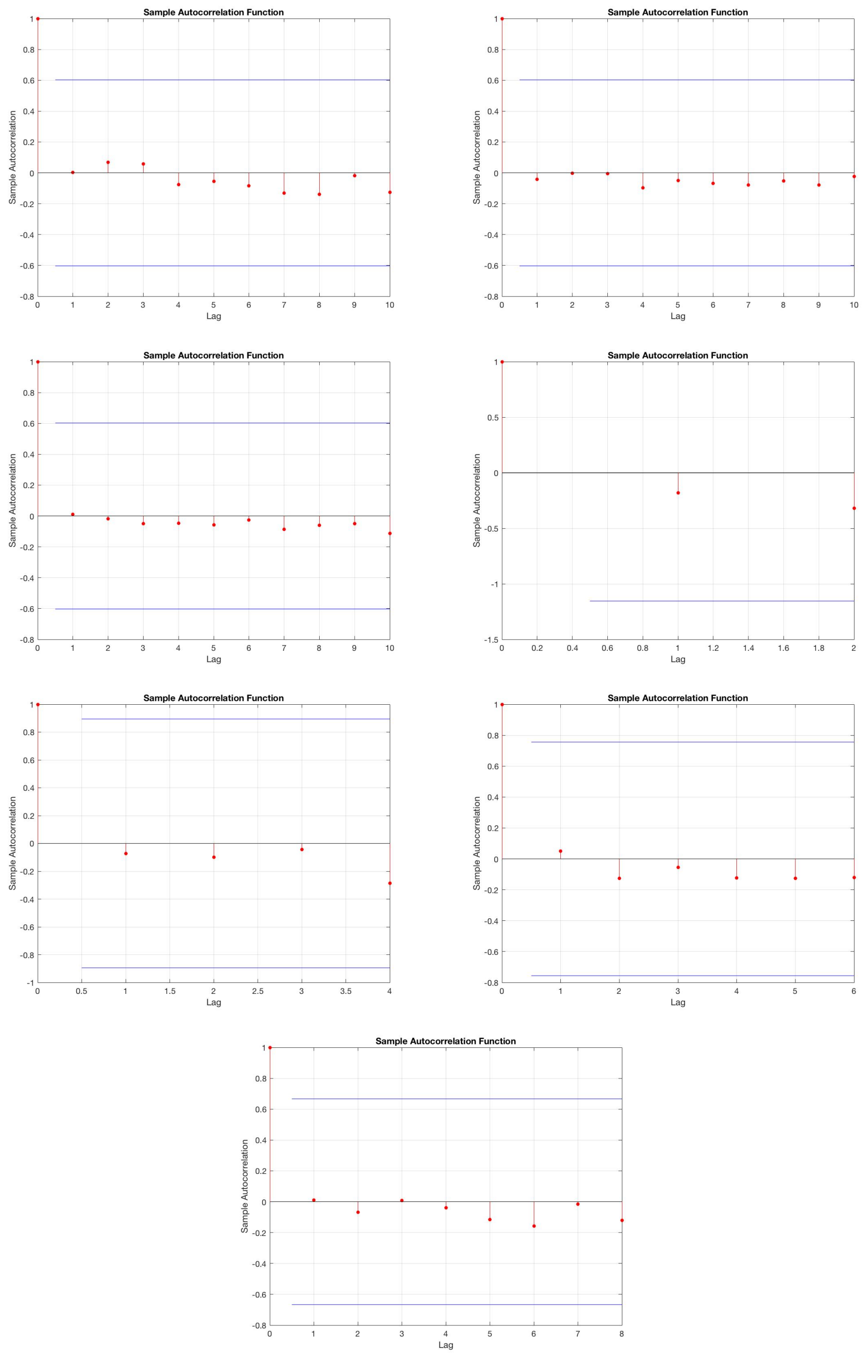

Figure 1: ACFs on $h = \{2, 4, 6, 8, 10\}$ and $n = \{500, 1000, 2000\}$.

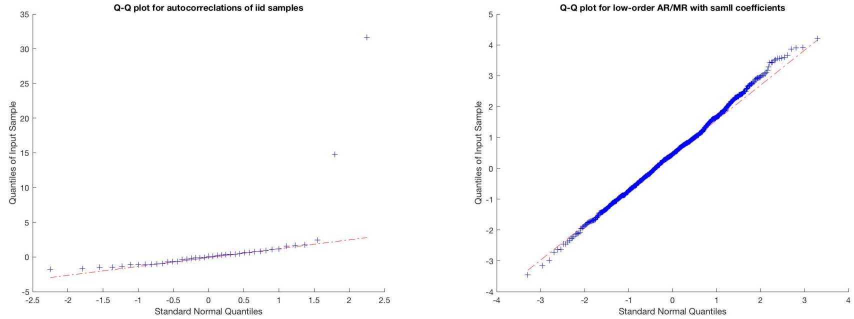

Figure 2: Q-Q plot of .

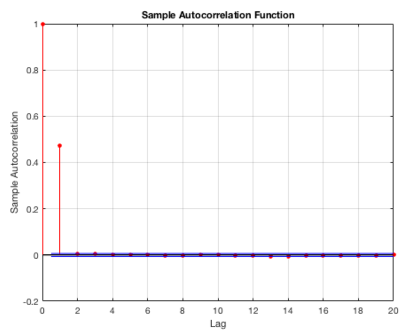

Figure 3: ACFs on $AR(1)$ process with $n = 100000$.

Problem 2. Parameter estimation.

1. Scatter plots

We have tried $m = 50$ and $n = 200$ and scatter plots of the estimated parameters within MLE / Yule- Walker methods are shown.

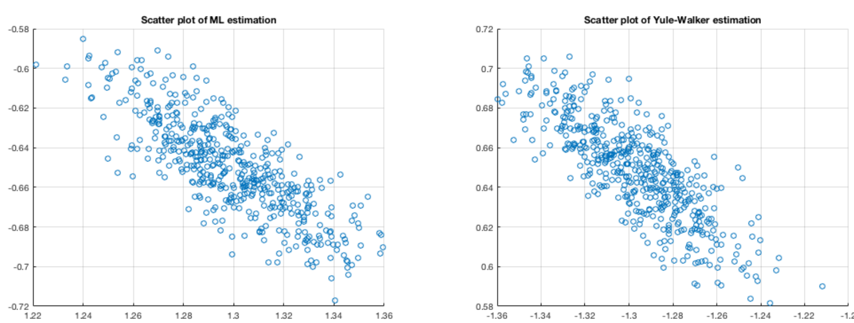

Figure 4: Scatter plots.

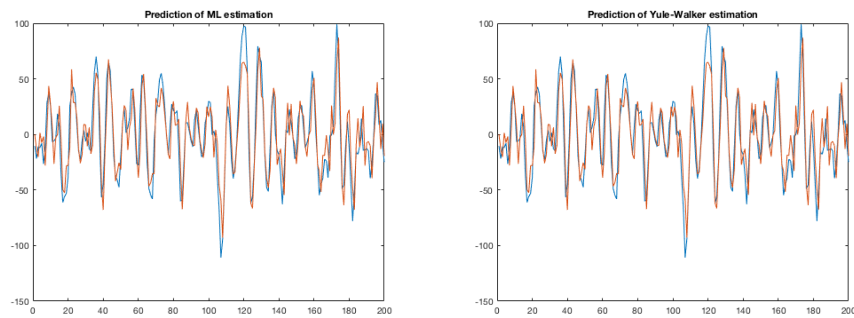

Figure 5: Predictions comparing with the sampled data.

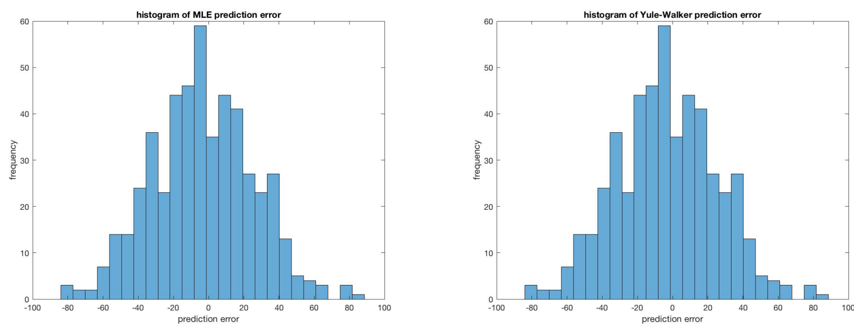

Figure 6: histogram of prediction error for MLE and Yule-Walker.

2. Accuracy, prediction and residuals

Taking the mean values of the estimated parameters we saw that MLE reached somehow closer to the correct parameter values than Yule-Walker did. Hence we conclude that **MLE is more accurate**.
We have written a `prediction` function dedicated for this problem only. By checking the mean-squared error w.r.t. the sampled data we conclude that **MLE has more accurate predictions**. Some comparing histograms are shown below.

A `residual` function calculating the residuals according to the fitted models is also written. Using the sampled data and fitted models above we have found that MLE has

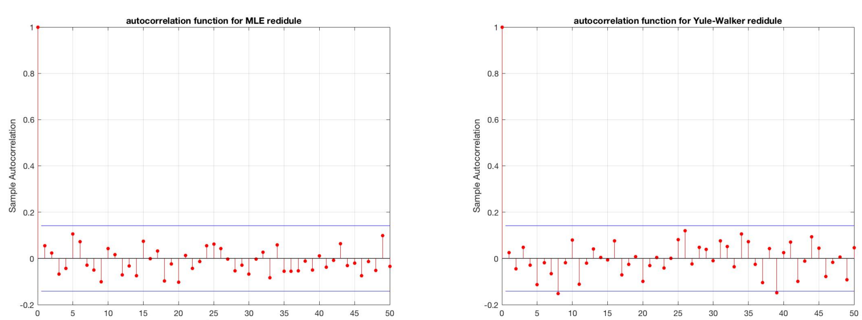

Figure 7: autocorrelation function for prediction residual of MLE and Yule-Walker.

Steps	#data	**AR(2)**	**AR(10)**
1	500	283.86	279.00
1	1000	279.53	277.28
3	500	335.95	339.38
3	100	334.00	354.39
3	300	337.33	345.32

Table 1: Prediction errors.

the mean sample variance of 281.1992, and Yule-Walker 281.3552, which are close to 280 the original white noise variance. Both models have validated that the residuals are uncorrelated and have the same variance as the original sampled data. However **MLE is slightly more accurate**.
We did ks-test and plotted ACF graph for prediction residual for MLE and Yule-Walker estimation, and both indicated that the residuals from the fitted model are uncorrelated and have the same distribution as the original noise sequence $\{Z_t\}$.

3. Some explanations on applied `MATLAB` functions

arma mle, aryule, prediction, residual

Problem 3: Prediction.

a) Predictions on AR(2) model.

The one-step and 3-step prediction errors are shown on the table below.
Some conclusions:

1. Overall speaking, one-step predictions are more accurate than 3-step predictions.
2. Overall speaking, prediction errors decrease as the dataset size increases.
3. **AR(1)** does better than **AR(10)** in one-step predictions but worse in 3-step predictions.

b) Prediction on AR(1) model.

Since **AR(1)** can be written as **MA(∞)** model we use ϕ^j ($j = 1, 2, \ldots 10$) as the parameters of **MA(10)** models (that's also why sometimes we attempt to use an *MA* model for **AR(1)** process). The table below shows the one-step and 3-step prediction errors:

c) Prediction on non-zero mean model.

We subtracted the dataset sample with their mean value and did the similar estimating process. The scatter plot of the estimated **AR(1)** parameters and the variances is shown

Steps	#data	**AR(1)**	**MA(10)**
1	100	0.9995	5.4400
1	500	0.9885	5.4320
1	2000	1.0021	5.5016
3	100	1.0346	5.2581
3	500	1.0672	5.6175
3	2000	1.0697	5.3958

Table 2: Prediction errors.

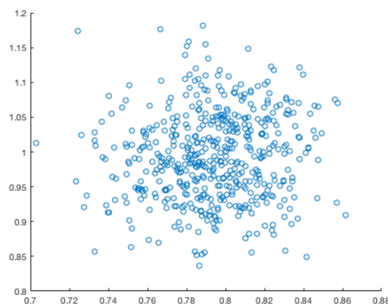

Figure 8: Scatter plot of the estimated **AR(1)** parameters and the variances.

below. The mean of those estimated parameters (**AR(1)** parameter 0.7905 and variance 1.0017) are close to the original values.

Below is a table showing the performance comparisons between zero and non-zero mean models in one-step predictions. Those fitted models have similar performance which shows that adding a constant term will not have any impact on the fitting performance.

Problem 4: Model selection.

1) Dataset 1.

For Dataset 1, we found trend component in the plot of raw data. we first subtracted data mean and plotted graph of ACF with lag of 100, a seasonal component with period of 12 was shown. A trend component was found after deseasonalization of the data. Finally, both trend and seasonal components were eliminated by differencing, we fitted the data into

Steps	#data	zero mean **AR(1)**	non-zero mean **AR(1)**
1	100	0.9995	0.9499
1	500	0.9885	0.9986
1	2000	1.0021	1.0065

Table 3: Performance comparison on one-step predictions.

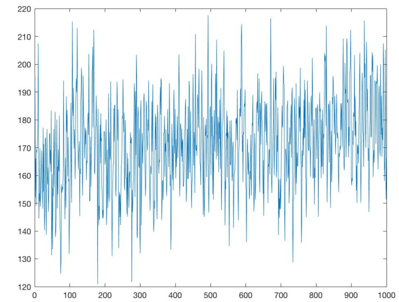

Figure 9: Dataset 1.

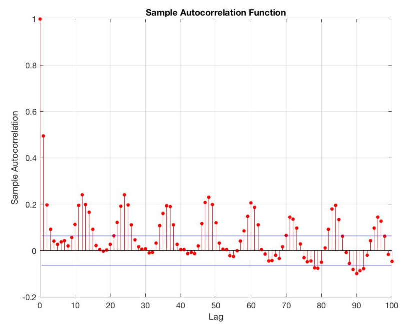

Figure 10: Seasonal component in ACFs for Dataset 1.

AR/MA/ARMA models and found that best model is ARMA(3, 3) with parameter as below:

$$X_t + 1.72X_{t-1} - 1.02X_{t-2} + 0.19X_{t-3} = Z_t - 1.75Z_{t-1} + 0.65Z_{t-2} + 0.11Z_{t-3}, \{Z_t\} \sim \mathcal{WN}(0, 1) \tag{1}$$

2) Dataset 2.

The same analysis was made for dataset 2, both trend and seasonal components were identified and eliminated by differencing. We plotted ACF and PACF, and found from the graph

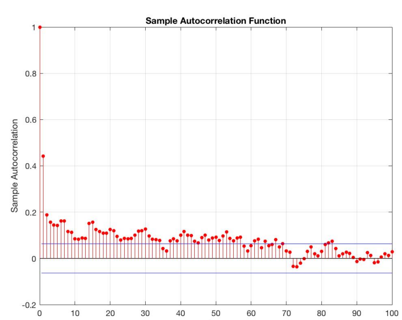

Figure 11: Trend in ACFs for Dataset 1.

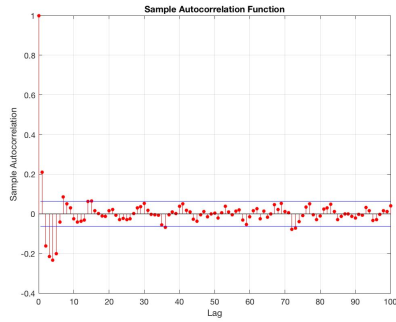

Figure 12: ACFs without trend and seasonal components for Dataset 1.

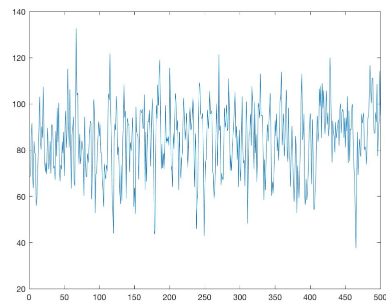

Figure 13: Dataset 2.

that the stationary sequence of sample data was best fitted into model AR(1) with parameter as below:

$$X_t + 0.42X_{t-1} = Z_t, \{Z_t\} \sim \mathcal{WN}(0, 1) \tag{2}$$

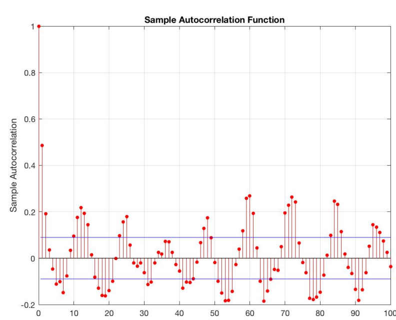

Figure 14: ACFs for Dataset 2.

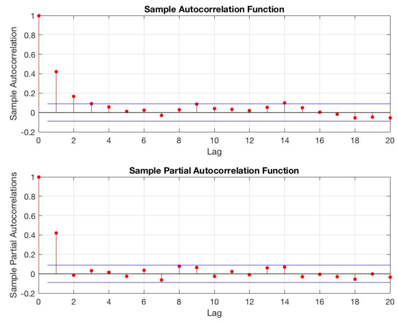

Figure 15: ACF and PACF for Dataset 2 after stationalization.

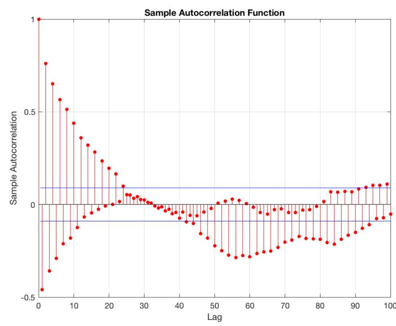

Figure 16: ACFs for Dataset 3.

3) Dataset 3.

After plotting ACF graph for dataset 3, we found that it was similar to textbook figure 3-3. Therefore, it is AR(2) model, the parameter estimated as follows:

$$X_t - 0.14X_{t-1} + 0.69X_{t-2} = Z_t, \{Z_t\} \sim \mathcal{WN}(0, 1) \tag{3}$$

4) Dataset 4.

After transferring the dataset to zero mean, we plotted graph for raw data and ACF graph and found trend as shown in figures. By stationalizing it and plotting ACF and PACF, the best model fitted and identified for the dataset is AR(2) with estimated parameter as below :

$$X_t + 1.12X_{t-1} - 0.52X_{t-2} = Z_t, \{Z_t\} \sim \mathcal{WN}(0, 1) \tag{4}$$

5) Dataset 5.

From the graph plotted for dataset 5, we could see that there were trend of going up then going down. After removing the trend component, we fitted it to different model and found

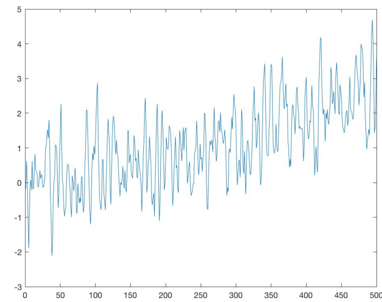

Figure 17: Dataset 3.

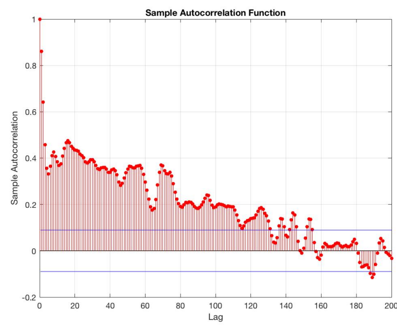

Figure 18: ACFs for Dataset 4.

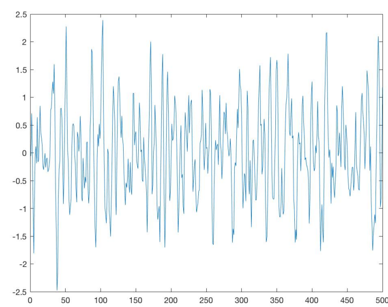

Figure 19: graph for Dataset 4 after stationalization.

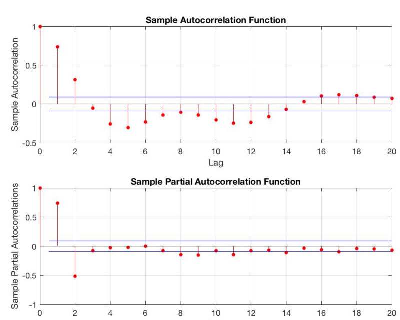

Figure 20: ACF and PACF for Dataset 4 after stationalization.

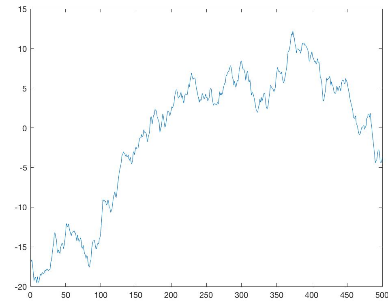

Figure 21: graph for Dataset 5.

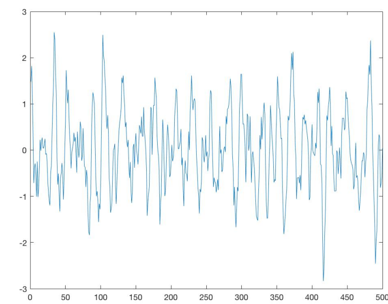

Figure 22: stationalized graph for Dataset 5.

that the best model for this dataset is ARMA(3, 1) with estimated parameter as below :

$$X_t + 1.93X_{t-1} - 1.14X_{t-2} + 0.15X_{t-3} = Z_t - 0.91Z_{t-1}, \{Z_t\} \sim \mathcal{WN}(0, 1) \tag{5}$$

6) Dataset 6.

No trend or seasonal components were identified in graph for dataset 6. We tried different models and found that best model for this dataset is ARMA(1, 2). But as estimated pa-

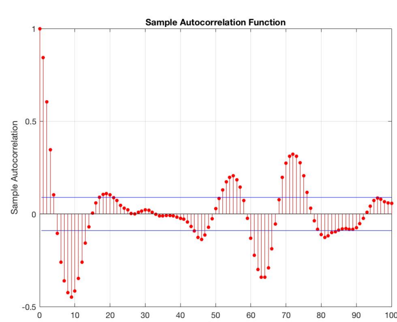

Figure 23: ACFs for Dataset 5.

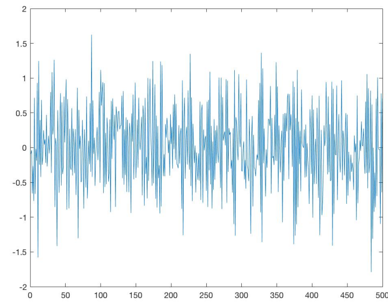

Figure 24: graph for Dataset 6.

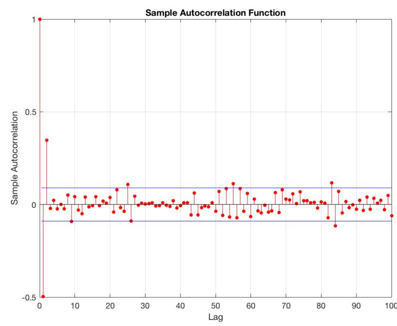

Figure 25: ACFs for Dataset 6.

rameter for AR model is around 0.03, we thought it could be fitted in MA(2) model. And the ACF graph indicates as well that MA(2) model fit for data. The estimated parameter as below :

$$X_t = Z_t - 0.52Z_{t-1} + 0.54Z_{t-2}, \{Z_t\} \sim \mathcal{WN}(0, 1) \tag{6}$$

Problem 5: Model selection for self chosen dataset.

We selected data of 'CO2 (ppm) mauna loa, 1965-1980' for analysis. The trend and seasonal components can be easily seen from both data plot and ACF plot. After elimination to make the data stationary, we identified best model according to the ACF graph for this dataset is MA(2) with parameters as below:

$$X_t = Z_t + 0.52Z_{t-1} + 0.35Z_{t-2}, \{Z_t\} \sim \mathcal{WN}(0, 1) \tag{7}$$

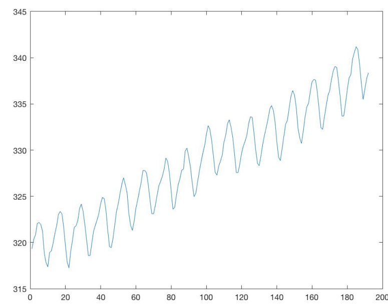

Figure 26: 'CO2 (ppm) mauna loa, 1965-1980' data.

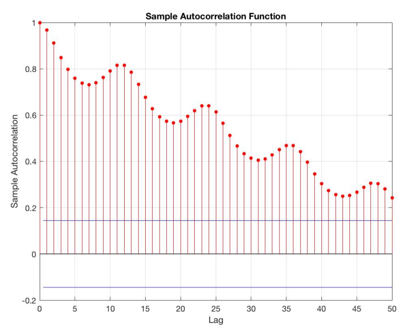

Figure 27: ACFs for 'CO2 (ppm) mauna loa, 1965-1980' data.

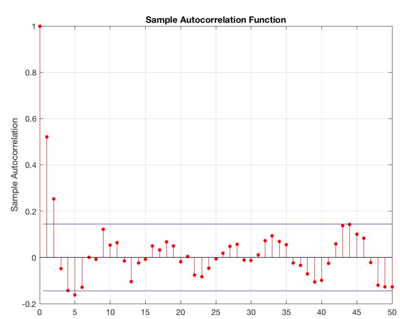

Figure 28: ACFs after trend and seasonal elimination for 'CO2 (ppm) mauna loa, 1965-1980' data.

Approach	buying on 2017/05/04	selling on 2017/05/04
Naive	1633.99	1498.60 - 1762.90
GARCH	1633.99	1576.40 - 1686.00

Table 4: 0.05-quantiles of one day investment on 2017/05/04.

Approach	buying on 2011/08/18	selling on 10 days
GARCH	893.08	630.37 - 1208.00

Table 5: 0.05-quantiles of ten days investment on 2011/08/18.

Problem 6: GARCH models.

a) One day predictions on OMXS30.

We used OMXS30 index from 2017/01/01 to 2017/05/03 to fit the model of GARCH(2, 2) and naive approach. According to the index of 1633.99 on 2017/05/03, we have 0.05-quantiles investment as shown in table. It can be seen that the GARCH process confidence interval is more accurate than naive approach.

b) Ten days predictions on OMXS30.

i) 2011-08-18.

We used OMXS30 index from 2010/08/25 to 2011/08/18 to fit the model of GARCH(2, 2). According to the index of 893.08 on 2011/08/18, we have 0.05-quantiles investment as shown in the table.

ii) 2014-04-24.

We used OMXS30 index from 2013/04/25 to 2014/04/24, totally 250 business days to fit the model of GARCH(2, 2). According to the index of 1365.27 on 2014/04/24, we have 0.05-quantiles investment as shown in the table.

Approach	buying on 2014/04/24	selling on 10 days
GARCH	1365.27	1135.90 - 1609.60

Table 6: 0.05-quantiles of ten days investment on 2014/04/24.

Approach	buying on 2015-06-18	selling on 10 days
GARCH	1562.42	1289.40 - 1823.30

Table 7: 0.05-quantiles of ten days investment on 2015-06-18.

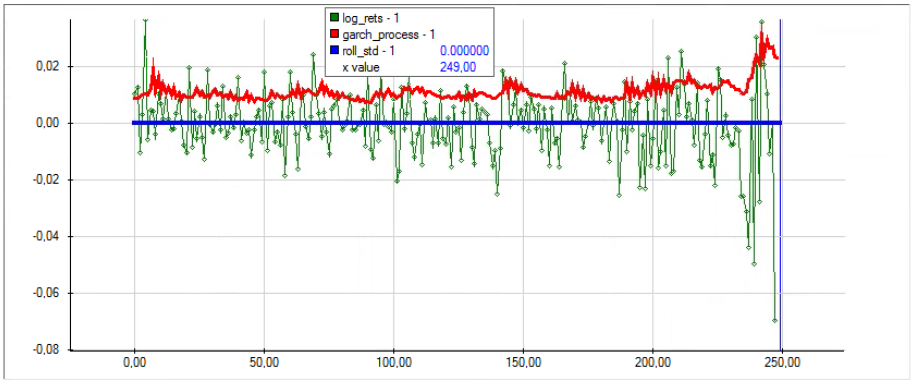

Figure 29: log-returns and GARCH graph for 2011-08-18.

iii) 2015-06-18.

We used OMXS30 index from 2013/04/25 to 2015-06-18, totally 250 business days to fit the model of GARCH(2, 2). According to the index of 1562.42 on 2015-06-18, we have 0.05-quantiles investment as shown in the table.

c) Graphs of long-return and GARCH.

We have plotted graphs for the three 10-day investments above. It can be seen that the GARCH processes plotted in red are the same trend as log-returns plotted in green. And the log-return and GARCH shows stock market turmoil in 2011 which corresponds to the larger 0.05-quantile in section b, and smooth stock market in 2014 and 2015 that match smaller 0.05-quantile.

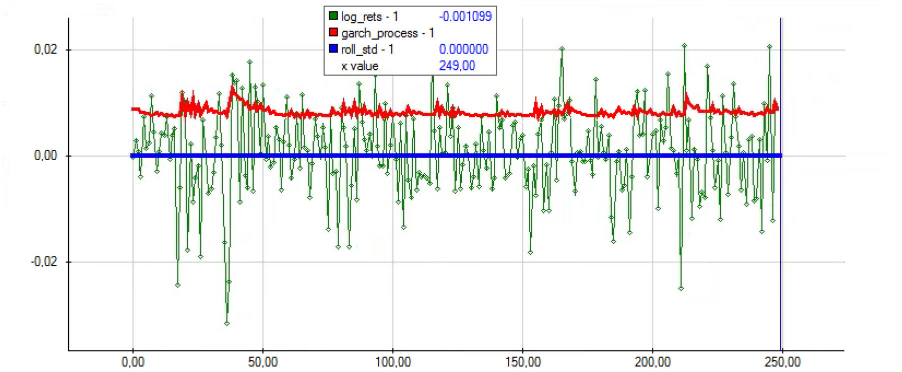

Figure 30: log-returns and GARCH graph for 2014-04-24.

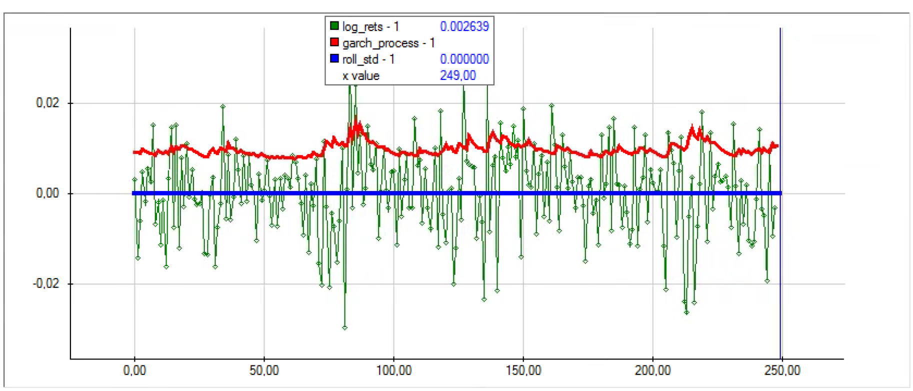

Figure 31: log-returns and GARCH graph for 2015-06-18.