Time Series - Assignment 1

Gabriela Zemenčíková

June 3, 2024

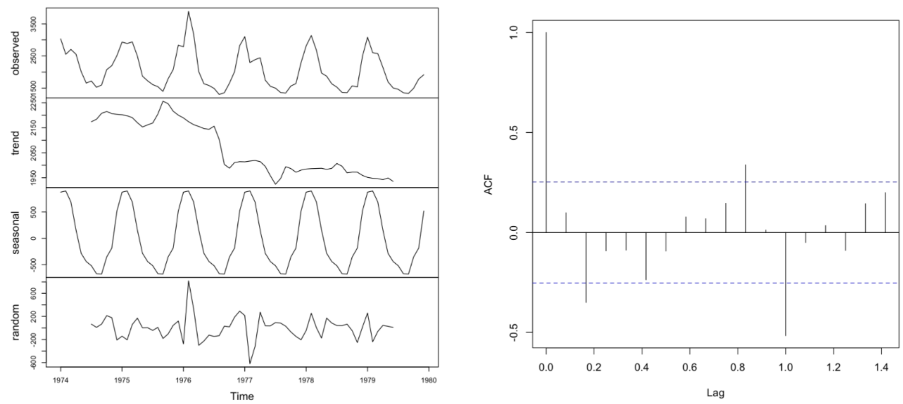
Read the file ldeaths in the folder datasets of R. Make the graphical representation. Identify and estimate the trend, the seasonal component and the residual component. Are the residuals a sample of an IID noise?

```r
require(stats)
require(graphics)
# Graphical representation
#plot(ldeaths, main="Deaths from Lung Diseases", xlab="Time",
ylab="Deaths", type="l")
# Decompose the time series using only moving average methods
decomp <- decompose(ldeaths) # Plot decomposed components
plot(decomp)
# Access the decomposed components
trend <- decomp$trend
seasonal <- decomp$seasonal
residual <- decomp$random
# Plot autocorrelation function (ACF) of residuals, handling
    missing values
na.omit_resid <- na.omit(residual)
acf(na.omit_resid, main = "ACF of Residuals")
# Verify the IID character
Box.test(residual, lag=log(length(residual)), type=c("Ljung-Box"))
```

Listing 1: TASK 1

Other scripts will not be displayed as they were too long. Please see the notebook.

With the plot Decomposition of additive time series, we obtain the graphical representation of the data and then decomposed parts of the observed values, i.e. trend, seasonality and residuals. Right away from the observed plot, we see that a seasonality is present (cycle/year). Next, there is a negative overall trend present in the time series. The decomposition of time series is displayed in Figure 1a. To confirm the residuals are IID, we use the Ljung-Box test which is a method that tests for the absense of autocorrelation in residuals. If the p-value > 0.05, this implies that the residuals of the data are are independent. Since in our example, we have p-value $= 0.05827$ it doesn't suggest any evidence against the null hypothesis of independence in the residuals. In other words, it means that there does not seem to be significant autocorrelation present in the residuals but with a low confidence.



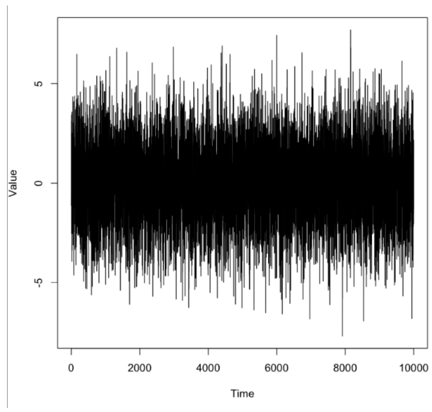
(a) Decomposition of additive time series

(b) ACF of residuals

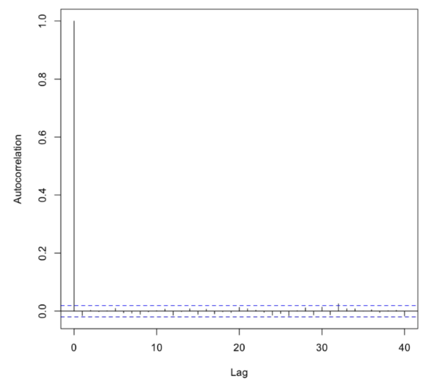
Figure 1: Decomposition & ACF plot

Simulate a Gaussian white noise of n = 10.000 data. Verify by testing that it is an IID noise and a Gaussian white noise. Simulate a Gaussian Random Walk. Simulate IID noises of 10.000 data that are not a Gaussian white noise: a Poisson noise and an exponential noise. Test all what you can.

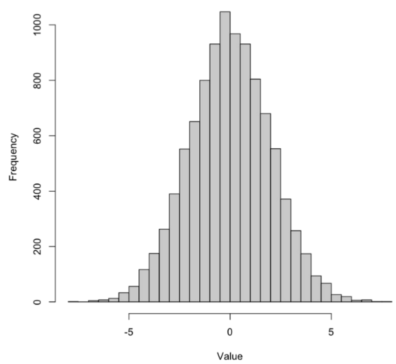
After generating 10000 sample, we check the properties of the data. Starting by plotting the sample, histogram and autocorrelation function, we can observe that the correlation is close to zero which suggests IID Gausian white noise. With a p-value of 0.7926, which is much greater than 0.05, we do not have enough evidence to reject the null hypothesis. Therefore, we cannot conclude that there is significant autocorrelation in the series, which only further confirms the hypothesis. In other words, this suggests that the autocorrelations in simulated Gaussian white noise series are not statistically different from zero at the given significance level. Thus, the series exhibits characteristics consistent with white noise, where each value is independent of others. To support our evidence, we use Q-Q-plot which shows a linear trend further supporting the hypothesis. All the figures are present in Figure 2.



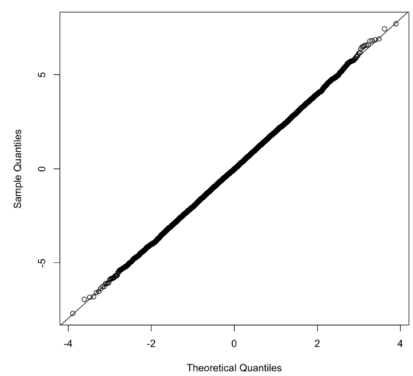
(a) Sample of Gaussian White Noise



(b) ACF of residuals of Gaussian White Noise



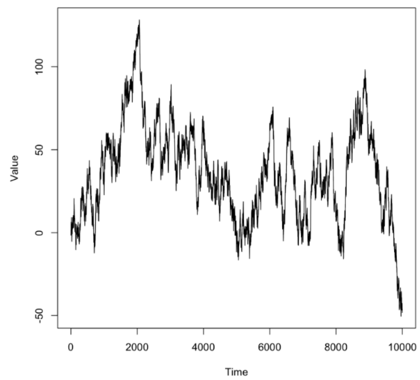
(c) Histogram of Gaussian White Noise



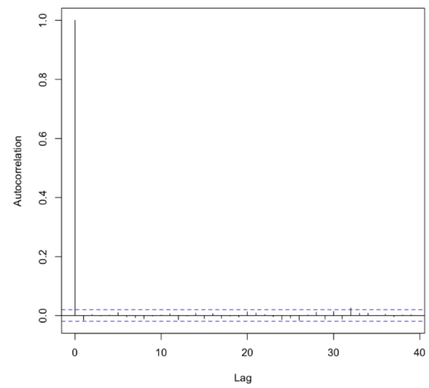
(d) Normal Q-Q Plot

Figure 2: Plots for Gaussian White Noise

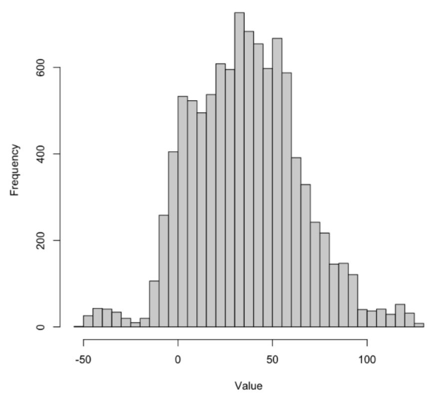
We use our previous result to simulate the Gaussian Random Walk. We take the cumulative sum of the data. Then to see the IID, we plot the correlation function with differenced data. We have that the differenced has no correlation as expected. Same plots are displayed for Gaussian Random Walk in Figure 3. From the ACF plot used for the differenced data we can observe that it is IID.



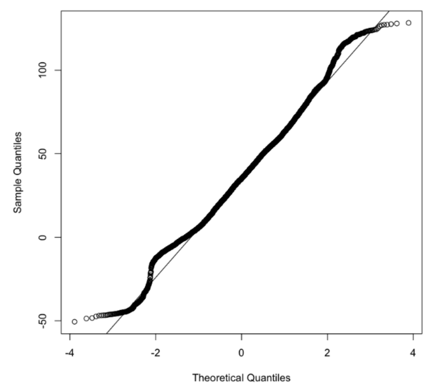
(a) Sample of Gaussian Random Walk



(b) ACF of residuals of Gaussian Random Walk



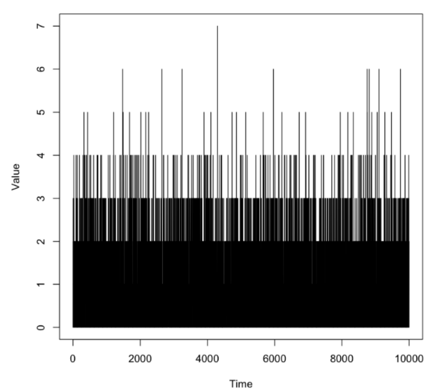
(c) Histogram of Gaussian Random Walk



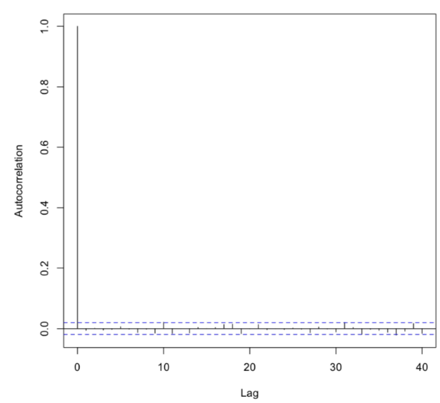
(d) Normal Q-Q Plot

Figure 3: Plots for Gaussian Random Walk

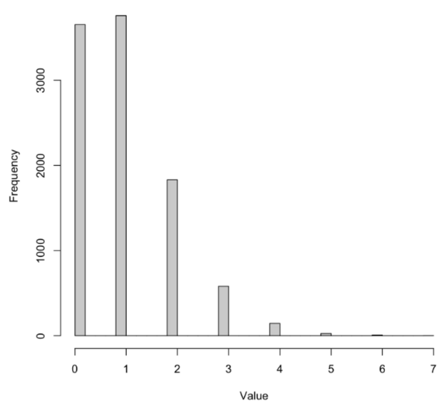
Lastly, we performed the same strategy with exponential and poisson noises, Figure 4 and Figure 5 respectively. With the Box- Ljung test confirming the IID character, since the p-values were 0.1417 and 0.8396 respectively. Also, the plots of autocorrelation function confirm it.



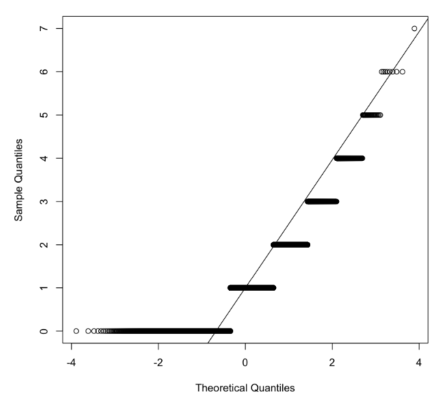
(a) Sample of Poisson Noise



(b) ACF of residuals of Poisson Noise

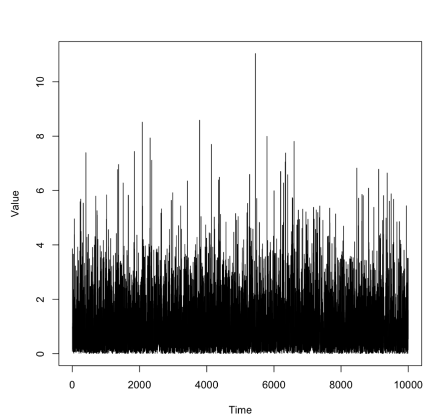


(c) Histogram of Poisson Noise

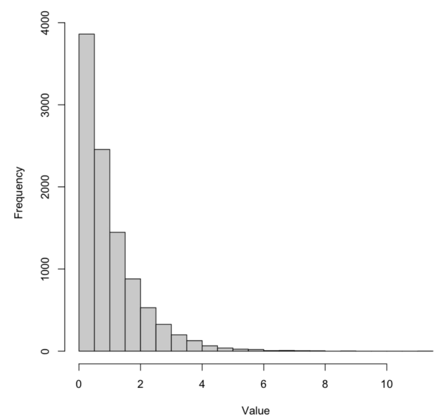


(d) Normal Q-Q Plot

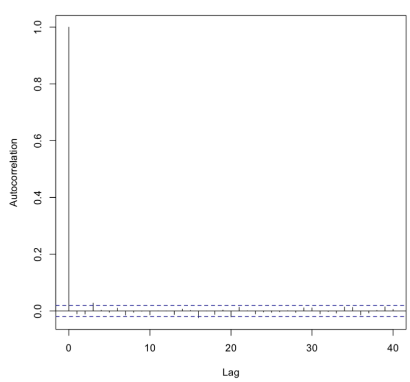
Figure 4: Plots for Poisson Noise



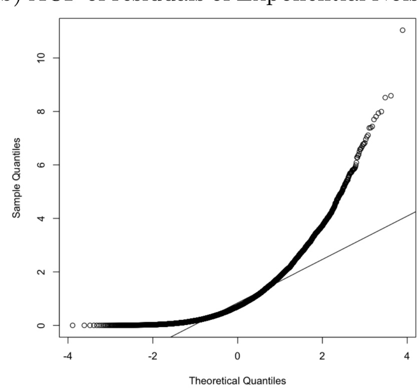
(a) Sample of Exponential Noise



(b) ACF of residuals of Exponential Noise



(c) Histogram of Exponential Noise



(d) Normal Q-Q Plot

Figure 5: Plots for Exponential Noise

Simulate an AR(p) model with 10000 data, for p=1 and p=2. Fit the best model to the data in both cases. Validate the model by showing the residuals are an IID noise.

In the next exercise, we simulate an AR(1) time series, plot the time series, compute the autocorrelation function, estimate AR parameters using maximum likelihood estimation, fit an AR(1) model, and plot a histogram of the simulated AR(1) process. After assessing AR(1), we run the same code for AR(2) model.

The AR parameter estimation using maximum likelihood estimation (MLE) for the AR(1) process resulted in the following:

- The estimated AR coefficient (lag 1) is 0.6923.
- The order selected is 1, indicating an AR(1) model.
- The estimated variance (σ^2) is approximately 0.9971.
- Log-likelihood is -14175.16.
- AIC is 28356.33.

We can see that the last two parameters, i.e. likelihood and AIC are in absolute terms very high. However it may be due to high n.

Residual Calculation: Residuals are obtained from the fitted AR(1) model (arima1_model).

Plotting ACF and PACF: Autocorrelation and partial autocorrelation of the AR(1) residuals are plotted to assess autocorrelation patterns.The value of the Ljung-Box test for residuals is 0.8639 together with ACF of AR(1) Residuals suggest that the noise is IID.

Ljung-Box Test: The Ljung-Box test is performed to formally test for the absence of autocorrelation in residuals.

- For the original series (x1):

  p-value $\leq 2.2e - 16$ (Highly significant)

  Interpretation: Strong evidence of autocorrelation in the original AR(1) series.

- For the AR(1) residuals (ar1_residuals):

  p-value = 0.8639 (Not significant)

  Interpretation: Lack of evidence of autocorrelation in the AR(1) residuals.

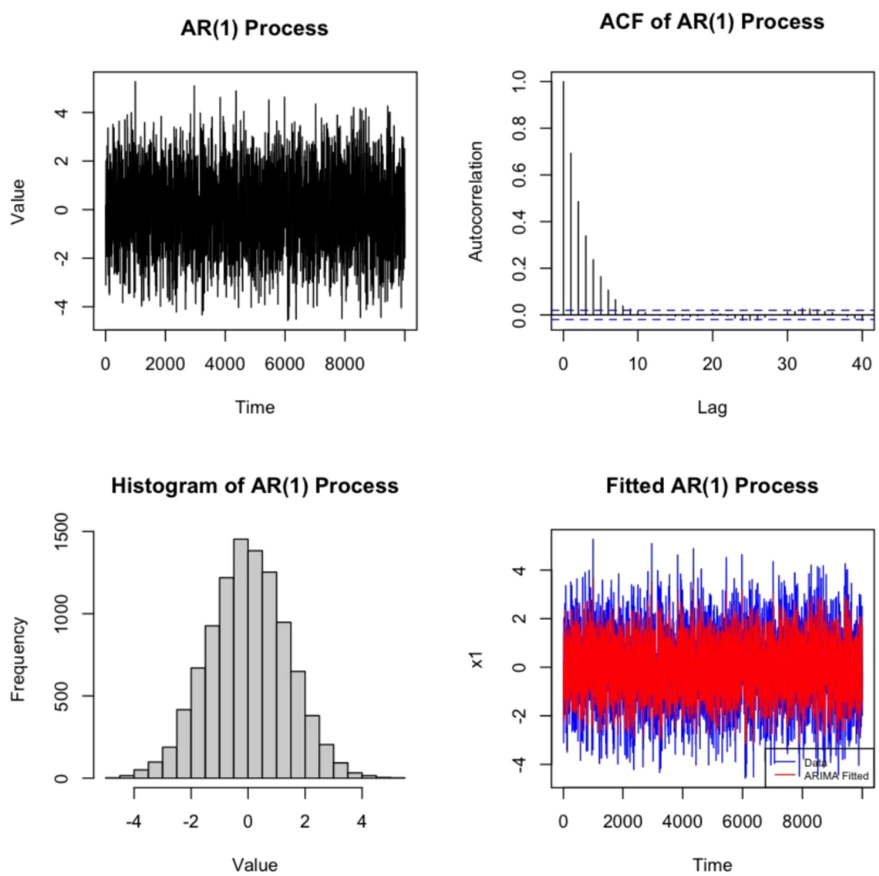

AR(1) Process
Value
Time
ACF of AR(1) Process
Autocorrelation
Lag
Histogram of AR(1) Process
Frequency
Value
Fitted AR(1) Process
x1
Time
Data
ARIMA Fitted

Figure 6

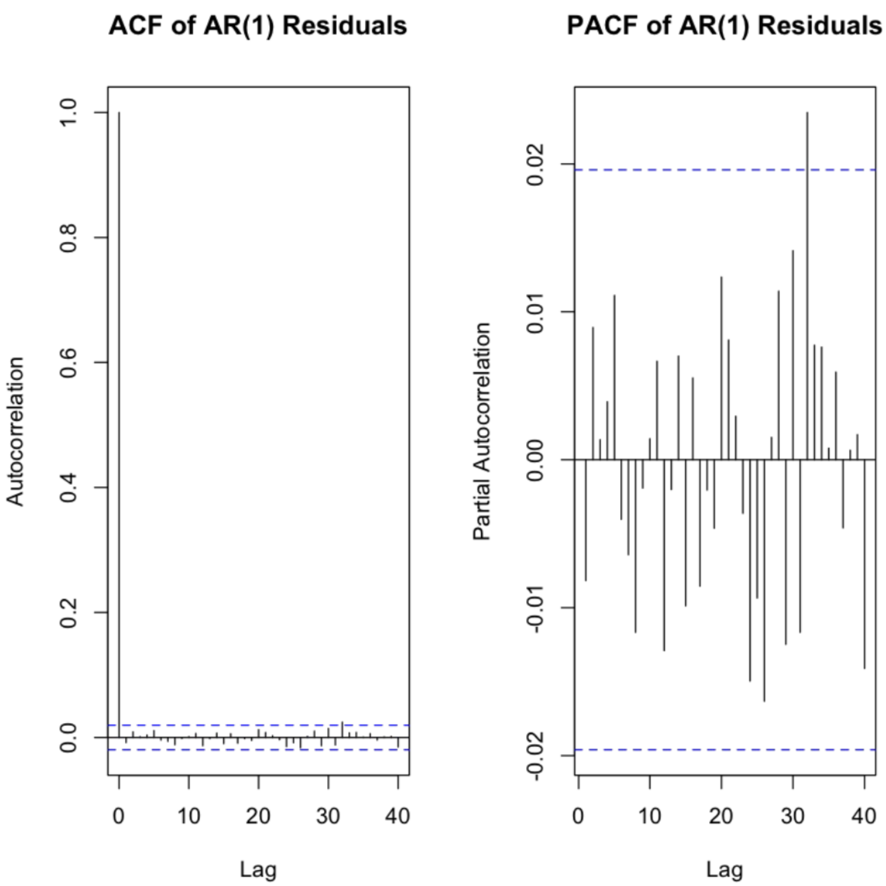

ACF of AR(1) Residuals
Autocorrelation
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
Lag
PACF of AR(1) Residuals
Partial Autocorrelation
-0.02
-0.01
0.00
0.01
0.02
0
10
20
30
40
Lag

Figure 7

These results suggest that while the original AR(1) series exhibits significant autocorrelation, the residuals from the AR(1) model appear to be adequately capturing the autocorrelation structure, as indicated by the non-significant p-value from the Ljung-Box test. Now we do the same for AR(2) process.

The AR parameter estimation using maximum likelihood estimation (MLE) for the AR(2) process resulted in the following:

- Estimated AR Coefficients (Lag 1 and Lag 2):

  Coefficient 1: 0.8970

  Coefficient 2: -0.4881

- These coefficients represent the strength and direction of the relationship between each observation and the two previous observations in the AR(2) process.
- Order Selected: 2 - The order selected indicates the number of lag terms included in the autoregressive model. In this case, an order of 2 suggests an AR(2) model, which considers the two previous observations.
- Log-likelihood is -14204.25.
- AIC is 28416.5.

The AIC and log-likelihood high values in absolute term are possibly due to the same result as with AR(1).

Residual Calculation: Residuals are obtained from the fitted AR(2) model (`arima2_model`).

Plotting ACF and PACF: Autocorrelation and partial autocorrelation of the AR(2) residuals are plotted to assess autocorrelation patterns. The value of the Ljung-Box test for residuals is 0.8996 together with ACF of AR(1) Residuals suggest that the noise is IID.

Ljung-Box Test: The Ljung-Box test is performed to formally test for the absence of autocorrelation in residuals.

- For the original series (`x2`):

  p-value $\leq 2.2e-16$ (Highly significant)

  Interpretation: Strong evidence of autocorrelation in the original AR(2) series.

- For the AR(2) residuals (`ar2_residuals`):

  p-value = 0.8996 (Not significant)

  Interpretation: Lack of evidence of autocorrelation in the AR(2) residuals.

These results suggest that while the original AR(2) series exhibits significant autocorrelation, the residuals from the AR(2) model appear to be adequately capturing the autocorrelation structure, as indicated by the non-significant p-value from the Ljung-Box test.

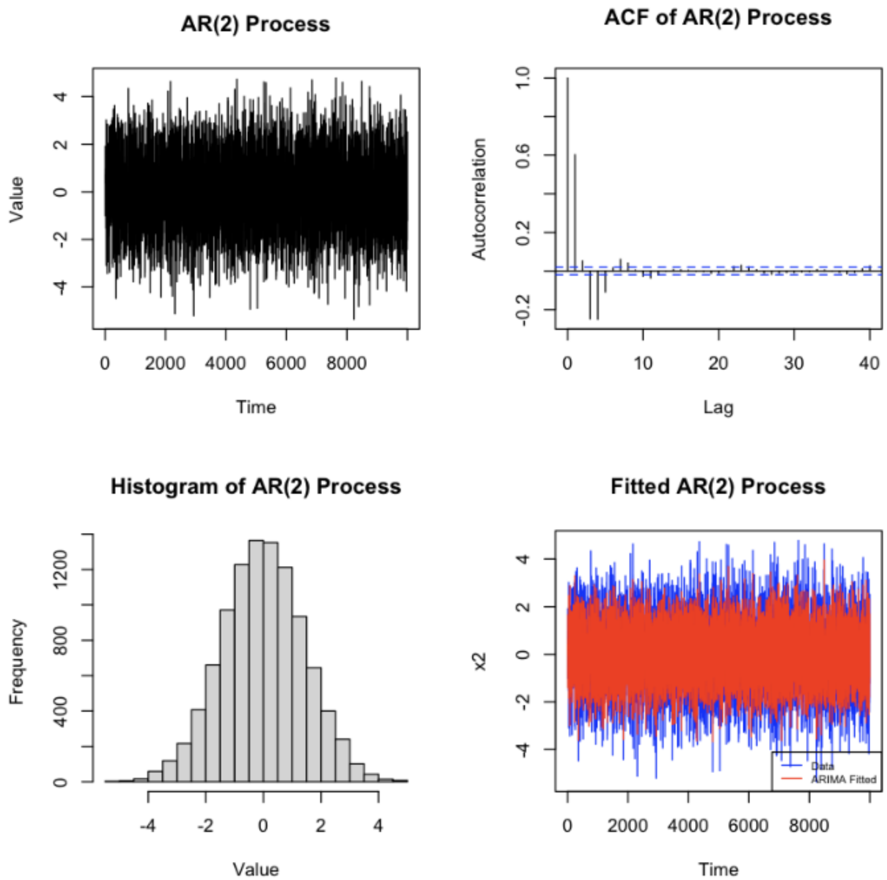


Figure 8

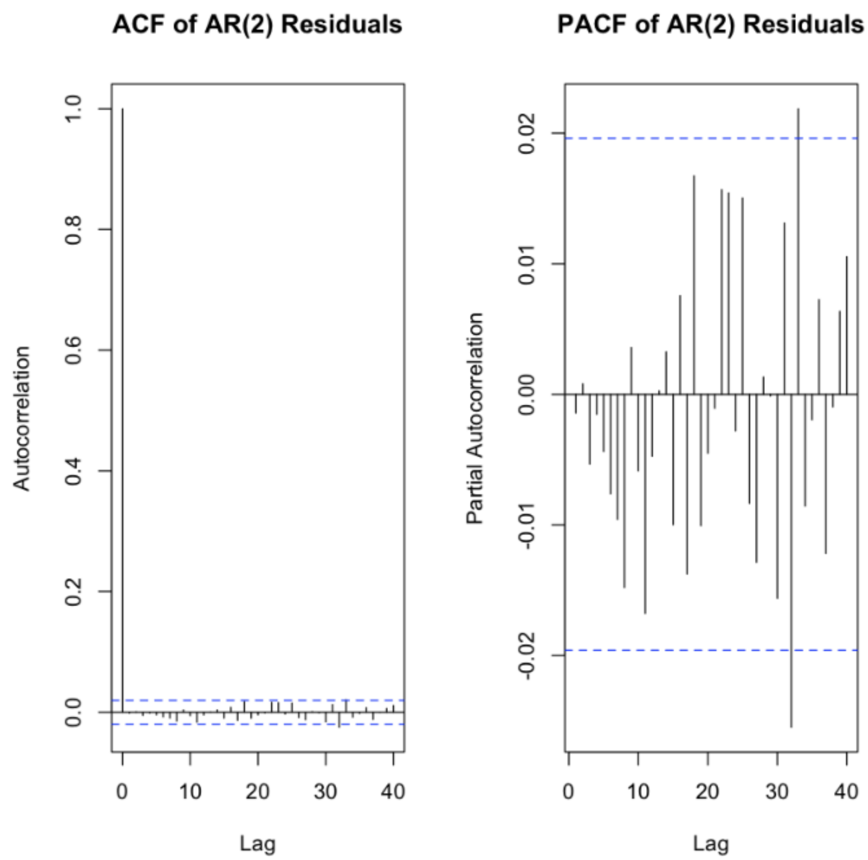


Figure 9

Simulate an ARMA (2,1). Compute the autocorrelation and the partial autocorrelation. Fit the best ARMA model. Validate it. Make the graphical representation of the forecasting.

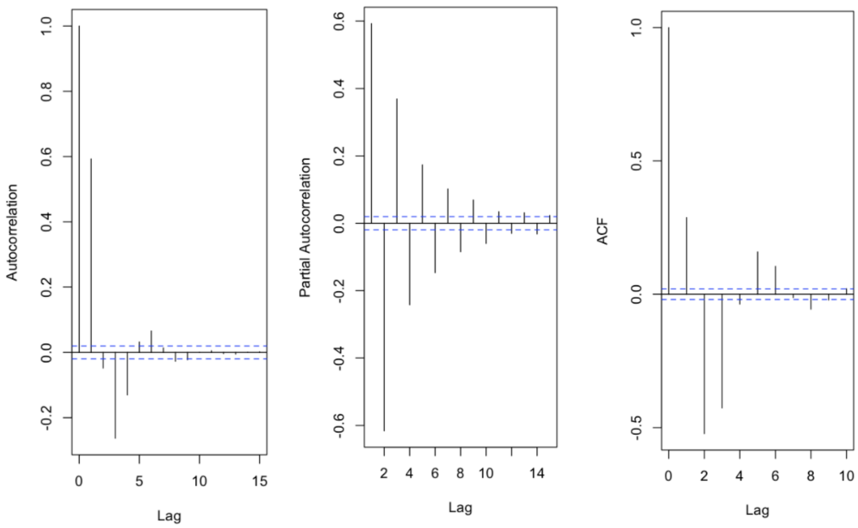
Simulating ARMA(2,1) Process:

- **Simulation:** An ARMA(2,1) process is simulated with specified coefficients using the `arima.sim` function.

Model Fitting:

- **Fitting the ARMA Model:** The best ARMA(2,1) model is fitted to the simulated process using the `arima` function.
  - The ARMA model is estimated with order $(p, d, q) = (2, 0, 1)$.

Model Validation:



(a) ACF (b) PACF (c) ACF diff values

Figure 10: ARMA(2,1) Process

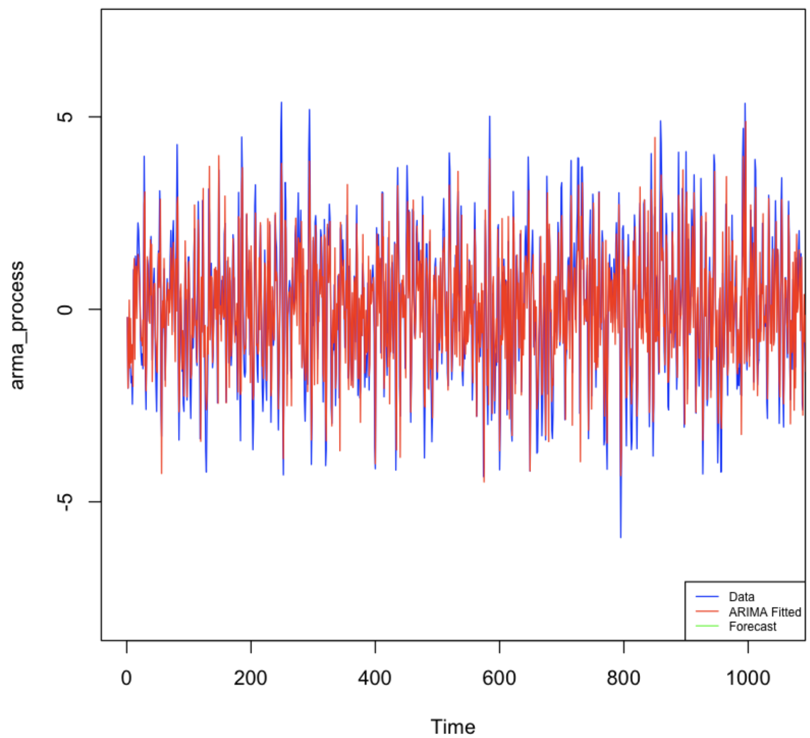
- **Residual Analysis:** Residuals of the fitted ARMA model are computed and analyzed.
  - Autocorrelation and partial autocorrelation plots of the residuals are created.
  - The Ljung-Box test is conducted to test for the absence of autocorrelation in residuals.
  - Results suggest that the residuals exhibit no significant autocorrelation.

Forecasting:

- **Forecasting:** Forecasting of the ARMA(2,1) process is performed using the `forecast` function.

Model Summary:

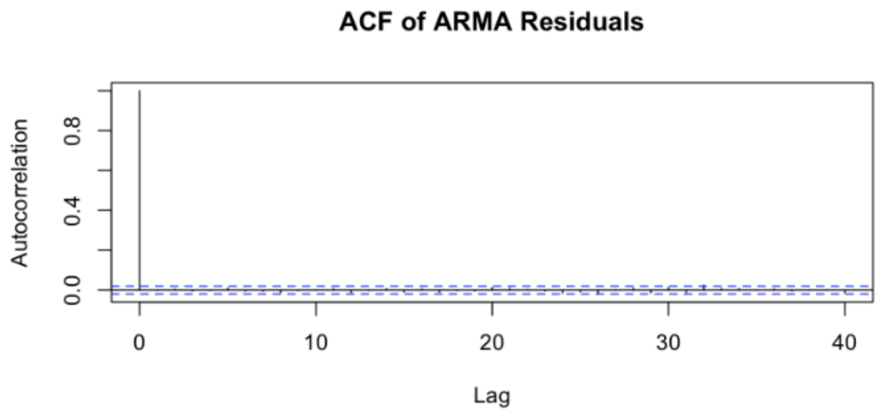
- **ARMA Model Coefficients:**
  - AR1: 0.5877
  - AR2: -0.3959
  - MA1: 0.7977
  - Standard errors (s.e.) are provided for each coefficient.
- **Model Information:**

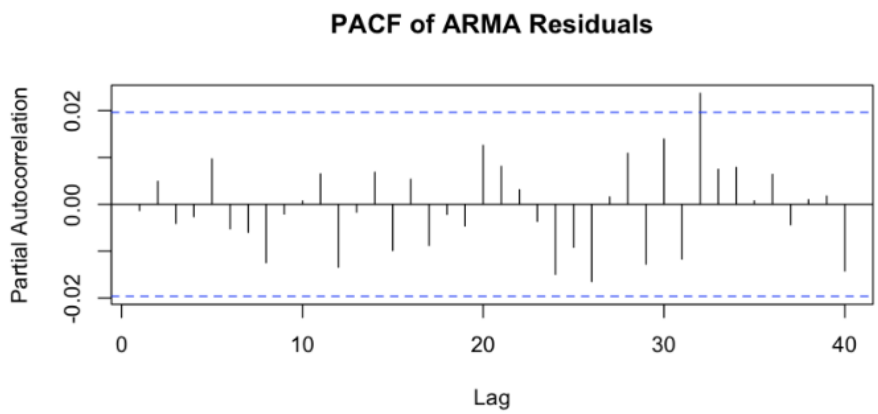


  - Sigma^2 (variance): 0.9977
  - Log likelihood: -14177.67

- **Box-Ljung Test Results:**
  - For the original ARMA(2,1) process: Highly significant evidence of autocorrelation.
  - For the ARMA residuals: No significant evidence of autocorrelation.

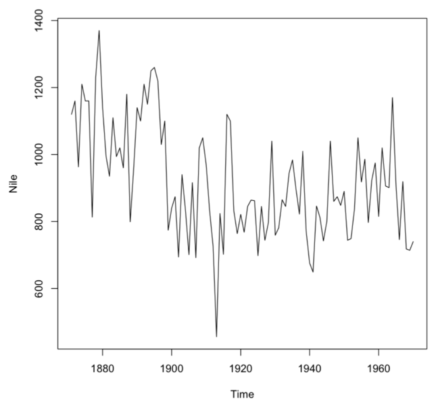
These results show how accurate is the model of the ARMA(2,1) model to the simulated process and its forecasting performance. Even when using `auto.arima` function the suggested model is $(2, 0, 1)$. Lastly, it is important to note, that we used CSS method for the model to converge.



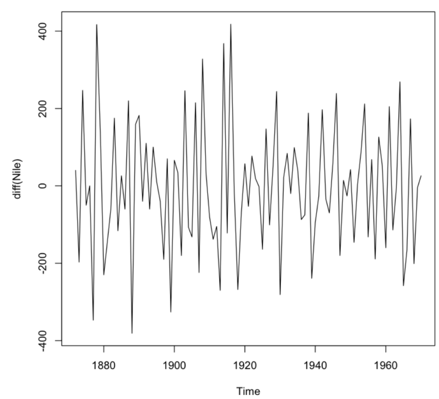


Take the file Nile in datasets. Fit the best ARIMA model to the process. Validate it. Make the graphical representation of the forecasting.

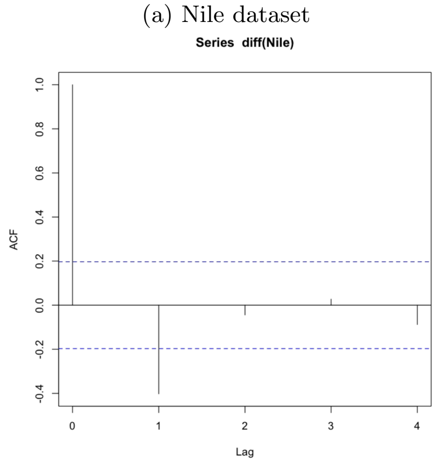
By plotting autocorrelation and partial autocorrelation functions, we observe that the most appropriate values for p and q are 1,1. After running auto.arima function, model ARIMA(1,0,1) is suggested. Validating the model and looking at residuals do not show any problems. The p-value 0.9204 indicates IID noise.



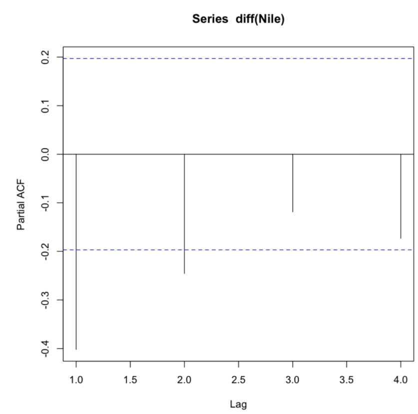
(a) Nile dataset



(b) Diff of Nile dataset



(c) ACF Diff of Nile



(d) PACF Diff of Nile

Figure 11: Plots for Nile dataset

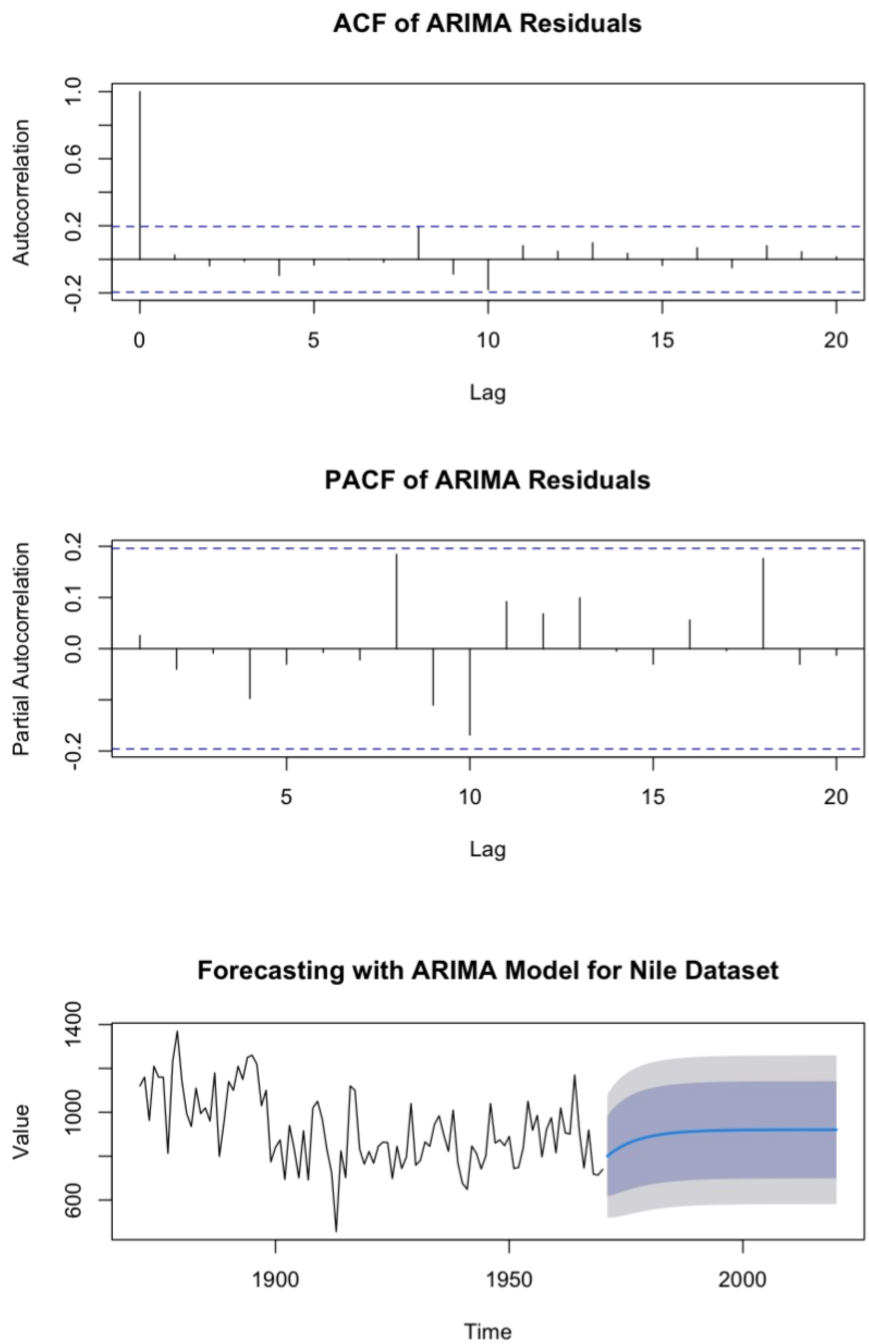

ACF of ARIMA Residuals
Autocorrelation
-0.2
0.2
0.6
1.0
0
5
10
15
20
Lag
PACF of ARIMA Residuals
Partial Autocorrelation
-0.2
0.0
0.1
0.2
5
10
15
20
Lag
Forecasting with ARIMA Model for Nile Dataset
Value
600
1000
1400
1900
1950
2000
Time

Figure 12

Simulate a FARIMA time series. Fit it the best model and test that the residuals of the fitted model are a white noise. Fit a FARIMA model to Nile data in datasets. Check that the fitted model is a good model.

Simulation and Fitting:

- We try to simulate and fit a fractionally differenced autoregressive moving average (FARIMA) time series, using the `fracdiff` package.
- Using the `fracdiff.sim` function with parameters `ar=0.5` and `d=0.4`. This corresponds to a $(1, d, 0)$ FARIMA model. Furthermore, we plot the simulated series. It fits a FARIMA model to the simulated series using the `fracdiff` function, specifying `nar=1` for an AR(1) model. Notice the poor estimation of d.

Residual Analysis:

- The estimated differencing parameter d is extracted from the summary of the fitted model. The differenced series y_1 is created by taking the d differenced series from the original simulated series. The autocorrelation function (ACF) of y_1 is plotted. An ARIMA$(1, 0, 0)$ model is fitted to y_1. The residuals of the ARIMA model y_2 are extracted and stored in z. The ACF of the residuals z is plotted. Finally, the Ljung-Box test is performed on the residuals z to check for white noise behavior which is confirmed by p-value > 0.05, i.e., 0.9902.

Autocorrelation Function (ACF) Analysis:

The ACF plot of the Nile dataset reveals a significant autocorrelation at lag 1, indicating a potential AR(1) process. FARIMA Model Fitting: The FARIMA model is fitted to the differenced data of the Nile dataset, with the differencing parameter d estimated to be approximately 4.58e-05. The FARIMA model summary shows a significant estimation of d (p $\leq$ 0.001), suggesting fractional differencing is necessary to achieve stationarity.

Residual Analysis:

The residuals obtained from the fitted FARIMA model are further analyzed. The ACF plot of the residuals reveals the presence of autocorrelation at various lags, indicating potential model misspecification or residual temporal dependencies. The Ljung-Box test on the residuals yields a p-value of 0.002541, indicating significant autocorrelation in the residuals. This suggests that the fitted FARIMA model might not adequately capture all temporal dependencies.

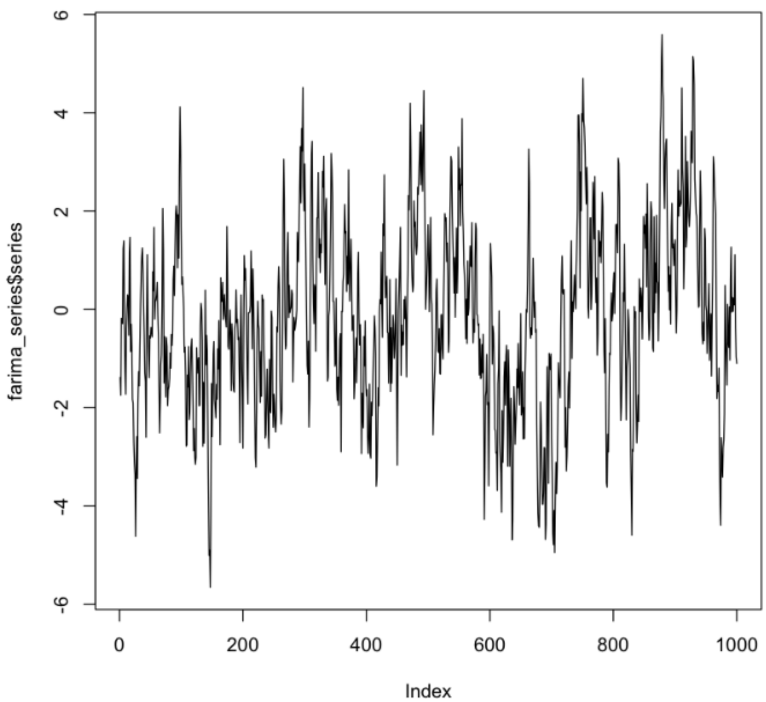


Figure 13: Farima

Simulate a GARCH(1,1) time series. Fit the best model to this series. Check that the fitted model it is a good model. Fit a GARCH model to the logarithmic transformation of series in EuStockMarkets of datasets. Check the stylized facts (un-correlation, correlation of the squares, heavy tails, volatility clustering). Check that the fitted model is a good model.

Model Simulation:

- A GARCH(1,1) time series is simulated with parameters ($a_0 = 0.1$), ($a_1 = 0.4$), and ($b_1 = 0.2$), along with white noise innovations.

Model Fitting:

- A GARCH(1,1) model is fitted to the simulated data using the `garch` function.
- The model summary displays the estimated coefficients for (a_0), (a_1), and (b_1), along with their standard errors and significance levels. The parameters are very closed to the original ones.
- The estimated coefficients are found to be statistically significant at a high

significance level ($p < 0.001$), indicating a strong effect of both the lagged squared error term and the lagged conditional variance on the current conditional variance.

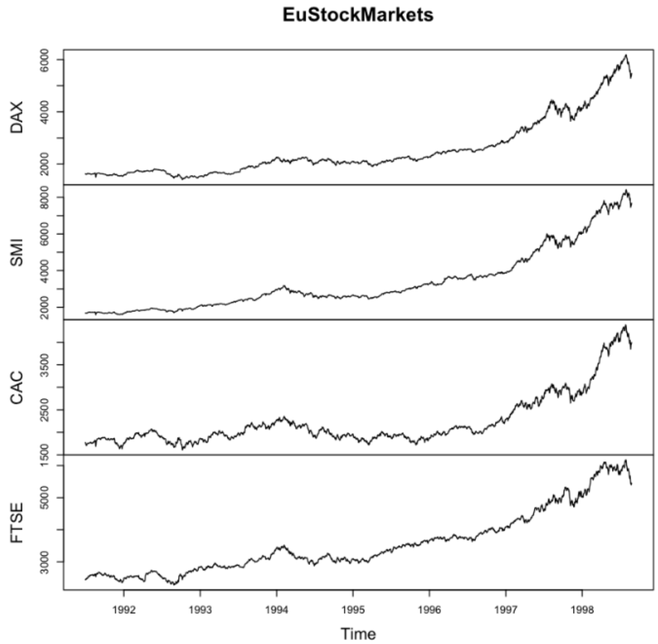
Residual Analysis:

- The residuals from the GARCH(1,1) model are extracted and plotted to visualize their behavior.
- The autocorrelation function (ACF) of the residuals and the ACF of the squared residuals are plotted to assess any remaining temporal dependencies or heteroscedasticity.
- Diagnostic tests, including the Jarque-Bera test and the Box-Ljung test, are conducted on the residuals and squared residuals, respectively, to evaluate the model's goodness-of-fit.
- The Jarque-Bera test statistic's p-value (0.9236) and the Box-Ljung test statistic's p-value (0.8364) suggest no evidence of non-normality in the residuals or significant autocorrelation in the squared residuals, indicating a satisfactory fit of the GARCH(1,1) model to the simulated data. Thus the above results confirm IID noise. Lastly, there are no heavy tails, however, high correlation is present.

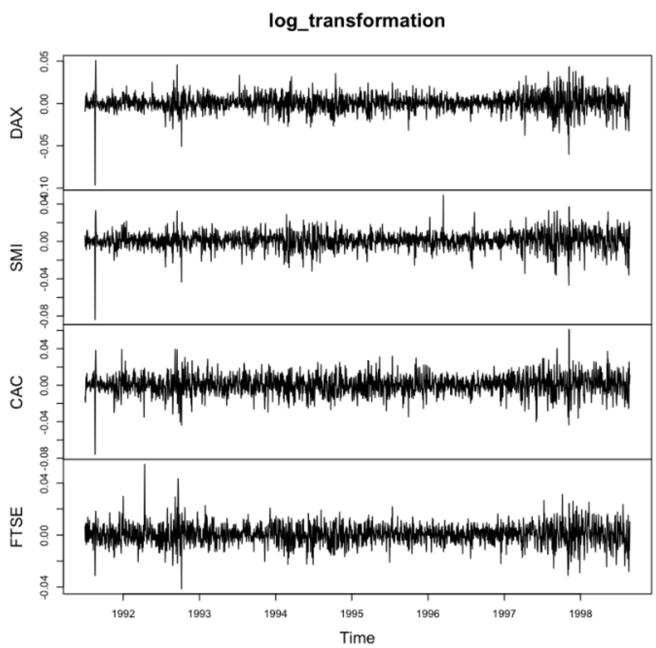
Parameter Confidence Intervals:

- The 95% confidence intervals for the estimated parameters (a_0), (a_1), and (b_1) are provided, indicating the range of good values for the parameters.

For the last part, we fitted GARCH(1,1) to EuStockMarkets dataset, where we worked with logarithimic differnced series and checked the stylized facts.

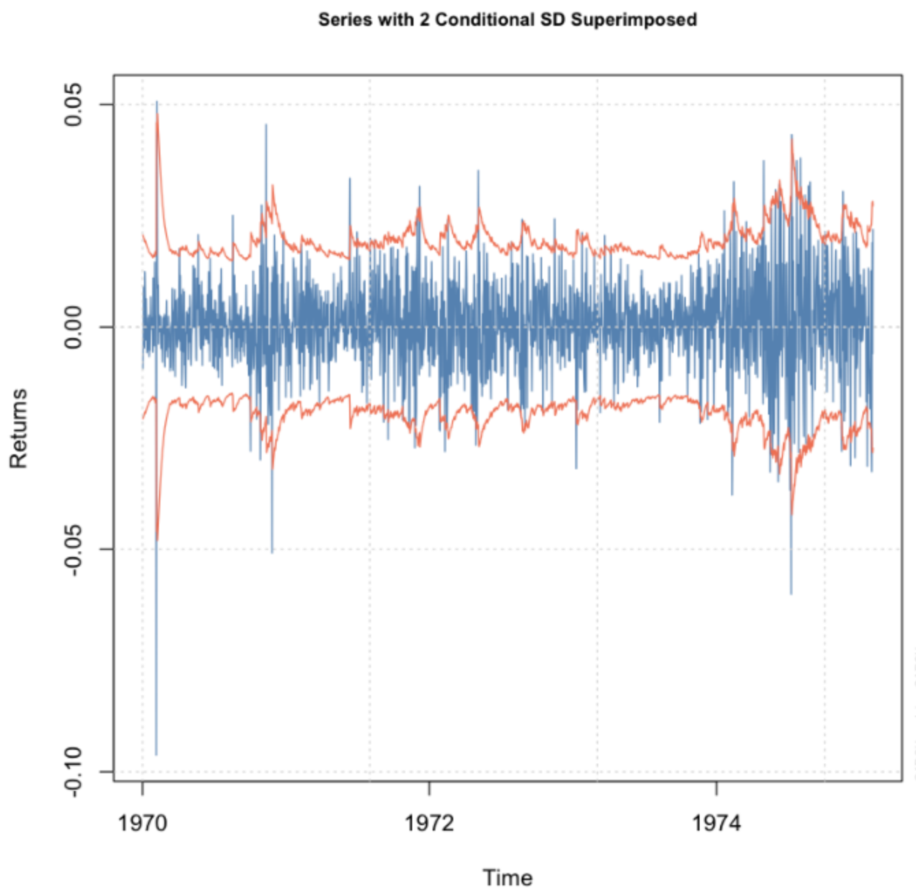

EuStockMarkets
DAX
SMI
CAC
FTSE
Time

(a)


log_transformation
DAX
SMI
CAC
FTSE
Time

(b)

Figure 14



Series with 2 Conditional SD Superimposed
0.05
0.00
-0.05
-0.10
Returns
1970
1972
1974
Time
GARCH model : sGARCH

Figure 15