Uppsala University
Department of Statistics

Homework assignment 2

Time series Analysis

Mayara Latrech & Roua Ben Ammar

Contents

1 Introduction

After getting more familiar with how the time series plot, ACT and PACF look like for MA(1), MA(2), AR(1), AR(2), and ARMA(1,1) models with the homework assignment 1, the this assignment we try to figure out which model corresponds best to our data. In the first task, we have five simulations of five stochastic processes, we need to identify for each process what model can capture best its variation then we use it to forecast future values. In the second task, we choose real data and try to find the best model for it using the same Box-Jenkins method.

2 Task 1

2.1 Process Y1

Identification

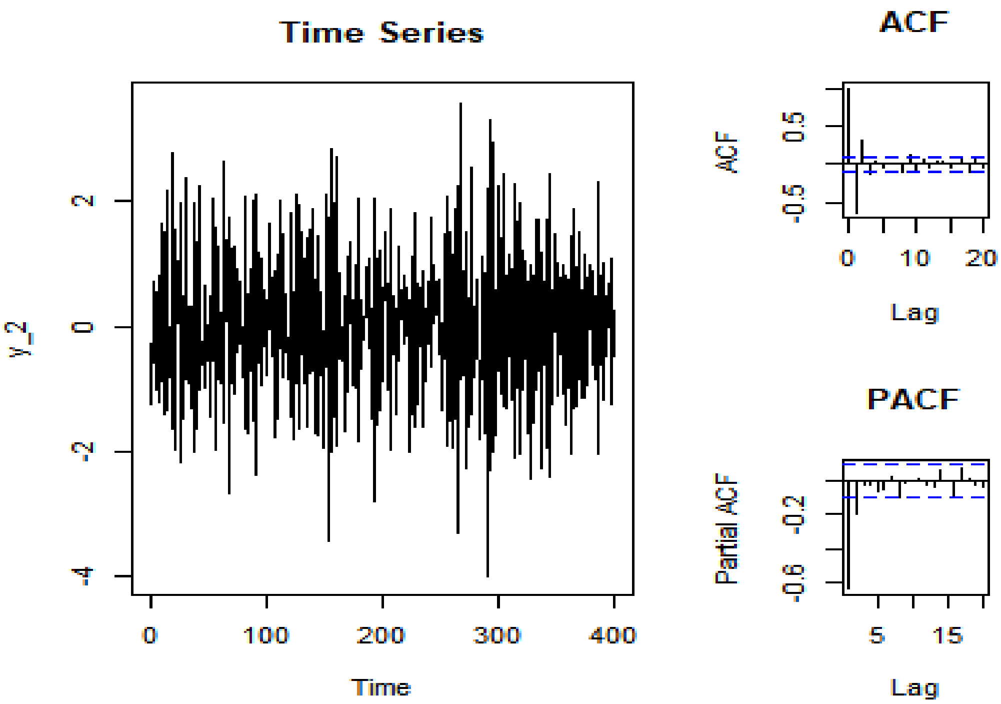


Figure 1: Time series plot , ACF and PACF of process Y1

The time series is choppy, it oscillates around the mean. The overall mean is stable as no visible drift nor trend can be observed. The variance is not stable throughout the process.

The ACF has alternating upward and downward spikes and it dies off quickly.
The PACF also has alternating upward and downward spikes, it does not cut off until the end but the spikes become less and less pronounced compared to the first 2 spikes.
Based on the time series plot and the correlogram, our best guess is that the process is an AR (1) process with a parameter ∅ that is negative and smaller than 1 inabsolute value.
The instability in the variance suggests that the parameter's value might be close to 1 in absolute value.

Estimation

We fit an AR(1) model with intercept to the data.
The fitted model is the following:

$$Y_t = 0.01 + 0.0188Y_{t-1} + e_t \tag{1}$$

Evaluation

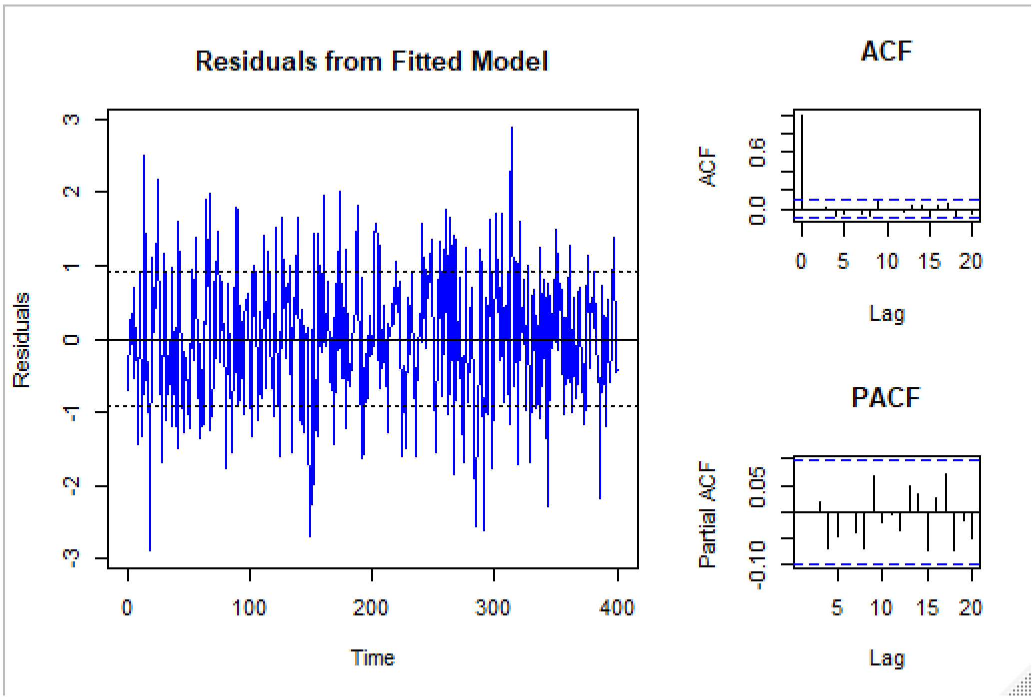


Figure 2: The plot, ACF and PACF of the residuals of the process Y1

The plot shows that the residuals are choppy with a constant mean of zero and no visible drift or trend.
The variance of the residuals seem to surpass the error bounds of the AR(1) process.
According to the correlogram, the ACF only has one significant spike at lag zero , which represents the variance, after that it completely dies off indicating that the residuals are uncorrelated. .)
The PACF does not represent any significant correlations as non of the spikes seem to exceed the cutoff values for the Z test for individual autocorrelation.
To confirm that the residuals are in fact uncorrelated, we conduct the Box-Ljung test of autocorrelation:

- Hypothesis:

$$H_0: \forall i \in \{1{,}20\}, \rho_i = 0$$
$$H_1: \textit{At least one} \;\; \rho_i \neq 0, i \in \{1{,}20\}$$

- Significance level: $\alpha = 0.05$
- Estimators: $\widehat{\rho_i}$, $i \in \{1{,}20\}$
- Assumption: “Large” T: number of observations

❖ Test statistic: $Q_{LB} = T(T+2)\sum_{j=1}^{20}\frac{\widehat{\rho_j}^2}{T-j} \sim \mathcal{X}^2_{0.05,18}$ under H0.

❖ Decision rule and figure:

Reject if ${Q_{LB}}^{Obs} > \mathcal{X}^2_{0.05,18}$ or P-value<0.05.

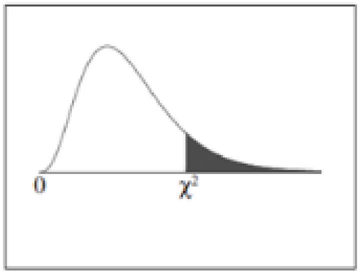


Figure 3: Chi2 distribution

❖ Calculations (R output): P-value = 0.9998 > 0.05. We do not reject H_0.

❖ Conclusion: At the 5% significance level, we fail to reject that the residuals are uncorrelated.

And thus we can conclude that the model does in fact capture the systematic variations in the data leaving behind uncorrelated residuals.
Next , we check the normality of the residuals by plotting them against the theoritical quantiles of the normal distribution:

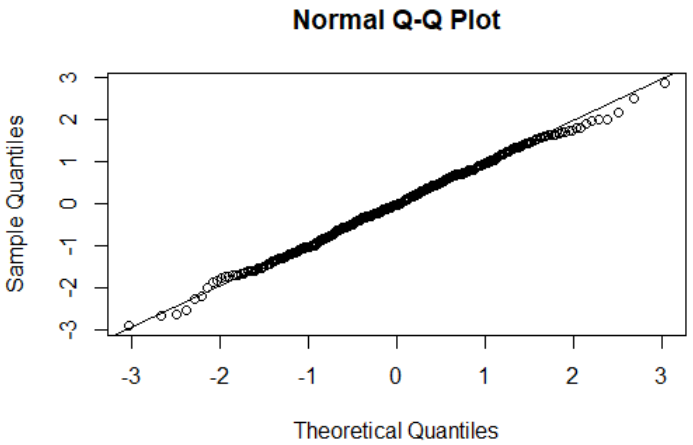


Figure 4: Q-Q plot of the residuals of Y1 process

The residuals quantiles coincide with the theoritical quantiles, thus we suspect that they are in fact normally distributed.

In order to confirm their normality, we conduct the Shapiro-Wilk normality test:

- Hypothesis:

$$H_0\text{: }the\ residuals\ are\ normally\ distributed$$

$$H_1\text{: }the\ residuals\ are\ not\ normally\ distributed$$

- Significance level: $\alpha = 0.05$
- Estimators: $\widehat{e_t}$
- Test statistic = W
- Assumption: "Large" T: number of observations
- Decision rule: Reject if or P-value<0.05.
- Calculations (R output): P-value = 0.7792 > 0.05. We do not reject H_0.
- Conclusion: At the 5% significance level, we fail to reject that the residuals are normally distributed

Forecasting

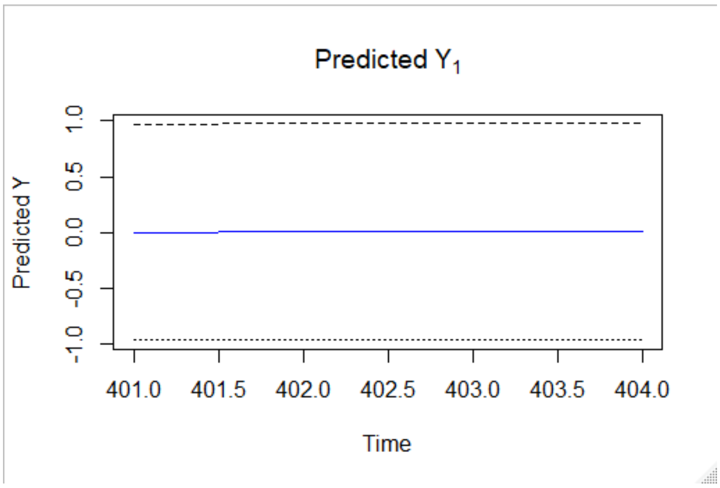


Figure 5: Predictions of model 1

The figure above, shows that the variance of the forecasted values using th e model Y1 is significantly large, but stable. This indicates that using this model to forecast future values comes with large uncertainty.
Next, we compare the fitted values to the actual values:

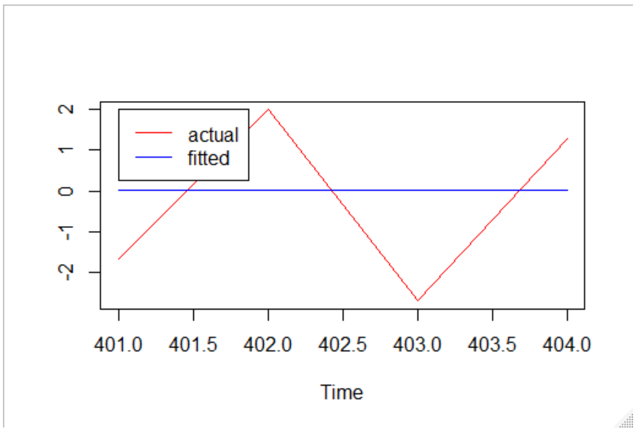


Figure 6: Actual vs fitted values of model 1

The actual values are quite different from the fitted values, in fact, the actual values fall beyong the forecasting bounds of the model.
This indicates that even though the systematic variations are in fact captured by Y5, the model fails to provide good enough forecasts for future values.

2.2 Process Y2

Identification

In order to identify the process, we first plot the time series, its ACF and PACF.

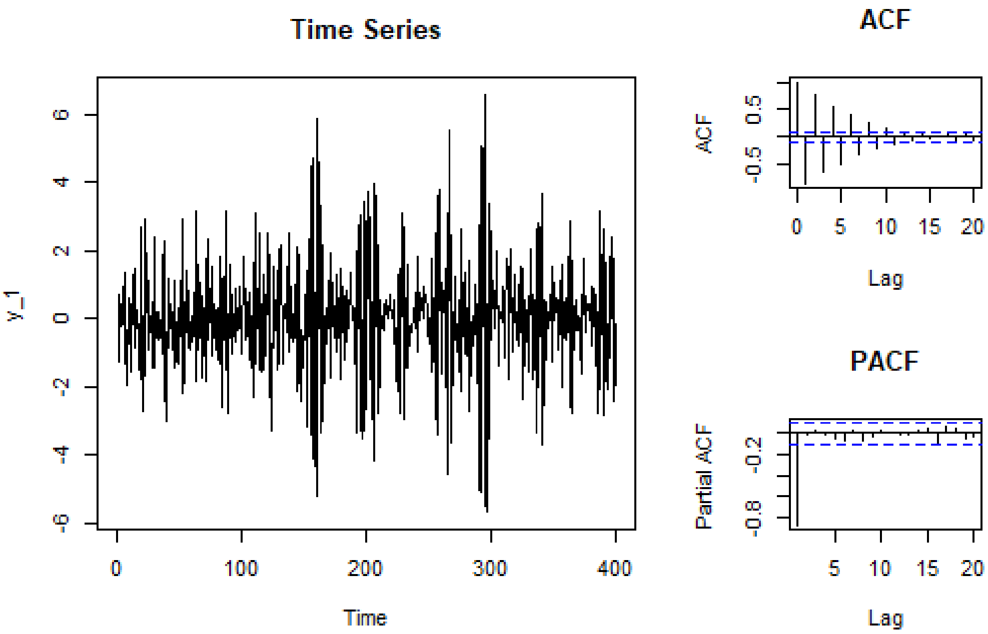


Figure 7: Time series plot, ACF and PACF of process Y2

After plotting the time series of the process Y2, we notice that it is choppy, with no visible drift or trend. The variance does not look stable throughout the process. The plot of the ACF starts with a positive spike then it keeps oscillating between positive and negative spikes. We notice that the first 11 correlations are actually significant so we fail to reject that they are 0. The other correlations are under the Z-test significance level so we can practically consider them 0. The PACF starts with a negative spike then it dies off completely. These plots look like they belong to an AR(2) process with a negative ϕ_1 and a positive ϕ_2.

Estimation

In order to estimate the model AR(2), we used the default procedure in R and we got:

$$\widehat{Y_t} = 0.0064 - 0.8769\ Y_{t-1} - 0.0079\ Y_{t-2} + e_t \tag{2}$$

Evaluation

In order to evaluate the process, we first plot the time series, its ACF and PACF of the residuals.

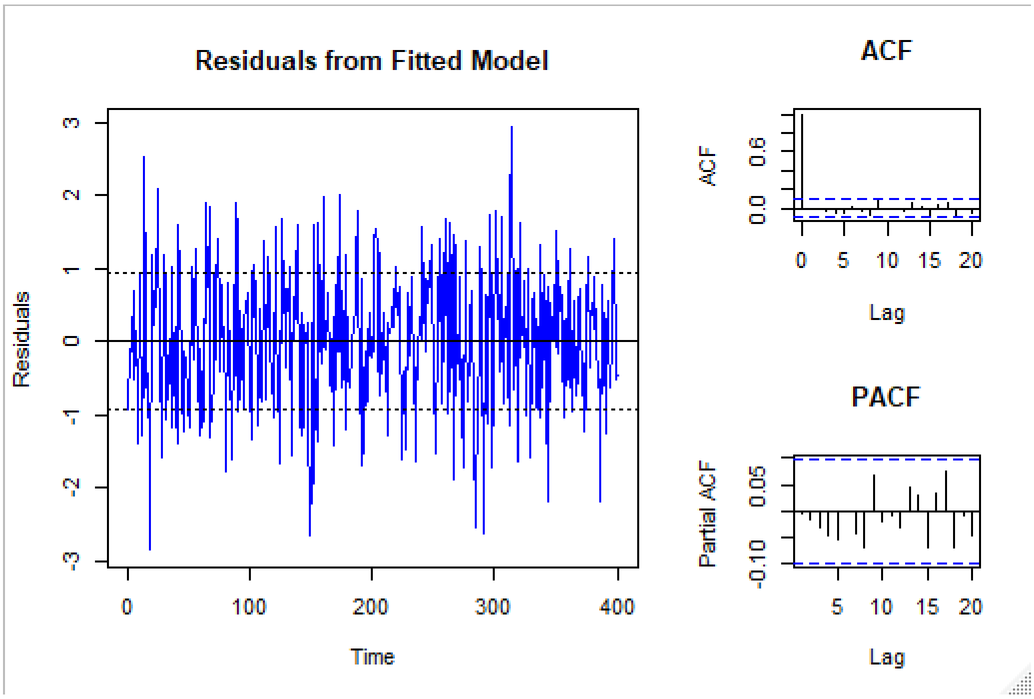


Figure 8: Time series plot, ACF and PACF of the residuals of process Y2

The plot of the residuals of the fitted model looks choppy. It has a constant mean=0 and the variance of the residuals seems to surpass the error bounds. However, the variance still seems stable. There is no visible drift or trend. The first spike of the ACF corresponds to the variance of the residuals which we already know is equal to 1. The ACF and PACF do not have any significant spikes which makes us suspect that the residuals are in fact uncorrelated. We test the correlation using an Ljung-Box Q test.

- Hypothesis:

$$H_0\text{: } \forall i \in \{1{,}20\}, \rho_i = 0$$
$$H_1\text{: } \textit{At least one } \ \rho_i \neq 0, i \in \{1{,}20\}$$

- Significance level: $\alpha = 0.05$
- Estimators: $\widehat{\rho_i}$, $i \in \{1{,}20\}$
- Assumption: "Large" T: number of observations
- Test statistic: $Q_{LB} = T(T+2)\sum_{j=1}^{20} \frac{\widehat{\rho_j}^2}{T-j} \sim \mathcal{X}^2_{0.05,18}$ under H0.
- Decision rule and figure:

  Reject if ${Q_{LB}}^{Obs} > \mathcal{X}^2_{0.05,18}$ or P-value<0.05.

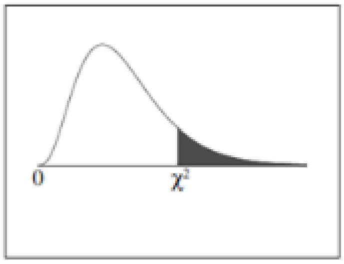


Figure 9: Chi2 distribution

- Calculations (R output): P-value= 0.9885 > 0.05. We do not reject H0.
- Conclusion: At the 5% significance level, we fail to reject that the residuals are uncorrelated.

Now we test if the residuals are normally distributed. First, we evaluate the normality visually through a Q-Q plot.

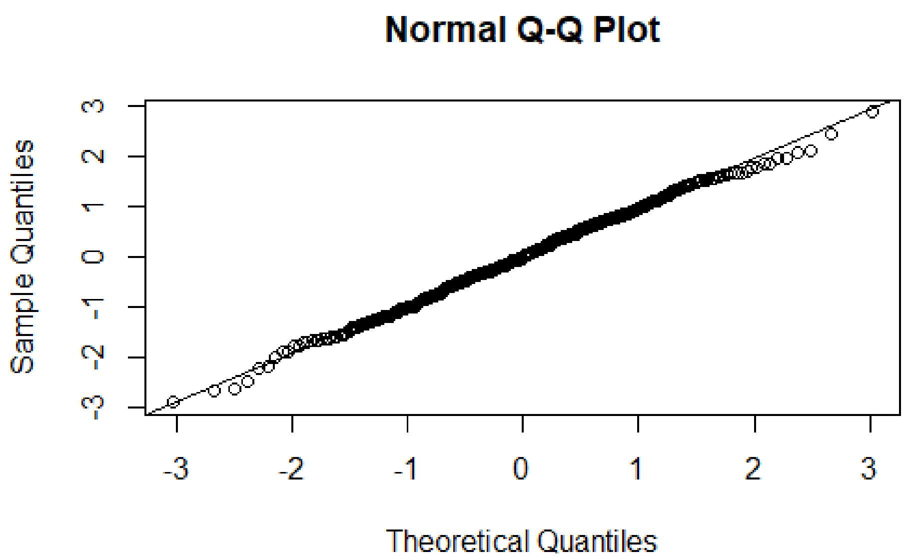


Figure 10: Q-Q plot of the residuals of Y2 process

We notice that the sample quantiles are very close the theoretical quantiles of the normal distribution. Let us then perform a normality test: The Shapiro-Wilk test:

- Hypothesis:

$$H_0\text{: } \textit{the residuals are normally distributed}$$
$$H_1\text{: } \textit{the residuals are not normally distributed}$$

- Significance level: $\alpha = 0.05$
- Estimators: $\widehat{e_\imath}$
- Assumption: “Large” T: number of observations
- Test statistic: W
- Decision rule: Reject if P-value<0.05.
- Calculations (R output): P-value= $0.733 > 0.05$. We do not reject H0.
- Conclusion: At the 5% significance level, we fail to reject that the residuals are not normally distributed.

Since the residuals are uncorrelated, we conclude that the model can capture the variation of the dataset.

Forecasting

After finishing up with the model diagnosis, we use it to forecast future values of the process.

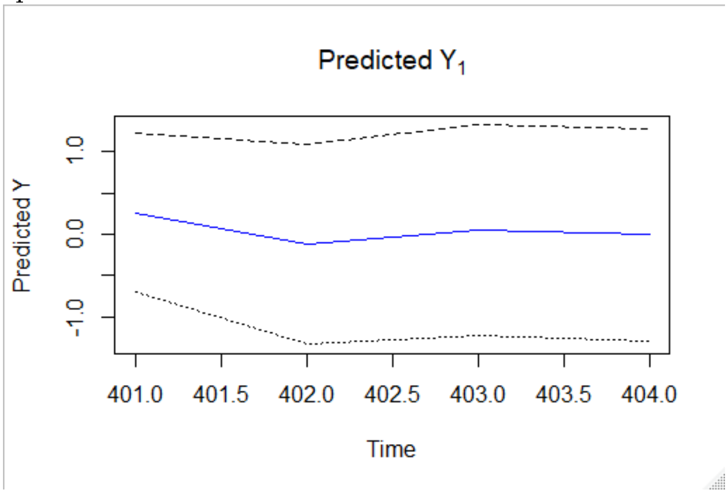


Figure 11: Predictions of model 2

The model can give us prediction values that we plotted with forecasting interval. We predict that the values we get are inside the forecasting interval. We also get point predictions that are indeed inside the forecasting interval. Let us now see if the predicted values are close to the real values.

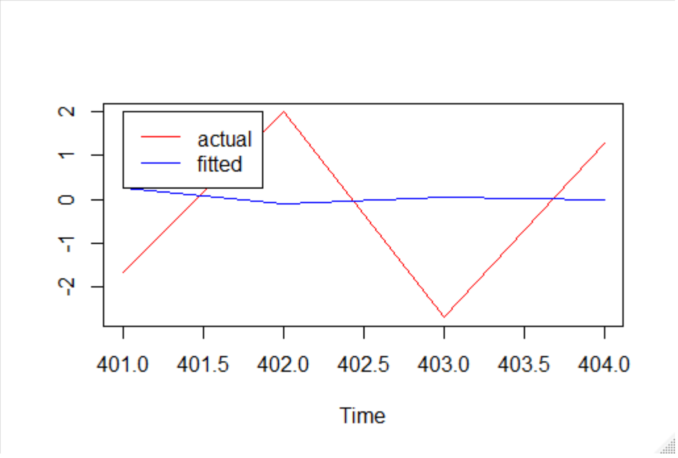


Figure 12: Actual vs fitted values of model 2

The forecasted values are not equal to the actual values of course due to having residuals. The actual values are not even inside the forecasting bounds which makes our model not accurate at forecasting.

2.3 Process Y3

Identification

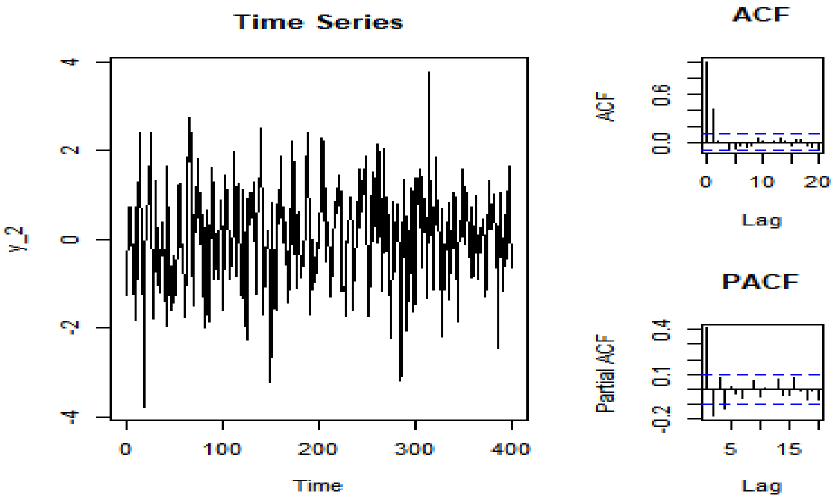


Figure 13: Time series plot, ACF and PACF of process Y3

The time series is choppy, it oscillates around the mean. The overall mean is stable as no visible drift nor trend can be observed. The variance is not stable throughout the process.
The ACF has 2 upward spikes then it dies off gradually rather quickly.
The PACF has alternating upward and downward spikes that do not die off completely but become less and less pronounced after the few first lags.
Based on the time series plot and the correlogram, the process seems to be an MA (2) with a positive parameter θ that is less than 1 in absolute value but close enough to result in the slight instability in the variance.

Estimation

The estimated model is an MA (2) with an intercept:

$$Y_t = 0.0136 + e_t + 0.5169e_{t-1} - 0.0096e_{t-2} \tag{3}$$

Evaluation

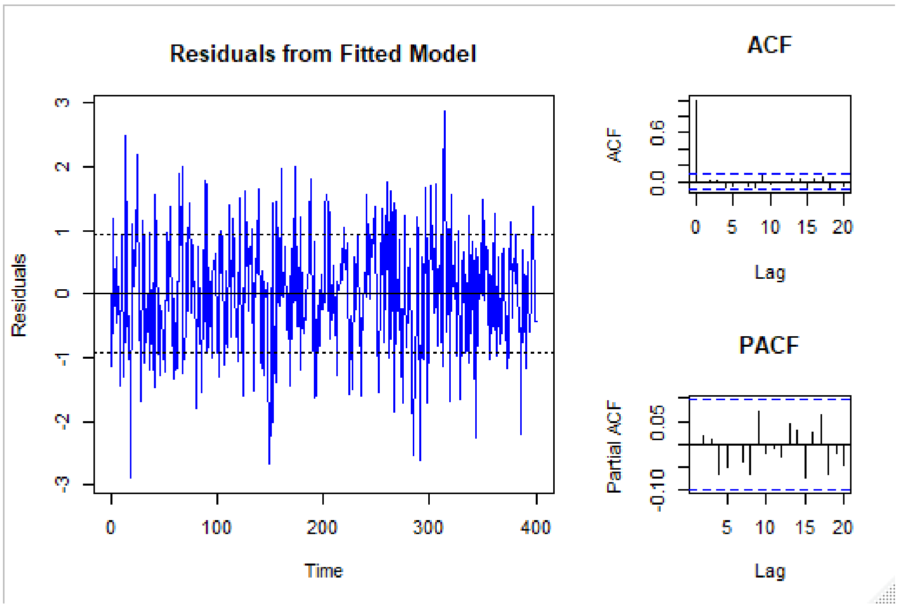


Figure 14: The plot, ACF and PACF of the residuals of Y3

The plot shows that the residuals are choppy with a constant mean of zero and no visible drift or trend.
variance of the residuals seem to surpass the error bounds of the ARMA(1,1) process.
According to the correlogram, the ACF only has one significant spike at lag zero , which reperesent the variance, after that it completely dies off indicating that the residuals are uncorrelated. . The PACF does not represent any significant correlations as non of the spikes seem to exceed the cutoff values for the Z test for individual autocorrelation.
To confirm that the residuals are in fact uncorrelated, we conduct the Box-Ljung test of autocorrelation:

- Hypothesis:

$$H_0: \forall i \in \{1{,}20\}, \rho_i = 0$$
$$H_1: \textit{At least one } \ \rho_i \neq 0, i \in \{1{,}20\}$$

- Significance level: $\alpha = 0.05$
- Estimators: $\widehat{\rho_i}$, $i \in \{1{,}20\}$

- Assumption: "Large" T: number of observations
- Test statistic: $Q_{LB} = T(T+2)\sum_{j=1}^{20}\frac{\widehat{\rho_j}^2}{T-j} \sim \mathcal{X}^2_{0.05,18}$ under H0.
- Decision rule and figure:

  Reject if ${Q_{LB}}^{Obs} > \mathcal{X}^2_{0.05,18}$ or P-value<0.05.

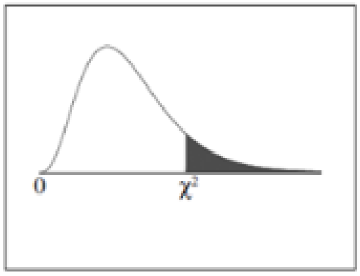


Figure 15: Chi2 distribution

- Calculations (R output): P-value = 0.9688 > 0.05. We do not reject H_0.
- Conclusion: At the 5% significance level, we fail to reject that the residuals are uncorrelated.

And thus we can conclude that the model does in fact capture the systematic variations in the data leaving behind uncorrelated residuals.

Next , we check the normality of the residuals by plotting them against the theoritical quantiles of the normal distribution

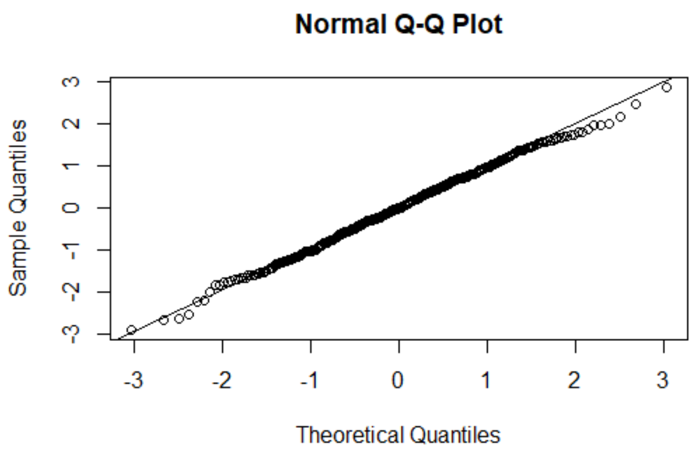


Figure 16: Q-Q plot of the residuals of Y3 process

The residuals quantiles coincide with the theoritical quantiles, thus we suspect that they are in fact normally distributed.

In order to confirm their normality, we conduct the Shapiro-Wilk normality test:

- Hypothesis:

$$H_0\text{: } \textit{the residuals are normally distributed}$$
$$H_1\text{: } \textit{the residuals are not normally distributed}$$

- Significance level: $\alpha = 0.05$
- Estimators: $\widehat{e_t}$
- Test statistic = W
- Assumption: "Large" T: number of observations
- Decision rule: Reject if or P-value<0.05.
- Calculations (R output): P-value =0.7625 > 0.05. We do not reject H_0.
- Conclusion: At the 5% significance level, we fail to reject that the residuals are normally distributed.

Forecasting

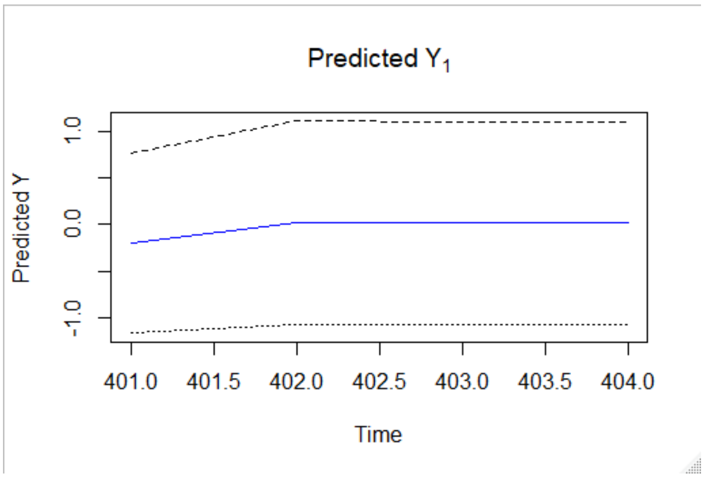


Figure 17: Predictions of model 3

The figure above, shows that the variance of the forecasted values using th e model Y1 is significantly large, but stable. This indicates that using this model to forecast future values comes with large uncertainty.
Next, we compare the fitted values to the actual values:

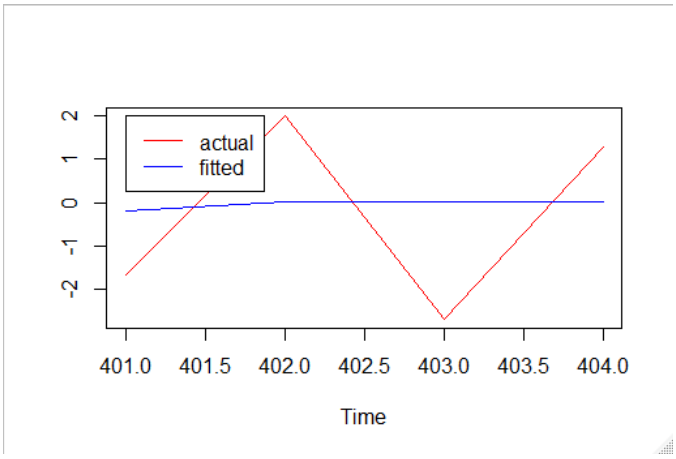


Figure 18: Actual vs fitted values of model 3

The actual values are quite different from the fitted values, in fact, the actual values fall beyong the forecasting bounds of the model.
This indicates that even though the systematic variations are in fact captured by Y1, the model fails to provide good enough forecasts for future values.

2.4 Process Y4

Identification

In order to identify the process, we first plot the time series, its ACF and PACF.

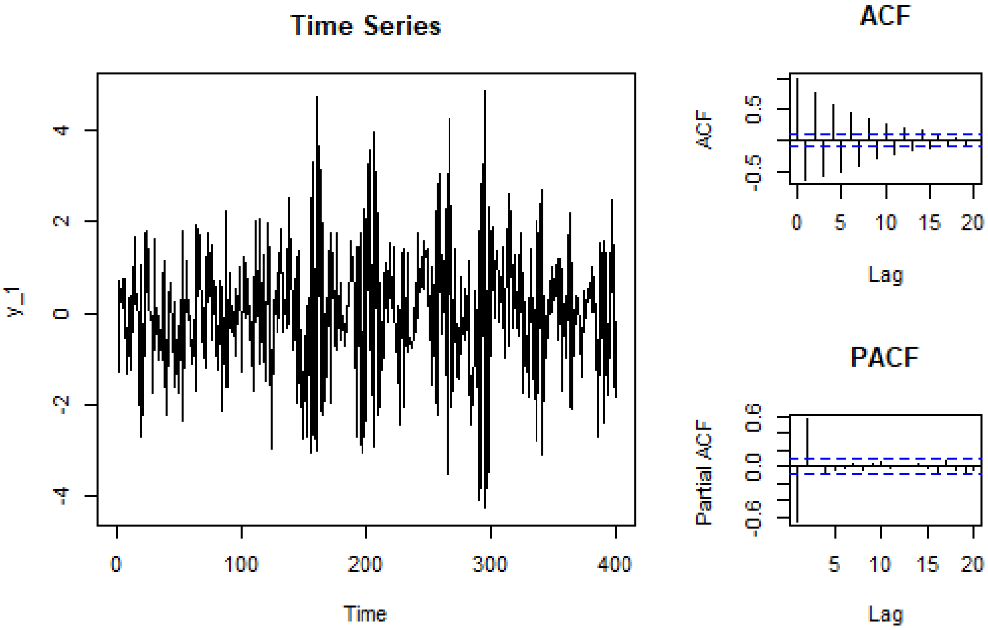


Figure 19: Time series plot, ACF and PACF of process Y4

After plotting the time series of the process Y4, we notice that it is choppy, with no visible drift or trend. The variance does not look stable throughout the process. This instability is less pronounced in process Y4 than Y2. The plot of the ACF starts with a positive spike then it keeps oscillating between positive and negative spikes. We notice that the first 15 correlations are actually significant so we fail to reject that they are 0. The other correlations are under the Z-test significance level so we can practically consider them 0. The PACF starts with a negative spike, a positive spike, then it dies off completely. These plots look like they belong to an AR(2) process.

Estimation

In order to estimate the model AR(2), we used the default procedure in R and we got:

$$\widehat{Y_t} = 0.0174 - 0.282Y_{t-1} + 0.568Y_{t-2} + e_t \tag{4}$$

Evaluation

In order to evaluate the process, we first plot the time series, its ACF and PACF of the residuals.

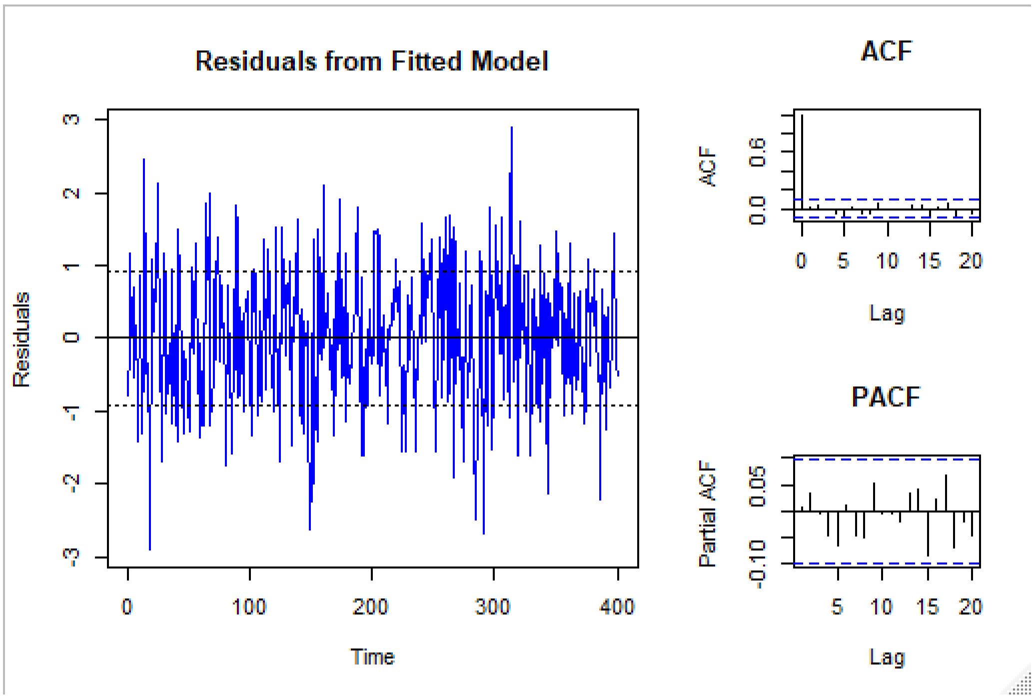


Figure 20: Time series plot, ACF and PACF of the residuals of process Y4

The plot of the residuals of the fitted model looks choppy. It has a constant mean=0 and the variance of the residuals seems to surpass the error bounds. However, the variance still seems stable. There is no visible drift or trend. The first spike of the ACF corresponds to the variance of the residuals which we already know is equal to 1. The ACF and PACF do not have any significant spikes which makes us suspect that the residuals are in fact uncorrelated. We test the correlation using the Ljung-Box Q test.

- Hypothesis:

$$H_0\colon \forall i \in \{1{,}20\}, \rho_i = 0$$
$$H_1\colon \textit{At least one } \ \rho_i \neq 0, i \in \{1{,}20\}$$

- Significance level: $\alpha = 0.05$
- Estimators: $\widehat{\rho_\imath}$, $i \in \{1{,}20\}$
- Assumption: "Large" T: number of observations
- Test statistic: $Q_{LB} = T(T+2)\sum_{j=1}^{20} \frac{\widehat{\rho_\jmath}^2}{T-j} \sim \mathcal{X}^2_{0.05,18}$ under H0.
- Decision rule and figure:

  Reject if ${Q_{LB}}^{Obs} > \mathcal{X}^2_{0.05,18}$ or P-value<0.05.

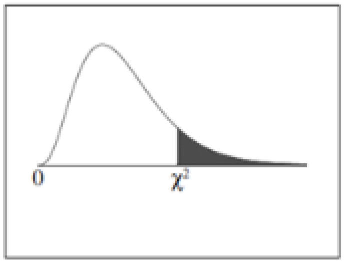


Figure 21: Chi2 distribution

- Calculations (R output): P-value= 0.9137 > 0.05. We reject H0.
- Conclusion: At the 5% significance level, we reject that the residuals are uncorrelated.

Forecasting

After finishing up with the model diagnosis, we use it to forecast future values of the process.

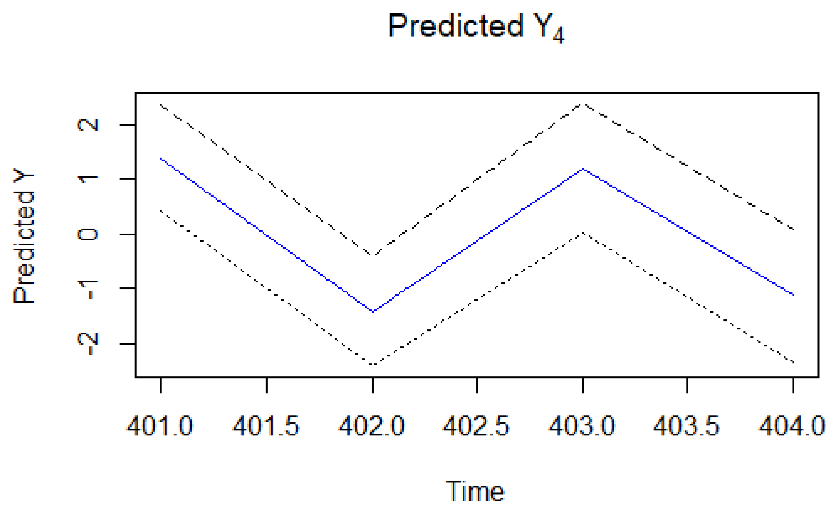


Figure 22: Predictions of model 4

The model can give us prediction values that we plotted with forecasting interval. We predict that the values we get are inside the forecasting interval. We also get point predictions that are indeed inside the forecasting interval. Let us now see if the predicted values are close to the real values.

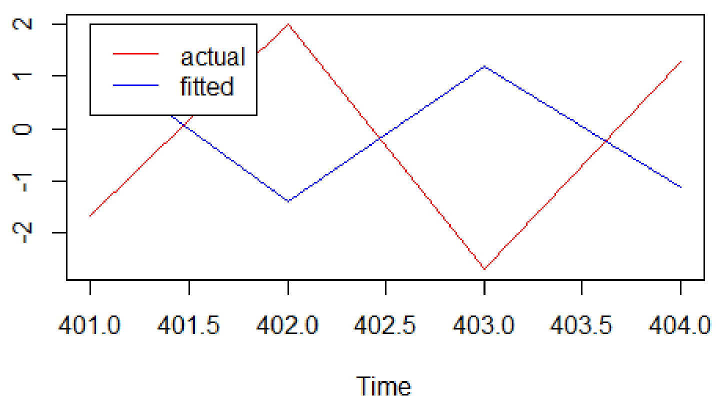


Figure 23: Actual vs fitted values of model 4

The forecasted values are not equal to the actual values of course due to having residuals. The actual values are not even inside the forecasting bounds which makes our model not accurate at forecasting.

2.5 Process Y5

Identification

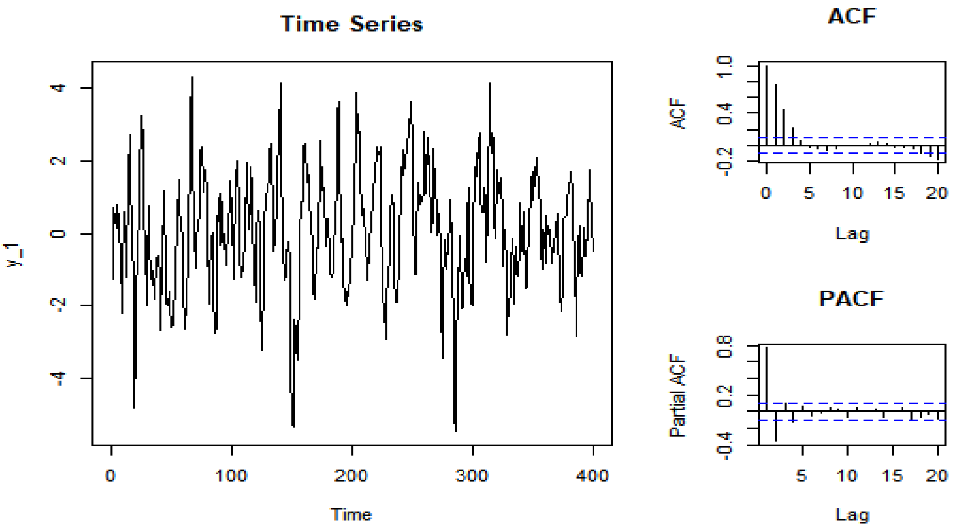


Figure 24: Time series plot, ACF and PACF of process Y5

The time series is choppy, it oscillates around the mean. No obvious trend can be seen. The process does exhibit instability in the mean and variance throughout the process.
The ACF has 4 upward spikes that exceed the cutoff value for the Z test for individual autocorrelation. The following spikes are barely noticeable and are well below the cutoff value and hence are non-significant.
The PACF starts with a significant upward spike at lag 0, then a downward significant spike at lag 1. After that the spikes are barely noticeable. They fall short of the cutoff line for the Z test for individual autocorrelation and hence are not significant. This looks like an ARMA(1,1) process.

Estimation

The estimated model is an ARMA (1,1) with an intercept:

$$Y_t = 0.0333 + 0.5829 Y_{t-1} + e_t - 0.5319 e_{t-1} \tag{5}$$

Evaluation

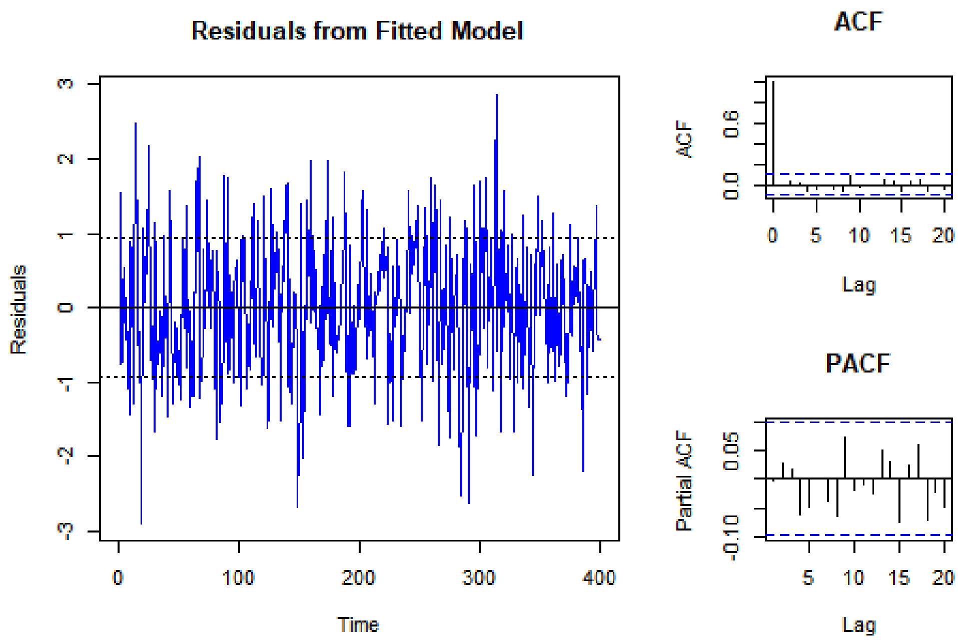


Figure 25: Time series plot, ACF and PACF of the residuals of process Y5

- Hypothesis:

$$H_0\colon \forall i \in \{1{,}20\}, \rho_i = 0$$
$$H_1\colon \textit{At least one } \rho_i \neq 0, i \in \{1{,}20\}$$

- Significance level: $\alpha = 0.05$
- Estimators: $\widehat{\rho_i}$, $i \in \{1{,}20\}$
- Assumption: "Large" T: number of observations
- Test statistic: $Q_{LB} = T(T+2)\sum_{j=1}^{20} \frac{\widehat{\rho_j}^2}{T-j} \sim \mathcal{X}^2_{0.05,18}$ under H0.
- Decision rule and figure:

  Reject if ${Q_{LB}}^{Obs} > \mathcal{X}^2_{0.05,18}$ or P-value<0.05.

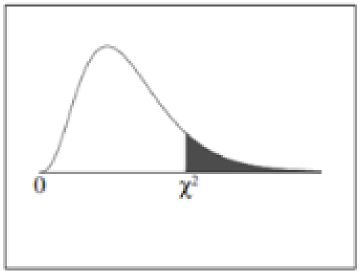


Figure 26: Chi2 distribution

- Calculations (R output): P-value = 0.8068 > 0.05. We do not reject H_0.
- Conclusion: At the 5% significance level, we fail to reject that the residuals are uncorrelated.

And thus we can conclude that the model does in fact capture the systematic variations in the data leaving behind uncorrelated residuals.

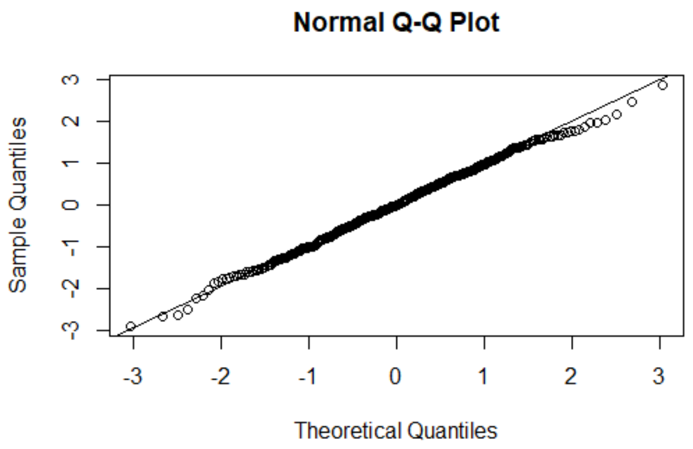


Figure 27: Q-Q plot of the residuals of Y5 process

The residuals quantiles coincide with the theoritical quantiles, thus we suspect that they are in fact normally distributed.

In order to confirm their normality, we conduct the Shapiro-Wilk normality test:

- Hypothesis:

$$H_0\text{: the residuals are normally distributed}$$

$$H_1\text{: the residuals are not normally distributed}$$

- Significance level: $\alpha = 0.05$
- Estimators: $\widehat{e_t}$
- Test statistic = W
- Assumption: “Large” T: number of observations
- Decision rule: Reject if or P-value<0.05.
- Calculations (R output): P-value =0.7694 > 0.05. We do not reject H_0.
- Conclusion: At the 5% significance level, we fail to reject that the residuals are normally distributed.

Forecasting

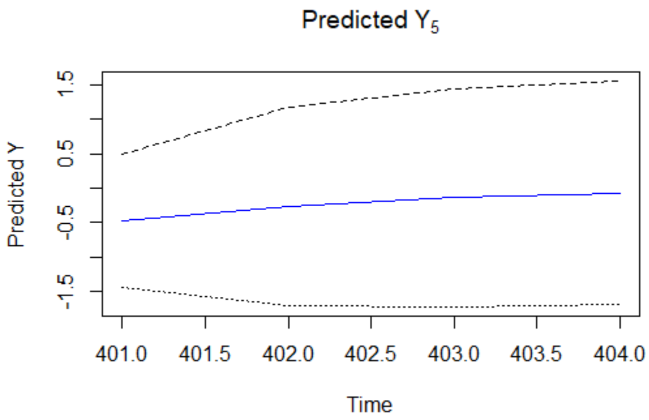


Figure 28: Predictions of model 5

The figure above, shows that the variance of the forecasted values using th e model Y5 is significantly large, but stable. This indicates that using this model to forecast future values comes with large uncertainty.
Next, we compare the fitted values to the actual values:

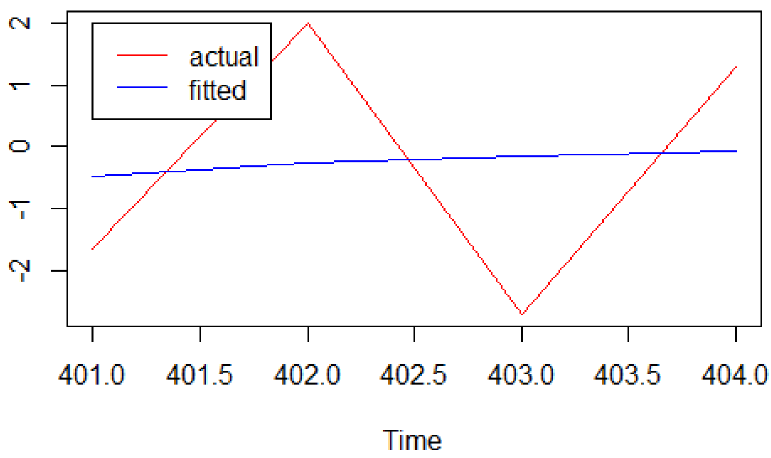


Figure 29: Actual vs fitted values of model 5

The actual values are quite different from the fitted values, in fact, the actual values fall beyond the forecasting bounds of the model.

This indicates that even though the systematic variations are in fact captured by Y5, the model fails to provide good enough forecasts for future values.

2.6 Forecasting Error Measures

		Evaluation measure		
Dat a	**Chosen Model**	**MSE**	**RMSE**	**MAPE**
$\mathbf{Y_1}$	AR (1)	3.908	1.977	99.806
$\mathbf{Y_2}$	AR(2)	4.330	2.081	105.956
$\mathbf{Y_3}$	MA(2)	3.744	1.935	96.577
$\mathbf{Y_4}$	AR(2)	10.440	3.231	171.518
$\mathbf{Y_5}$	ARMA(1,1)	3.3708	1.835	96.125

For each dataset we choose a model and evaluate it through 3 measures: MSE, RMSE, and MAPE. The model is better when it has lower value for these measures. For the fifth data set, we notice that the model has the lowest values for our three measures so it is the best model for that dataset.

3 Task 2

In this task we try to model real data. We chose to use a dataset for unemployment rates. According to the U.S. bureau of Labor Statistics:" This rate represents the number of unemployed as a percentage of the labor force. Labor force data are restricted to people 16 years of age and older, who currently reside in 1 of the 50 states or the District of Columbia, who do not reside in institutions (e.g., penal and mental facilities, homes for the aged), and who are not on active duty in the Armed Forces."[1]

We split our data into two datasets. One for modeling and one for evaluating forecasted values. The dataset for forecasting contains unemployment rates of the years 2019 and 2018.

In order to model the data, we first look at how it behaves.

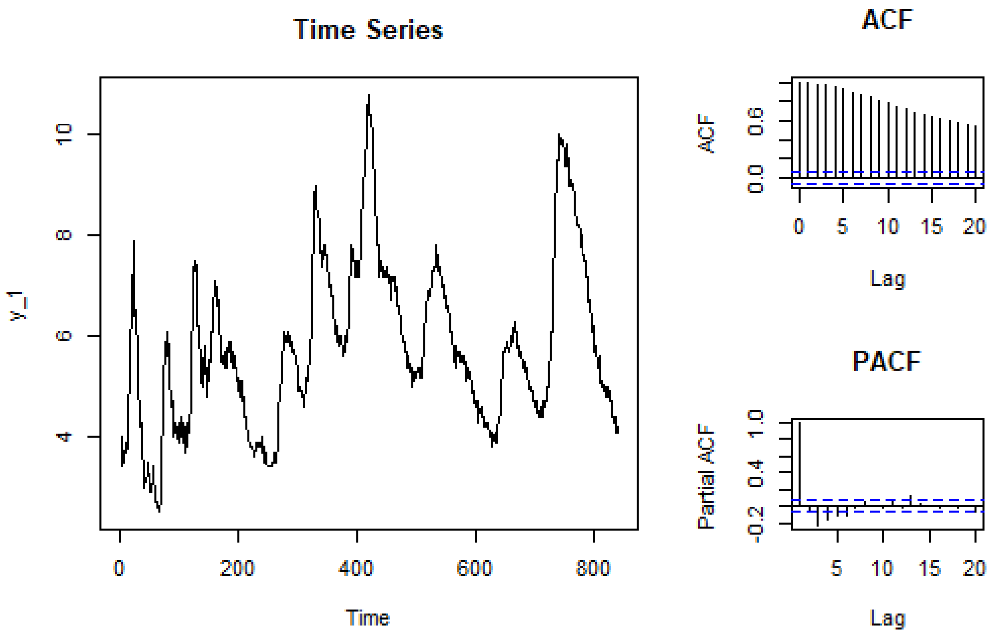


Figure 30:Time series plot, ACF and PACF of the unemployment rate process

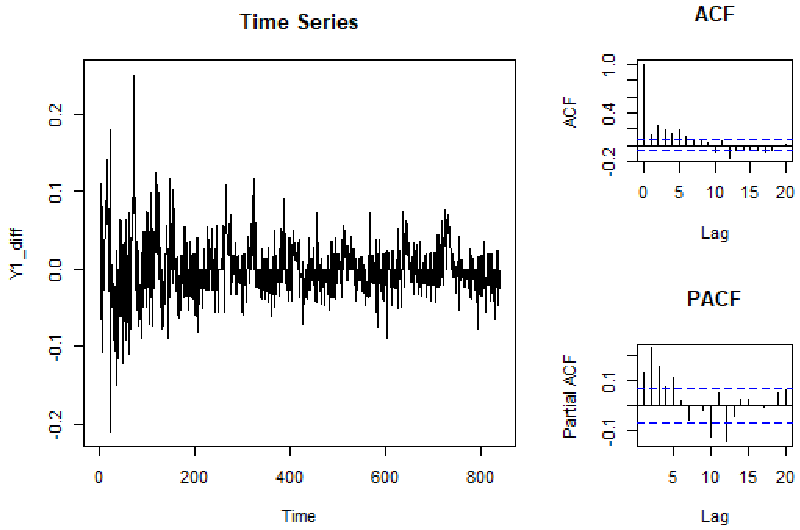
According to the plot, the time series is non-stationary, a long-term increase in the data can be noticed between 0 and the 400th lag, followed by a decrease in the following 200 lags and then another increase.

This indicates that the data exhibits a significant trend that changes direction twice.

The correlogram shows that the ACF decreases very slowly and does not die off confirming the non-stationarity of the time series.
The PACF starts with a significant upward spike and then oscillates between negative and positive values as it decays rather quickly.
The ACF and PACF resemble the ACF and PACF of an AR (2) or ARMA (1,1) process.

In order to model the data, we need to make it stationary.

In order to remove the trend component, we take the first difference and plot the series and the correlogram again



After detrending the time series using differencing, the plot of the time series shows no indication of a trend.

The series oscillates around the mean and has a finite and a relatively stable variance throughout time.

The correlogram shows that the ACF starts with an upward spike that reaches the value of 1 at lag 0, the dies off quickly after few lags indicating the stationarity of the series.

The fast decaying ACF resembles that of an AR process

The PACF starts with positives spikes for the first few lags and then oscillates between negative and positive values as it decays. It does not seem to cut off completely.

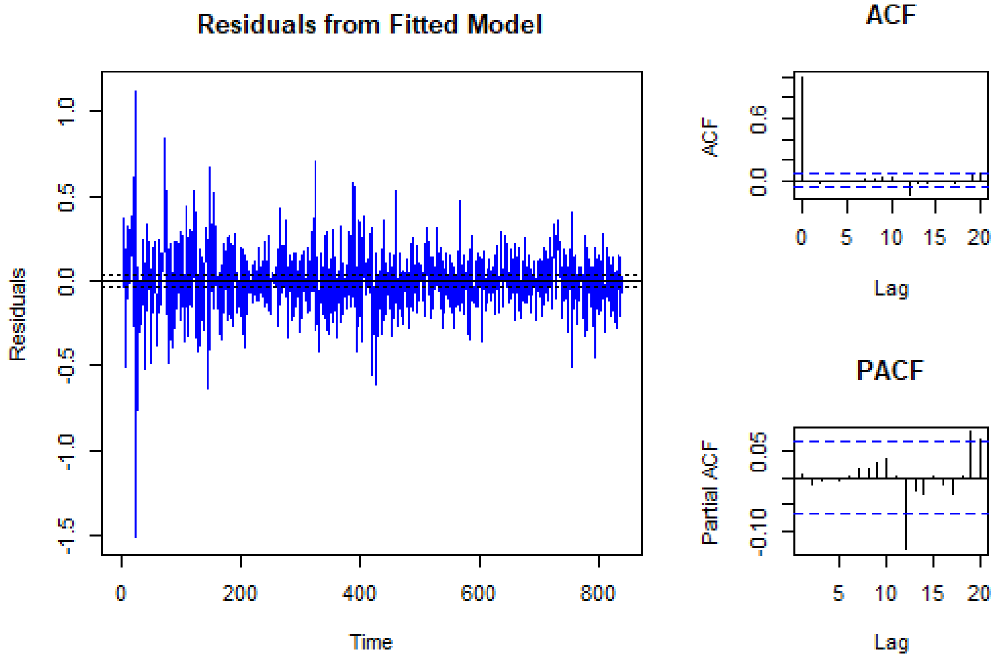
Theoretically, the order of an AR process can be identified through the PACF, where the order is equivalent to the lag where the PACF dies off.

The PACF has significant spikes until the 11^{th} lag (a significant autocorrelation is that that exceeds the bound of the Z test for individual correlation)

We fit an ARMA (11,0,0) to the differenced time series:

$$\begin{aligned}\widehat{Y_t} = 0.0124Y_{t-1} + 0.2239Y_{t-2} + 0.1501Y_{t-3} + 0.1029Y_{t-4} + 0.1424Y_{t-5} \\ + 0.009Y_{t-6} - 0.0552Y_{t-7} - 0.0051Y_{t-8} - 0.0227Y_{t-9} \\ - 0.1104Y_{t-10} + 0.0321Y_{t-11} + e_t\end{aligned}$$

In order to evaluate the process, we first plot the time series, its ACF and PACF of the residuals.



The plot of the residuals of the fitted model looks choppy. I has a constant mean=0 and the variance of the residuals seems to surpass the error bounds. The variance does not seem stable. There is no visible drift or trend. The first spike of the ACF corresponds to the variance of the residuals which we already know is equal to 1. The ACF and PACF do not seem to have any significant spikes which makes us suspect that the residuals are in fact uncorrelated. We test the correlation using the Ljung-Box Q test.

- Hypothesis:

$$H_0\colon \forall i \in \{1{,}20\}, \rho_i = 0$$
$$H_1\colon \textit{At least one } \ \rho_i \neq 0, i \in \{1{,}20\}$$

- Significance level: $\alpha = 0.05$
- Estimators: $\widehat{\rho_i}$, $i \in \{1{,}20\}$
- Assumption: "Large" T: number of observations
- Test statistic: $Q_{LB} = T(T+2)\sum_{j=1}^{20}\frac{\widehat{\rho_j}^2}{T-j} \sim \mathcal{X}^2_{0.05,18}$ under H0.
- Decision rule and figure:
  Reject if ${Q_{LB}}^{Obs} > \mathcal{X}^2_{0.05,18}$ or P-value<0.05.

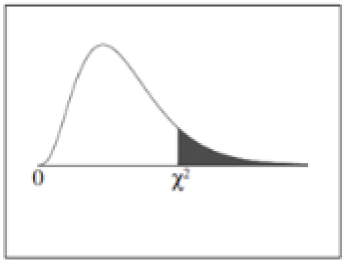
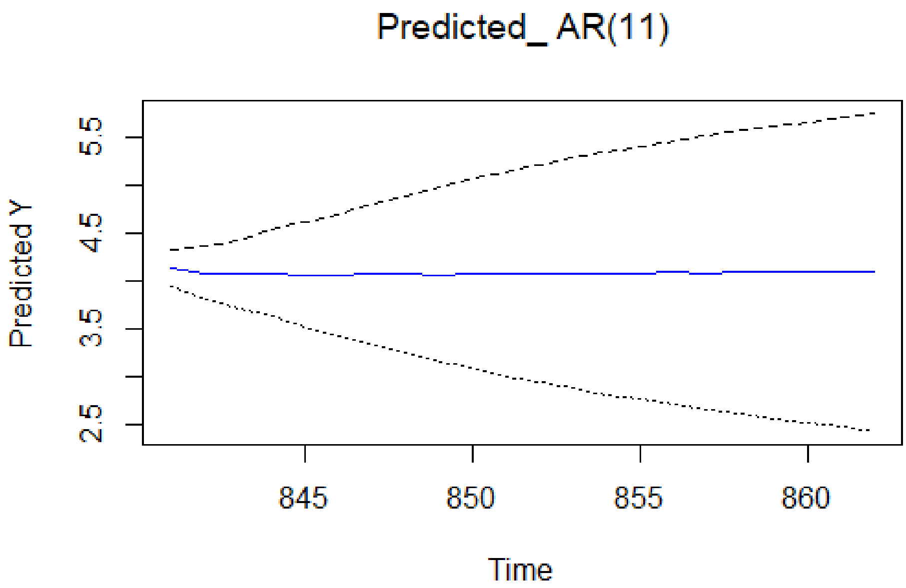


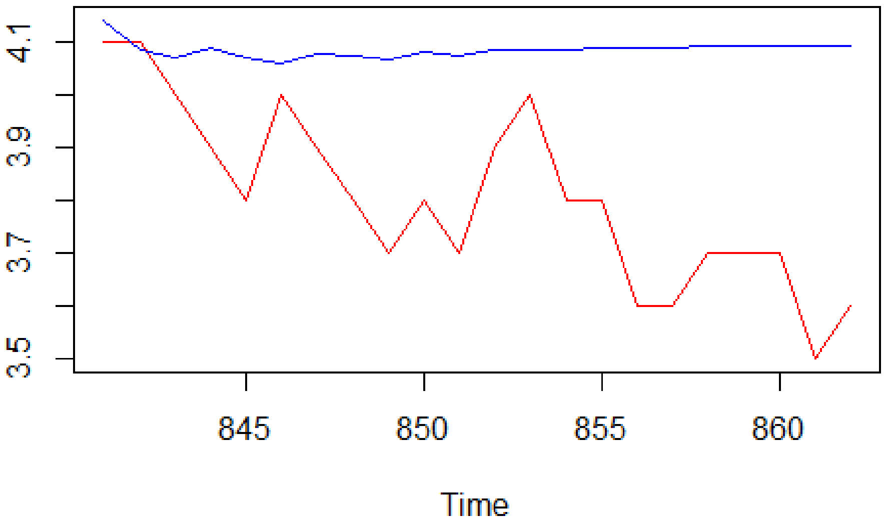
Figure 31: Chi2 distribution

- Calculations (R output): P-value= 0.0883 > 0.05. We do not reject H0.
- Conclusion: At the 5% significance level, we fail to reject that the residuals are uncorrelated.

After finishing up with the model diagnosis, we use it to forecast future values of the process.



The model can give us prediction values that we plotted with forecasting interval. We predict that the values we get are inside the forecasting interval. We also get point predictions that are indeed inside the forecasting interval. We notice that the forecasting uncertainty is bigger the more we forecast in the further future. Let us now see if the predicted values are close to the real values.



The red line corresponds to the actual values and the blue line corresponds to the fitted values. The forecasted values are not equal to the actual values of course due to having residuals. The actual values are close to the real values for the first months. However, the model's accuracy drops after the first couple of months.

	AIC	MSE	RMSE	MAPE
AR (11)	-350.93	0.1042	0.3229	7.55179

The error measure for the model are fairly low indicating an acceptable forecasting accuracy.

4 Conclusion

In this project, we have worked with five different simulations for five processes. We tried to choose the best model for each process using the Box-Jenkins method. When the expected model cannot capture the variation of our data we try a different model. We then tried to work with real data. We chose a dataset for monthly unemployment rates in the U.S. from 1948 until 2019. We modeled the series with an AR(11) process , the model seems to capture the variations in the data

5 Reference

[1] https://fred.stlouisfed.org/series/UNRATE