Time Series Analysis
Assignment 2

Authors

Daniel Gonzalvez Alfert- s240404
Spiliopoulos Charalampos - s222948
Maria Kokali - s232486
Marios-Dimitrios Lianos - s233558

July 5, 2024

Contents

1 Stability **1**

1.1 Determine if the process is stationary for $\phi_1 = -0.7$ and $\phi_2 = -0.2$ by analyzing the roots of the characteristic equation 1

1.2 Is the process invertible? 1

1.3 Write the autocorrelation $\rho(k)$ for the AR(2) process as function of ϕ_1 and ϕ_2 1

1.4 Plot the autocorrelation $\rho(k)$ up to nlag = 30 for the coefficient values above 2

1.5 Simulate 5 realizations of the process up to n = 200 observations with the coefficient values given above. Plot them in one plot 3

1.6 Calculate the empirical ACF of the simulations and plot them together with $\rho(k)$, up to lag 30 4

1.7 Keep $\phi_2 = -0.2$ and redo plots of simulations and ACF for the following $\phi 1 - values$, $\phi 1$ = -0.2, $\phi 1$ = 0.7, $\phi 1$ = -0.8, $\phi 1$ = -0.85 5

1.8 $\phi_1 = 0.7$ 7

1.9 $\phi_1 = -0.8$ 9

1.10 $\phi_1 = -0.85$ 11

1.11 Would you recommend always plotting the time series data or does it provide sufficient information just to examine the ACF? 13

2 Predicting Monthly Solar Power **14**

2.1 Calculation of residuals and model validation 15

2.2 With the specified model calculate $\hat{Y}_{t+k|t}$ for t = 36 and k = 1,...,12, i.e. predict the power for the following twelve months. 18

2.3 Calculate 95% prediction intervals for the twelve months ahead and add them to the plot. 20

2.4 Would you trust the forecast? Do you think the prediction intervals have correct width (all the time)? 22

3 Simulating Seasonal Processes **23**

4 Identifying ARMA Models **35**

4.1 Guess the ARMA model structure, and give a short reasoning of your guess . 35

4.2 Guess the ARMA model structure, and give a short reasoning of your guess . 36

4.3 Guess the ARMA model structure, and give a short reasoning of your guess . 37

List of Figures **I**

List of Tables **II**

1 Stability

1.1 Determine if the process is stationary for $\phi_1 = -0.7$ and $\phi_2 = -0.2$ by analyzing the roots of the characteristic equation

The process is stationary if the roots of the characteristic equation $\phi(z^{-1}) = 0$, with respect to z, all lie within the unit circle.
We introduce the backward shift operator B, defined as $BX_t = X_{t-1}$.

$\phi(B)$ = 1 - 0.7B - 0.2B^2

1 - 0.7z^{-1} $- 0.2z^{-2} = 0$

We need to solve this quadratic equation to find the roots. The AR(2) process is stationary if and only if all the roots of the characteristic equation lie inside the unit circle (i.e., their absolute values are greater than 1).

$z_1 = 0.9178908 - 0i$
$z_2 = -0.2178908 + 0i$

It can be observed that the roots of the characteristic equation (related to $\phi1$ and $\phi2$) lie inside the unit circle, thus the process is weakly stationary. This condition ensures that the mean and variance are constant over time and that the autocovariance function depends only on the lag. For linear processes like AR(2) with normally distributed errors, weak stationarity implies strong stationarity because the normal distribution is fully specified by its mean and variance.

1.2 Is the process invertible?

A time series model is said to be invertible when it can be rewritten as an infinite weighted sum of past white noise (error) terms.
An AR(p) process is always invertible (Theorem 5.9: Texts in Statistical Science Time Series Analysis, p. 120)

1.3 Write the autocorrelation $\rho(k)$ for the AR(2) process as function of ϕ_1 and ϕ_2

$\rho(k) + \phi1\ \rho(k-1) + \phi2\ \rho(k-2) = 0$, k = 1, 2, ...

Overall the autocorrelation function of the AR(2) process is given by:

$\rho(k) = 1, k = 0$

$(1) \rightarrow \rho(1) = -\frac{\phi_1}{\phi_2+1}, k = 1$

$(2) \rightarrow \rho(2) = \frac{\phi_1^2}{\phi_2+1} - \phi_2, k = 2$

(3) $\rho(k)$ = $-\phi1$ $\rho(k-1)$ $-\phi2$ $\rho(k-2)$, k > 2

1.4 Plot the autocorrelation $\rho(k)$ up to nlag = 30 for the coefficient values above

In Figure 1 we can see that the behavior of ACF exponentially decays to 0 as the lag increases, something we were expecting due to the value of $\phi1$.

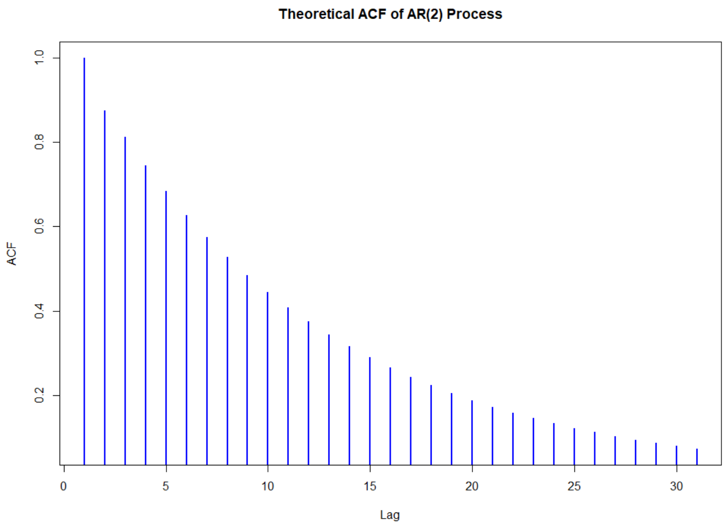

Figure 1: Theoritical Autocorrelation Function AR(2)

1.5 Simulate 5 realizations of the process up to n = 200 observations with the coefficient values given above. Plot them in one plot

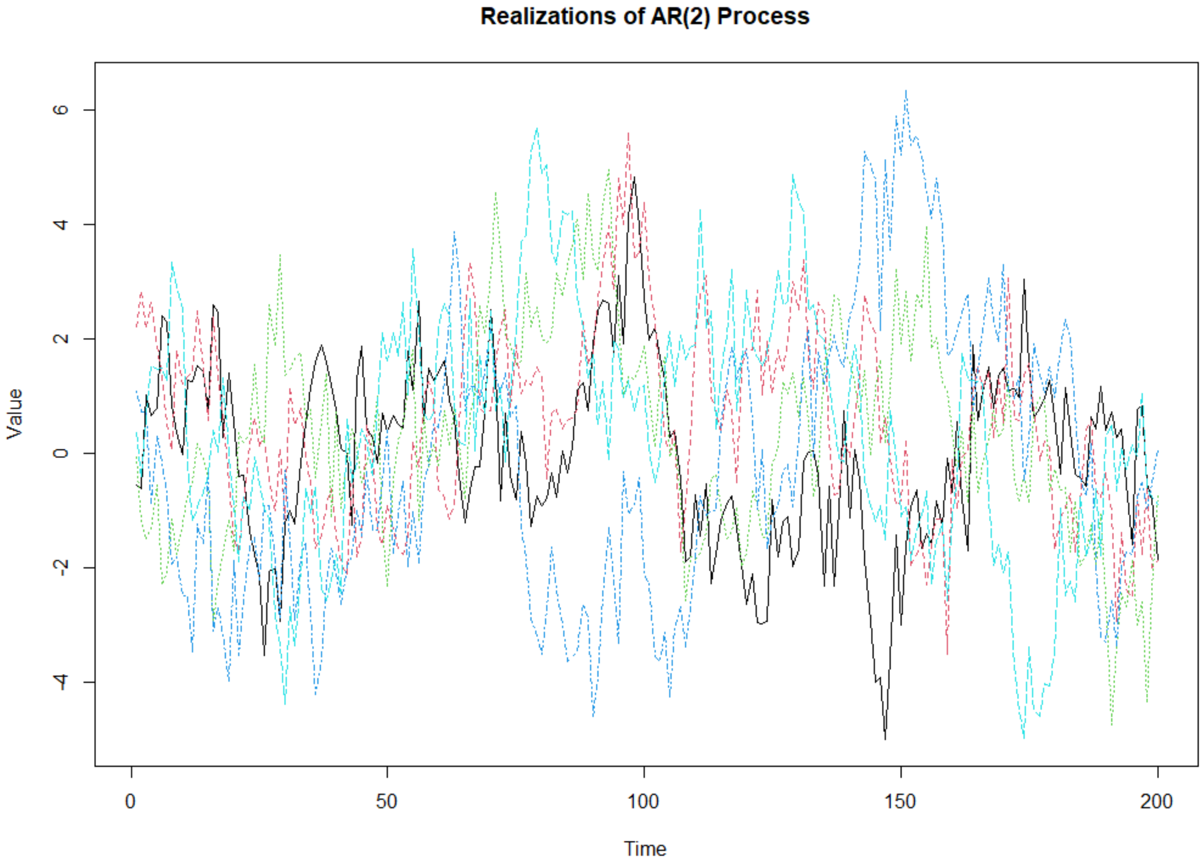

Figure 2: Realizations of AR(2) Process

1.6 Calculate the empirical ACF of the simulations and plot them together with $\rho(k)$, up to lag 30

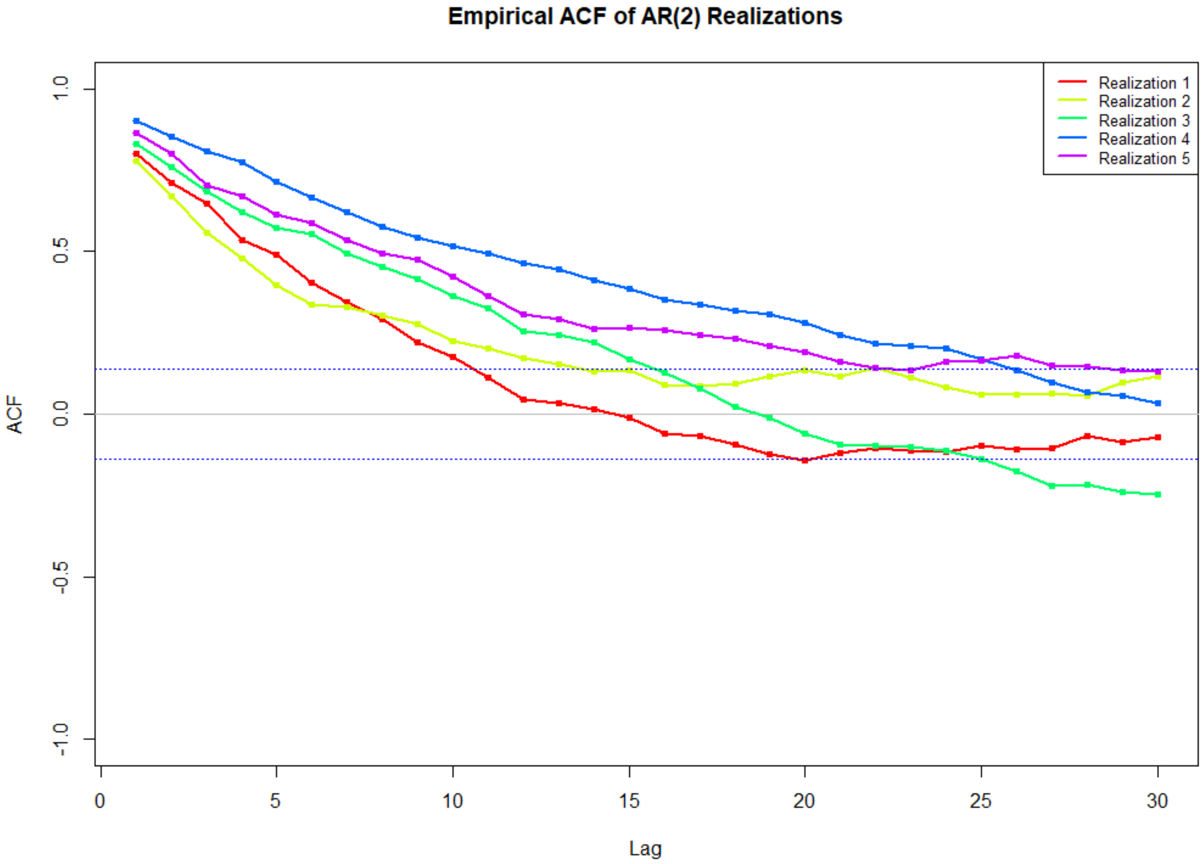

Figure 3: ACF for each realization of the process Xt

We estimate the ACF for each realization and then we plot the ACF for each realization in Figure 3. Observing this figure, we can conclude that the ACF is truncated (converges to zero) after lag 2, which is in compliance with the theoretical autocorrelation function, although with some noise.

1.7 Keep $\phi_2 = -0.2$ and redo plots of simulations and ACF for the following $\phi 1 - values$, $\phi 1$ = -0.2, $\phi 1$ = 0.7, $\phi 1$ = -0.8, $\phi 1$ = -0.85

$\underline{\phi 1 = -0.2}$
1 - 0.2B - 0.2B^2 = 0

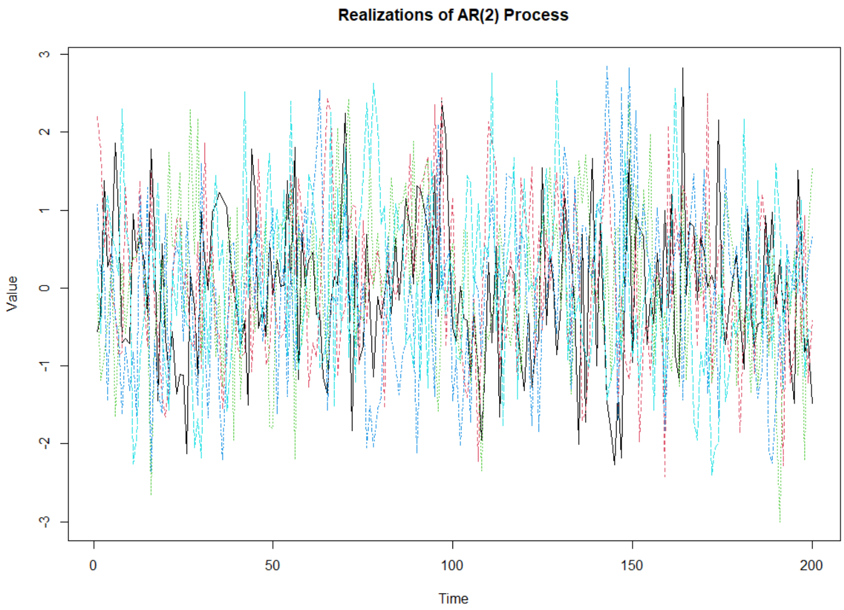

Figure 4: Realizations of AR(2) Process

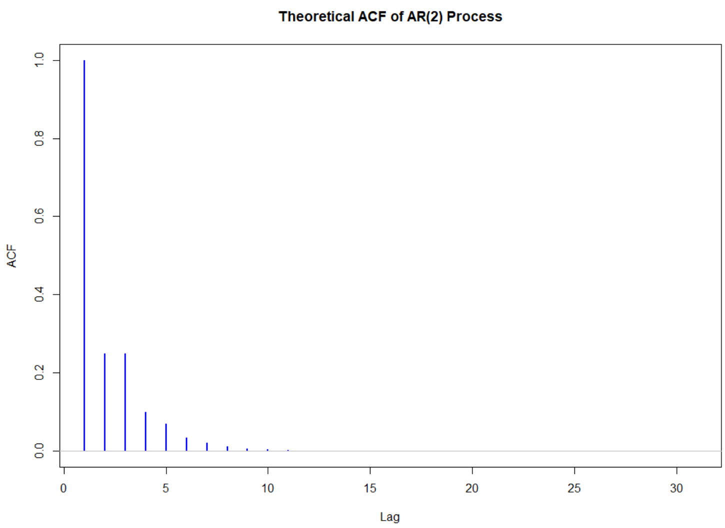

Figure 5: Theoretical ACF

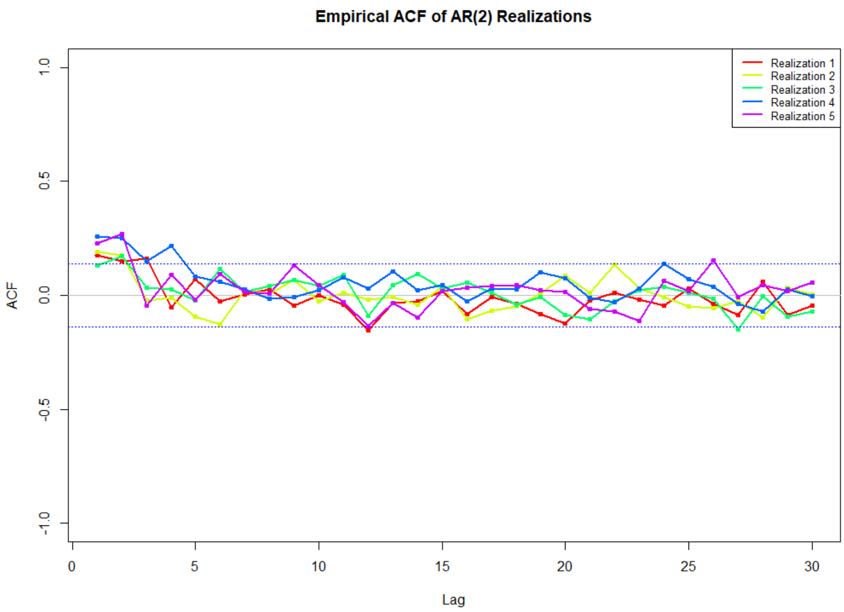

Figure 6: ACF for each realization of the process Xt

Parameter	$\phi 1$	z1	z2
Value	0.2	0.5582576+0i	-0.3582576-0i

The process is stationary. In Figure 6 in all realizations we can see that the ACF shows a quick drop-off after the second lag. ACF fluctuates and generally follows the pattern of the theoretical ACF, although with some noise.

1.8 $\underline{\phi_1 = 0.7}$

$\phi(B)$ = 1 + 0.7B - 0.2B^2

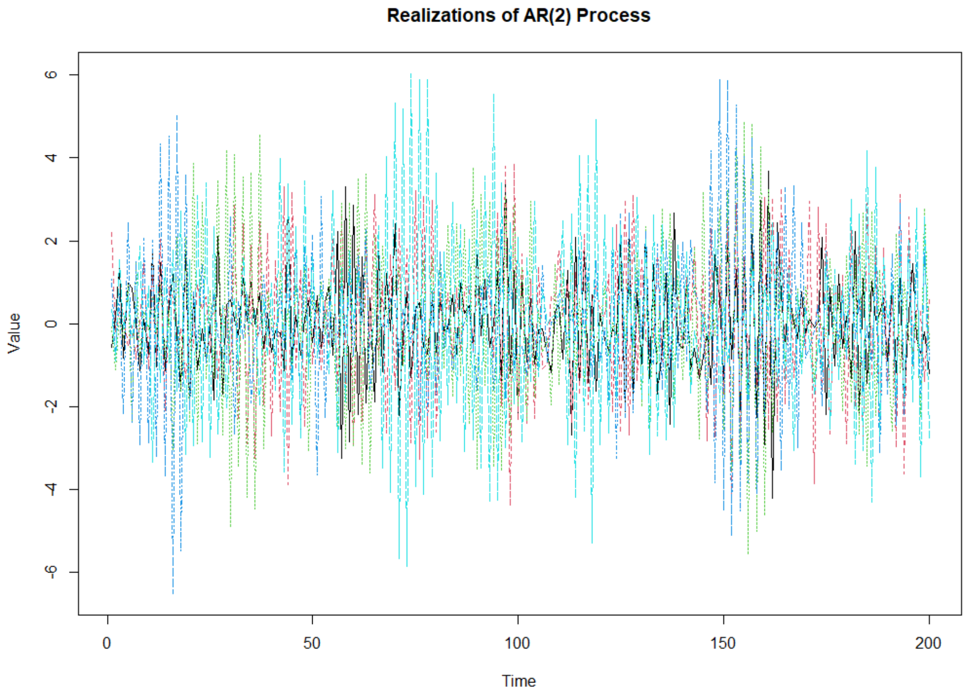

Figure 7: Realizations of AR(2) Process

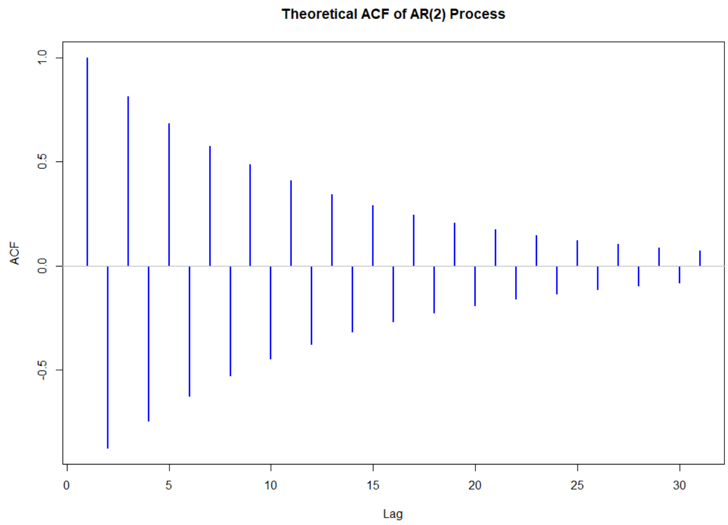

Figure 8: Theoretical ACF

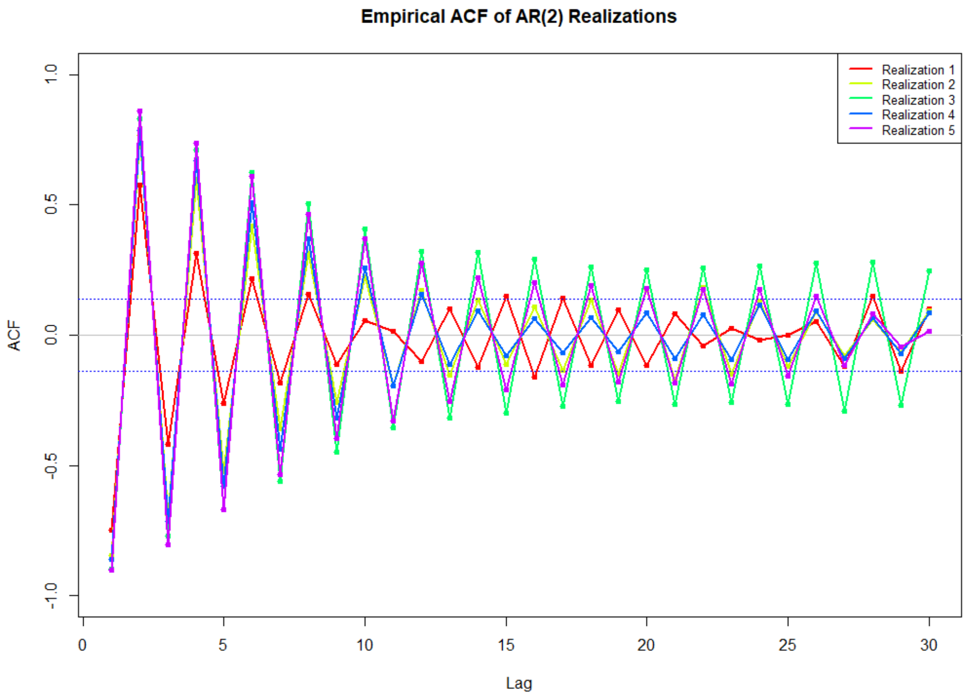

Figure 9: ACF for each realization of the process Xt

Parameter	$\phi 1$	z1	z2
Value	0.7	-0.9178908+0i	0.2178908+0i

The process is stationary. In Figure 9 in all realizations we can see that the ACF shows a drop-off, with the difference this time that the ACF exponentially decreases to 0 but the algebraic signs for the autocorrelations alternate between positive and negative. Again the empirical ACF aligns with the theoretical.

1.9 $\underline{\phi_1 = -0.8}$

$\phi(B)$ = 1 - 0.8B - 0.2B^2

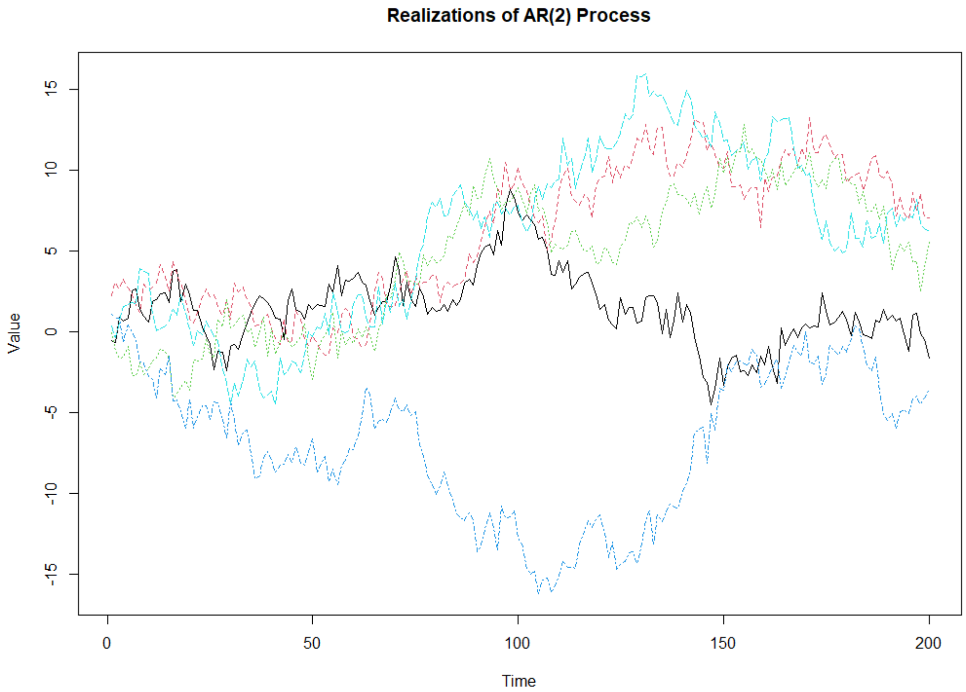

Figure 10: Realizations of AR(2) Process

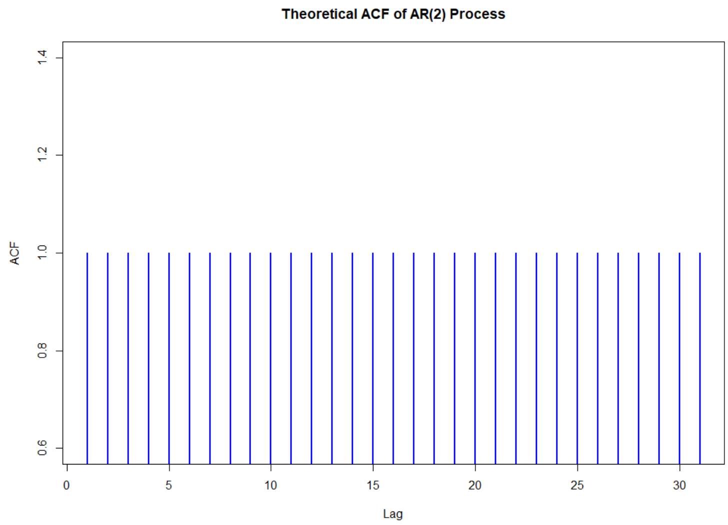

Figure 11: Theoretical ACF

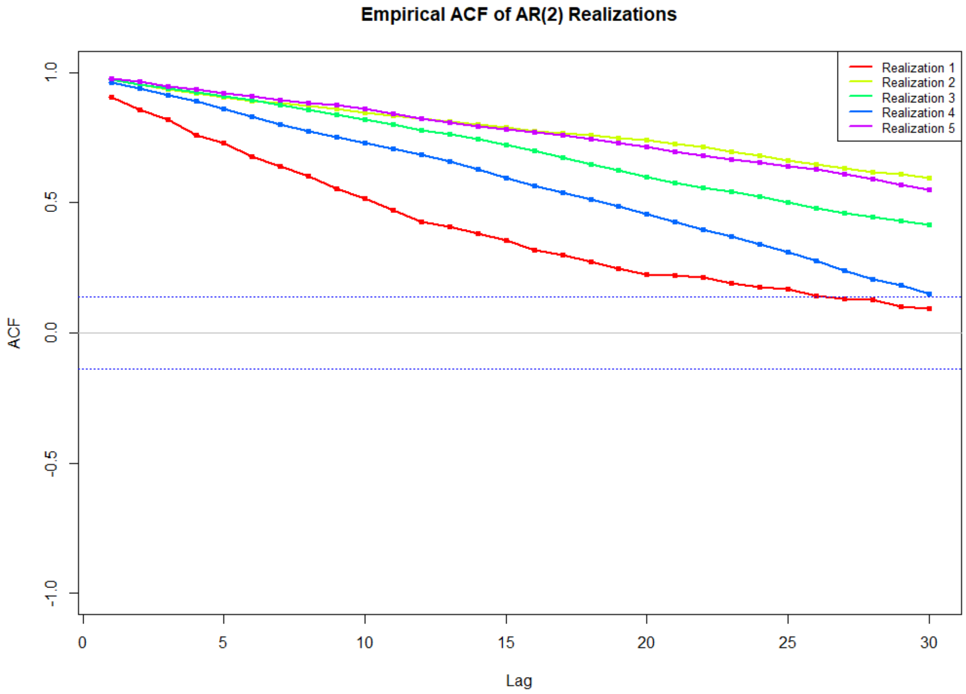

Figure 12: ACF for each realization of the process Xt

Parameter	$\phi 1$	z1	z2
Value	-0.8	1.0-0i	-0.2+0i

The process is not stationary $z_1 = -1 + 0i$. Figure 11 displays the theoretical autocorrelation function (ACF) for an AR(2) process, indicating regular oscillations between lags, which is characteristic of an AR process where one of the roots of the characteristic equation is on the unit circle (specifically, $z_1 = 1$). This indicates that the process is at the boundary between stationarity and non-stationarity In Figure 12, the empirical ACF from five realizations of an AR(2) process is shown, each revealing a decreasing trend in autocorrelation as the number of lags increases, with slight variations between each realization.

1.10 $\underline{\phi_1 = -0.85}$

$\phi(B)$ = 1 - 0.85B - 0.2B^2

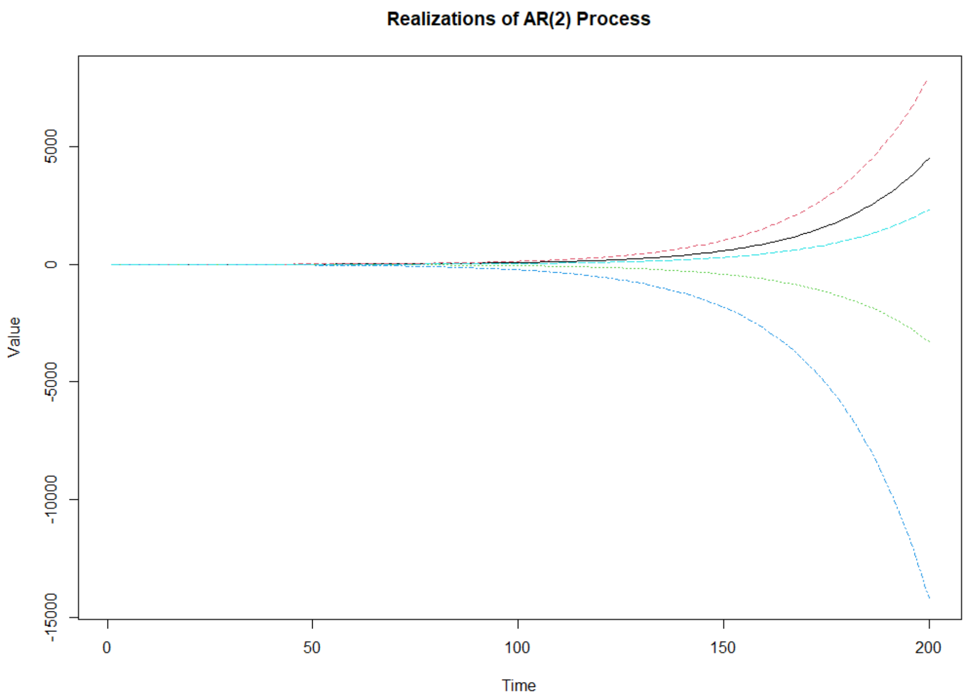

Figure 13: Realizations of AR(2) Process

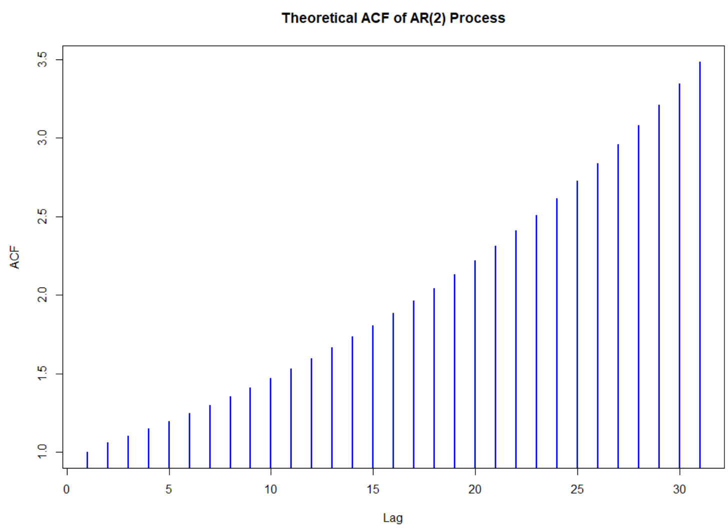

Figure 14: Theoretical ACf

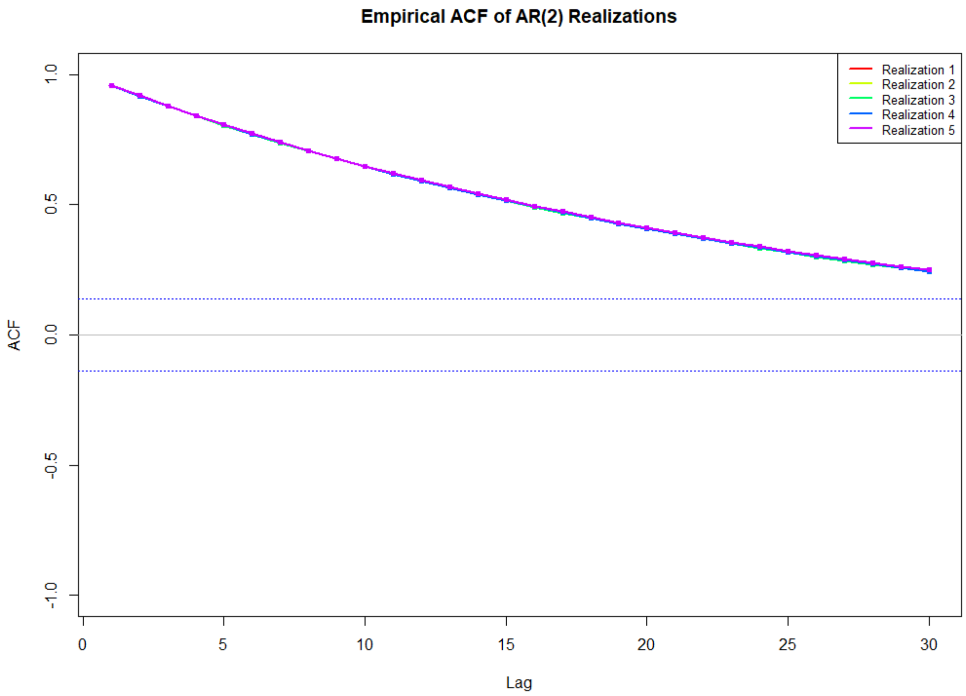

Figure 15: Autocorrelation $\gamma(k)$ up to nlag=30

Parameter	ϕ1	z1	z2
Value	-0.85	1.0419481-0i	-0.1919481+0i

The process is not stationary since $z_1 = 1.0419481 - 0i$. Figure 11 displays the theoretical autocorrelation function (ACF) for an AR(2) process, indicating regular oscillations between lags, which is characteristic of a certain type of AR process behavior. In Figure 12, the empirical ACF from five realizations of an AR(2) process is shown, each revealing a decreasing trend in autocorrelation as the number of lags increases, with slight variations between each realization. We can also see that the five ACF functions are almost identical which

1.11 Would you recommend always plotting the time series data or does it provide sufficient information just to examine the ACF?

We wouldn't recommend relying solely on either plotting the time series data or examining the ACF. Both provide valuable but complementary information for time series analysis. Here's a breakdown of why:

Why Plot the Time Series Data

- Trends and Seasonality: Visualizing the raw data helps reveal obvious trends (upward or downward), seasonality (regularly repeating patterns), and potential outliers that might influence the ACF.
- Non-Stationarity: Plotting the data can give us an initial indication of potential non-stationarity (e.g. if the mean or variance appears to change over time).
- Breaks: We might discover shifts in the time series or abrupt changes in behavior.
- Data Quality Issues: Sometimes, plotting can expose errors or inconsistencies in the data recording process.

Why Examine the ACF

- Autocorrelation Structure: The ACF specifically quantifies how correlated a time series is with itself at different lags. This is key for identifying potential AR or MA components in the process.
- Suitable Models: The pattern of decay (or cut-off) in the ACF suggests which types of models (AR, MA, or ARMA) might best fit the data.
- Seasonality: If strong seasonal patterns are present, we'll see them reflected in the ACF at lags corresponding to the seasonal period.

2 Predicting Monthly Solar Power

In this exercise we deal with forecasting in renewable energy, particularly for solar PV plants.

We have a seasonal AR(1) model, based on 36 observations of historical data, which incorporates seasonal effects and a white-noise process to account for variability, with parameters $\phi_1 = -0.38$, $\Phi_1 = -0.94$, $\mu = 5.72$, and noise variance $\sigma_\epsilon^2 = 0.22^2$. The AR model is the following:

$$(1 + \phi_1 B)(1 + \Phi_1 B^{12}) (\log(Y_t) - \mu) = \epsilon_t$$

where Y_t is the monthly energy from the plant in MWh, and ϵ_t is the white-noise process. We have observations for each month of the years 2008, 2009 and 2010.

A plot of the time series is

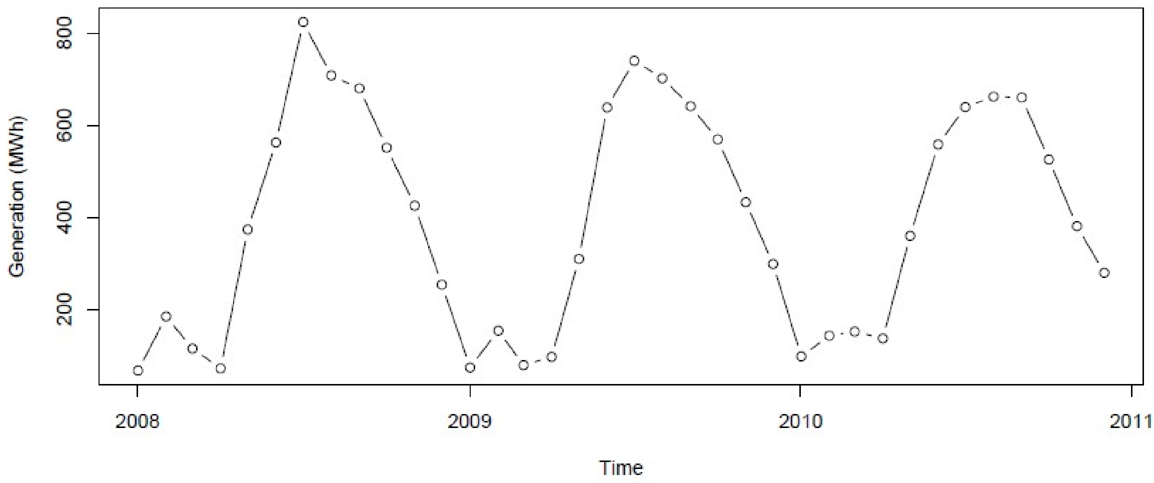

Figure 16: Our Time Series

To be more specific, in our AR(1) model, ϕ_1 is the parameter for the non-seasonal AR part, Φ_1 is the parameter for the seasonal AR part and the number 12 represents the number of periods in a season.

We can say that our seasonal AR(1) model is equivalent to an ARIMA$(1, 0, 0)x(1, 0, 0)_{12}$ process, which has one non-seasonal AR term and one seasonal AR term, without any differencing or moving average terms, both seasonally and non-seasonally.

2.1 Calculation of residuals and model validation

We introduce $X_t = \log(Y_t) - \mu$ and we substitute it in the above model equation for simplification and finally we have:

$$(1 + \phi_1 B)(1 + \Phi_1 B^{12})X_t = \epsilon_t.$$

To calculate the residuals $\hat{\epsilon}_{t+1|t}$, representing the difference between the observed values X_{t+1} and the forecasted values $\hat{X}_{t+1}$, we follow these steps:

1. Forecast X_{t+1} using the model based on available data.
2. Calculate the residual as $\hat{\epsilon}_{t+1|t} = X_{t+1} - \hat{X}_{t+1|t}$.

Therefore, our model equation will be written in the form:

$$X_t = -\phi_1 X_{t-1} - \Phi_1 X_{t-12} - \phi_1 \Phi_1 X_{t-13} + \varepsilon_t$$

or

$$X_{t+1} = -\phi_1 X_t - \Phi_1 X_{t-11} - \phi_1 \Phi_1 X_{t-12} + \varepsilon_{t+1}$$

and for our predictions we have:

$$\hat{X}_{t+1|t} = E[X_{t+1}|X_t, X_{t-1}, \ldots]$$

$$\hat{X}_{t+1|t} = E[-\phi_1 X_t - \Phi_1 X_{t-11} - \phi_1 \Phi_1 X_{t-12} + \varepsilon_{t+1}|X_t, X_{t-1}, \ldots]$$

$$\hat{X}_{t+1|t} = -\phi_1 X_t - \Phi_1 X_{t-11} - \phi_1 \Phi_1 X_{t-12}$$

Given that $s = 12$, we cannot calculate the residuals for the first 12 periods because we do not have a full season's worth of data prior to these points. Therefore, for our seasonal model, the first residual can only be calculated starting from $t = 13$, which corresponds to the first observation of the second year. So, we have:

Time (t)	Forecasted ($\hat{X}_t$)	Residual (ε_t)
13	-0.47134647	-0.958194085
14	-1.00639810	0.316836022
15	-1.53651765	0.173226474
16	0.27089078	-1.426542584
17	0.50787761	-0.494536336
18	0.99666039	-0.258322102
19	0.77370681	0.112943381
20	0.76994089	0.063992510
21	0.55207437	0.190955090
22	0.33787071	0.286009724
23	-0.16163441	0.510059996
24	-1.28581533	1.262908812
25	-0.57276629	-0.572522727
26	-1.32556785	0.561394904
27	-0.86637875	0.163658592
28	0.11854848	-0.925893597
29	0.75133499	-0.588012601
30	0.79937315	-0.195014184
31	0.74834965	-0.008445192
32	0.69516759	0.080097967
33	0.61448862	0.157751212
34	0.31116130	0.232236961
35	-0.06232467	0.282495919
36	-1.10212900	1.013340778

Table 1: Predicted values and residuals for our time series.

Model Validations:

We will do a model validation by checking the assumptions of i.i.d. errors, meaning that the residuals should be independent and identically distributed.

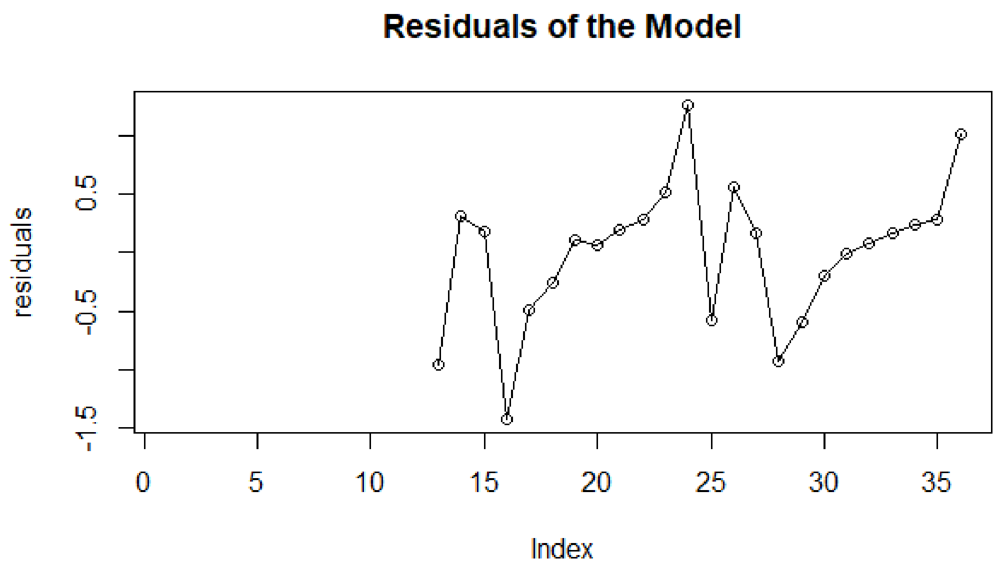

Figure 17: Residuals

In Figure 13, which is the plot of the residuals, there's no obvious pattern or trend and the residuals seem to vary around the same level without any clear signs of changing variance. However, some points seem to stand out, possibly indicating outliers.Also, it appears that the mean of the residuals could be around zero, as they oscillate above and below the zero line.

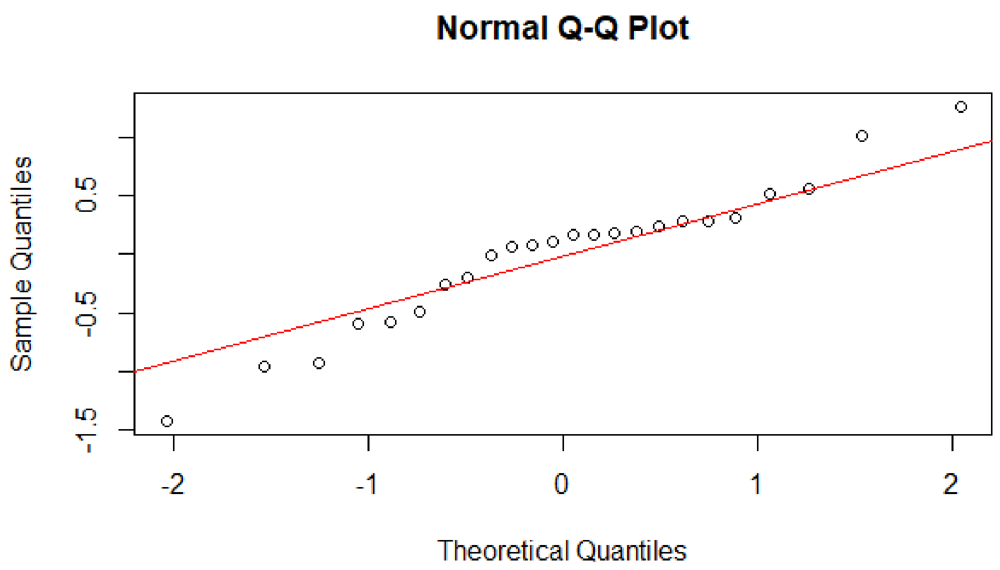

Figure 18: QQplot

In the QQplot, in Figure 14, the values of the residuals seem to fit approximately well with the expected values from the normal distribution (red line), suggesting that they are normally distributed and the deviations at the tails could be due to outliers.

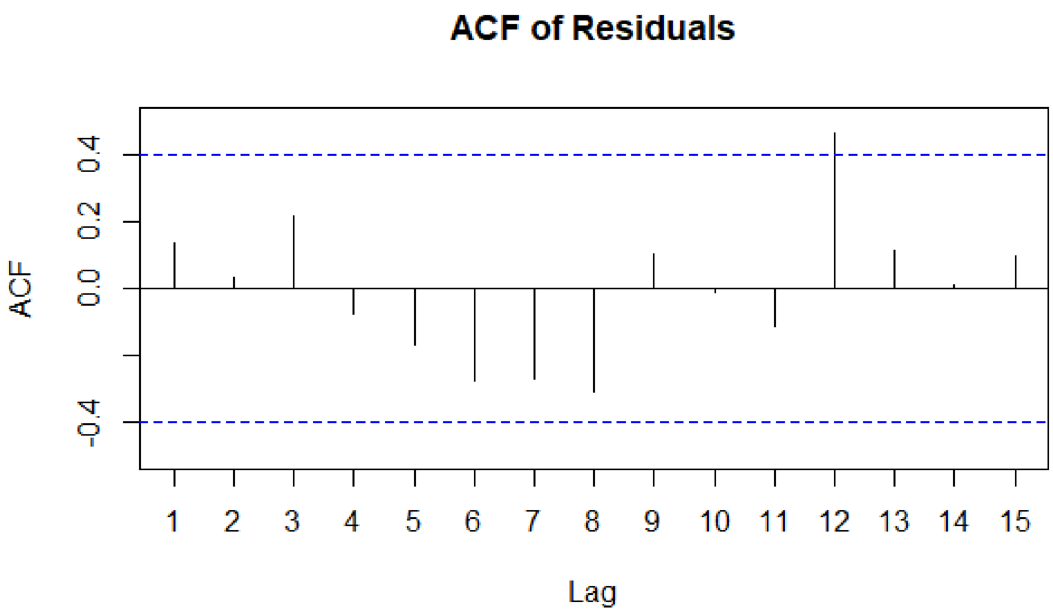

Figure 19: Autocorrelation in residuals

In Figure 15, the ACF plot shows the correlation of the residuals with themselves at different lags. All the autocorrelation coefficients at lags up to 15 are within the blue dashed confidence bounds. This suggests that there is no significant autocorrelation in the residuals, which supports the independence assumption of the i.i.d. errors.

Therefore, the residuals meet the i.i.d assumptions in general.

2.2 With the specified model calculate $\hat{Y}_{t+k|t}$ for t = 36 and k = 1,...,12, i.e. predict the power for the following twelve months.

First, we are going to predict $\hat{X}_{t+k|t}$ for t = 36 and k = 1,...,12.

For $k = 1$, we have that:

$$\hat{X}_{37|36} = -\phi_1 X_{36} - \Phi_1 X_{25} - \phi_1 \Phi_1 X_{24}$$

For $1 < k < 12$ we have that:

$$\hat{X}_{t+k|t} = -\phi_1 \hat{X}_{t+k-1|t} - \Phi_1 \mathrm{X}_{24+k} - \phi_1 \Phi_1 \mathrm{X}_{23+k}$$

For k=12 we have that:

$$\hat{X}_{48|36} = -\phi_1 \hat{X}_{47|36} - \Phi_1 X_{36} - \phi_1 \Phi_1 X_{35}$$

After, we have calculated the forecasted values $\hat{X}_{t+k|t}$, we get the values $\hat{Y}_{t+k|t}$ from:

$$\hat{Y}_{t+k|t} = e^{\hat{X}_{t+k|t}+\mu}$$

Index (k)	Forecasted Value ($\hat{Y}_{36+k|36}$)
1	101.2782
2	147.2255
3	156.9226
4	142.5501
5	355.3098
6	538.0196
7	611.2055
8	631.8932
9	630.1100
10	508.1568
11	375.0125
12	280.4901

Table 2: Forecasted values of $\hat{Y}_{36+k|36}$.

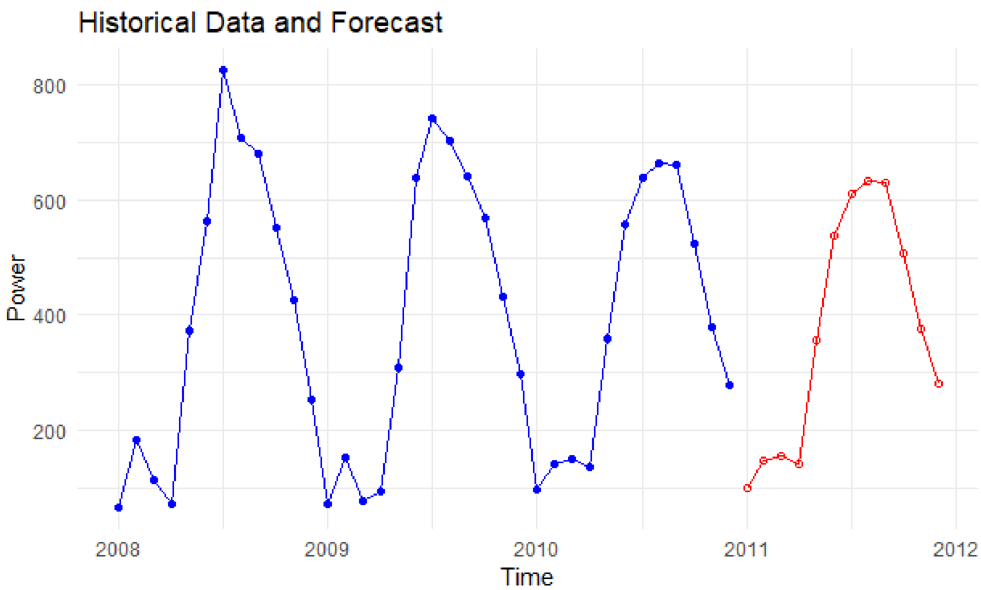

Figure 20: Historical data and forecasts

In Figure 16, we plotted the $\hat{Y}_{t+k|t}$ (red points) extending the observed time series (blue points).

2.3 Calculate 95% prediction intervals for the twelve months ahead and add them to the plot.

We are going to calculate the prediction intervals for our twelve predictions based on the variable X_t and then do the appropriate transformations in order to get the intervals for $\hat{Y}_t$.

We consider only the AR(1) part of our model, which is

$$(1 + \phi_1 B)X_t = \epsilon_t$$

This process can be formulated in MA-form by following this procedure:
We have that

$$X_t = -\phi_1 X_{t-1} + \epsilon_t \tag{*}$$

And from the above we can get the equivalent expression:

$$X_{t-1} = -\phi_1 X_{t-2} + \epsilon_{t-1},$$

which we can substitute in the previous equation using X_{t-1}:

$$X_t = -\phi_1(-\phi_1 X_{t-2} + \epsilon_{t-1}) + \epsilon_t = \phi_1^2 X_{t-2} - \phi_1 \epsilon_{t-1} + \epsilon_t.$$

If we repeat this process infinite times, we arrive at the expression:

$$X_t = \lim_{i \to \infty} \phi_1^i X_{t-i} + \sum_{i=0}^{\infty} (-1)^i \phi_1^i \epsilon_{t-i}.$$

But $\lim_{i \to \infty} \phi_1^i X_{t-i} = 0$ since $\phi_1 < 1$, therefore we arrive at the desired MA form of the linear process:

$$X_t = \sum_{i=0}^{\infty} (-1)^i \phi_1^i \epsilon_{t-i}.$$

From which we can conclude that the psi-weights of said linear process are $\psi_i = (-1)^i \phi_1^i$.

Now we will use said weights to calculate the prediction intervals, which are given by the following expression:

$$\hat{Y}_{t+k|t} \pm u_{\alpha/2} \sigma_\epsilon \sqrt{\sum_{i=0}^{k} \psi_i^2}.$$

Where $u_{\alpha/2}$ is the $\alpha/2$ quantile in the standard normal distribution, approximately equal to 1.96, σ_ϵ, is the square root of the variance of the associated white noise process, ϵ_t and

$\{\psi_i\}$ are the psi-weights.

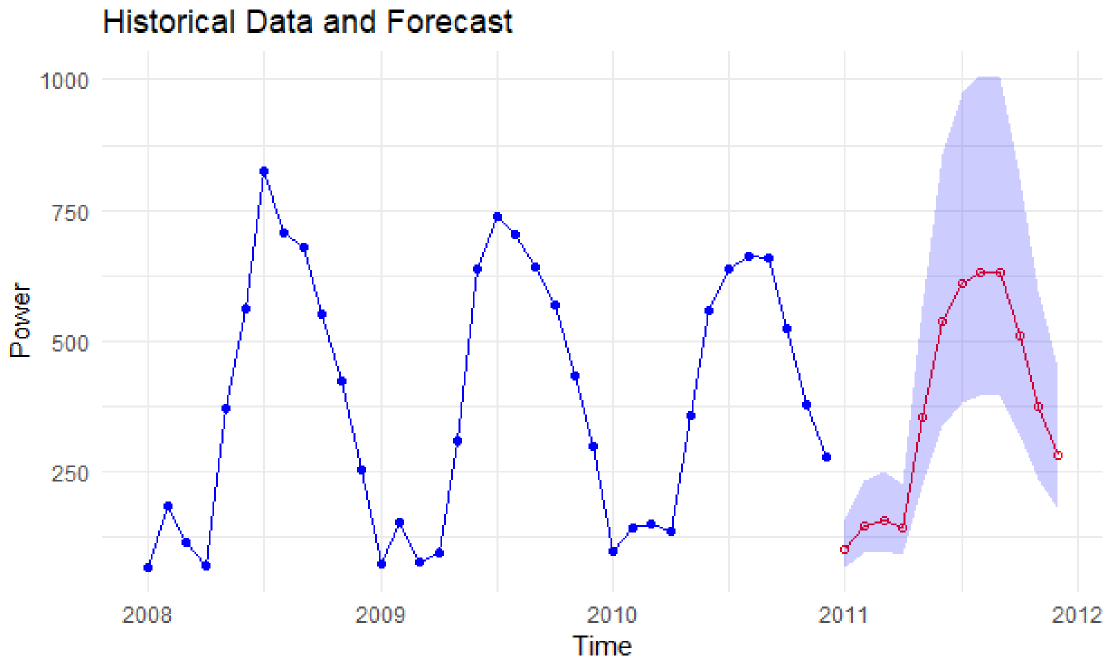

Figure 21: Historical data, forecasts and confidence intervals

In Figure 21, we can observe the prediction intervals in the light blue shade and, in Table 8, we can see their lower and upper bounds.

Month	Lower Prediction Bound	Upper Prediction Bound
1	65.80337	155.8776
2	92.82188	233.5156
3	98.52260	249.9396
4	89.44515	227.1841
5	222.92509	566.3116
6	337.55487	857.5348
7	383.47122	974.1857
8	396.45060	1007.1595
9	395.33178	1004.3173
10	318.81824	809.9391
11	235.28332	597.7235
12	175.97985	447.0665

Table 3: Prediction Intervals

2.4 Would you trust the forecast? Do you think the prediction intervals have correct width (all the time)?

The forecasted values align with previous fluctuations, especially those showing a steep increase, meaning that the forecasted values continue the pattern observed in the historical data and captured the seasonality well. Thus, we could say that we have reliable forecasts.

We notice, though, prediction intervals which are wide, and specifically too wide for the forecasts with high values. This might indicate excessive uncertainty and reduces the usefulness of the predictions. Furthermore, the intervals seem to widen as we go further out, which is expected as uncertainty increases with the forecast horizon.

Therefore, it is not simple to say that we fully trust the forecast and whether the intervals have the "correct" width all the time can't be fully assessed without further statistical tests or actual future values for comparison.

Also, the log transformation complicates the interpretation of prediction intervals on the original scale. While the transformation helps stabilize variance and achieve a better model fit, the back-transformed intervals need to be understood as representing proportional rather than absolute uncertainty. The prediction intervals on the log scale do not translate directly back to the original scale. When we exponentiate the upper and lower bounds of the log-scale prediction intervals, the intervals on the original scale become asymmetrical because the exponential function is non-linear.This might explain the widening of the prediction intervals as the forecast value increases.

3 Simulating Seasonal Processes

The general form of a multiplicative $(p, d, q) \times (P, D, Q)_s$ seasonal model is the following:

$$\phi(B)\Phi(B^s)\nabla^d\nabla_s^D Y_t = \theta(B)\Theta(B^s)\varepsilon_t \tag{1}$$

where ε_t is a white noise process, and $\phi(B)$ and $\theta(B)$ are polynomials of order p and q, respectively. $\Phi(B^s)$ and $\Theta(B^s)$ are polynomials in B^s. We have the six following models and for each model, we want to plot the simulation, as well as the ACF and PACF.

1. The $(1, 0, 0) \times (0, 0, 0)_{12}$ model with $\phi_1 = 0.6$ can be written as:

$$(1 + \phi_1 B)Y_t = \varepsilon_t \iff Y_t + 0.6Y_{t-1} = \varepsilon_t \iff Y_t = -0.6Y_{t-1} + \varepsilon_t \tag{2}$$

This process is not seasonal, because we can see that it does not contain a seasonal part.

To examine if the process is stationary, we should find the roots of the polynomial $\phi(z^{-1})$ and if the roots are within the unit circle then the process is stationary.

Generally, an AR process can be written in the following form:

$$\phi(B)Y_t = \varepsilon_t \tag{3}$$

Specifically, for the given Y_t process, the polynomial $\phi(B)$ is the following:

$$\phi(B) = 1 + 0.6B \tag{4}$$

Therefore the polynomial $\phi(z^{-1})$ is:

$$\phi(z^{-1}) = 1 + 0.6z^{-1} \tag{5}$$

The polynomial that is described by equation (5) has root $z = -0.6$ which is within the unit circle. **Therefore the process $Y_t = -0.6Y_{t-1} + \varepsilon_t$ is stationary.**

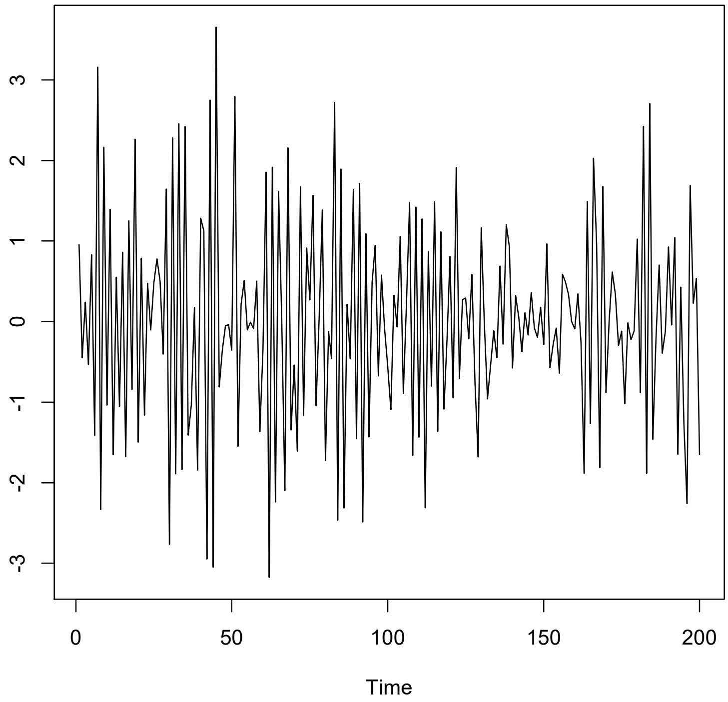

Figure 22: 1 realization with 200 observations of the process: $Y_t = -0.6Y_{t-1} + \varepsilon_t$

In Figure 22 we can observe the simulation of a realization with 200 observations of the process. In Figure 23, the autocorrelation and the partial correlation functions are depicted. Observing the PACF plot in Figure 27, we can conclude that after lag 1, the PACF is truncated (converges to zero), which is expected for an AR process of order 1 that captures the direct effect of $Yt-1$ on Yt after accounting for the intermediate values. The lack of significant spikes at higher lags in the PACF supports the notion that we have an AR(1) process without additional AR terms. On the other hand, observing the ACF plot, we can observe that it shows a gradual decay, which is characteristic of an AR(1) process. The autocorrelations decrease as the lag increases, but they do not cut off sharply after the first lag, which would have been the case for a MA(1) process. This gradual decline in the ACF suggests that the impact of a shock to persists for several periods into the future, although the impact diminishes over time.

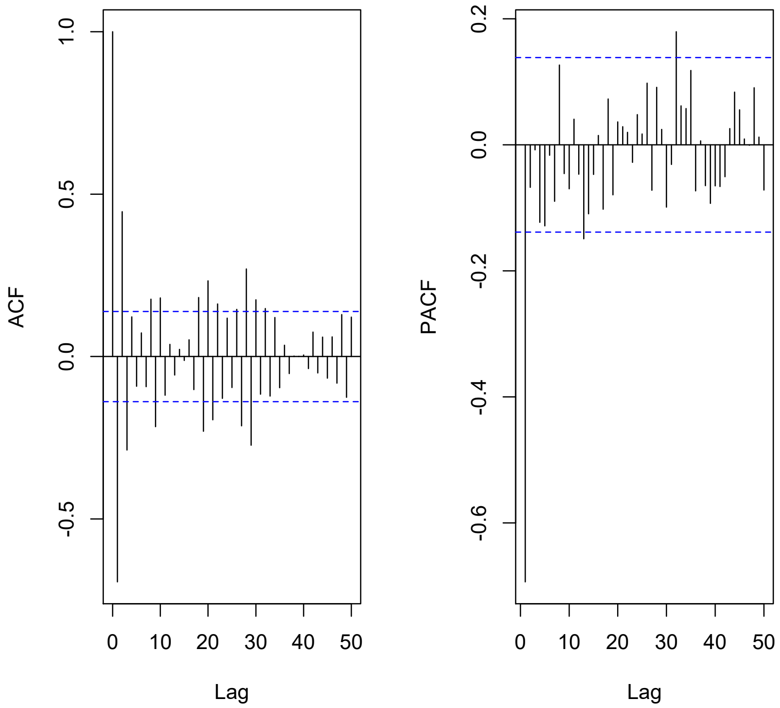

Figure 23: ACF and PCF of the process $Y_t = -0.6Y_{t-1} + \varepsilon_t$

2. The $(0, 0, 0) \times (1, 0, 0)_{12}$ model with $\Phi_1 = -0.9$ can be written as:

$$(1 - \Phi_1 B^{12})Y_t = \varepsilon_t \iff Y_t - 0.9Y_{t-12} = \varepsilon_t \iff Y_t = +0.9Y_{t-12} + \varepsilon_t \tag{6}$$

The process is seasonal, because it includes a seasonal AR part of order 1.
For the seasonal part, the polynomial $\Phi(B)$ is the following:

$$\Phi(B) = 1 - 0.9B^{12} \tag{7}$$

We replace B^{12} with z^{-12} to find the seasonal roots:

$$\Phi(z^{-1}) = 1 - 0.9z^{-12} = 1 - 0.9z^{-12} \tag{8}$$

The polynomial that is described by the equation (32) has roots

$$
\begin{aligned}
z = 0.4956292 \pm 0.8584549i, -0.8584549 \pm 0.4956292i,& \\
-0.4956292 \pm 0.8584549i, 0.8584549 \pm 0.4956292i,& \\
0.0000000 \pm 0.9912584i, -0.9912584 \pm 0.0000000i,& \\
0.9912584 \pm 0.0000000i& \quad (9)
\end{aligned}
$$

Therefore the process is stationary since all the roots are in the unit circle.

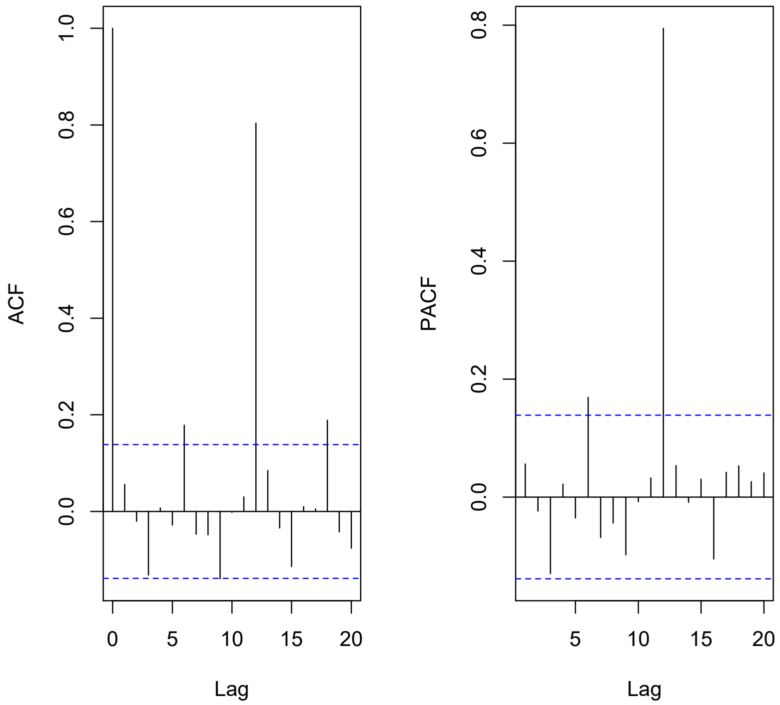

Figure 24: ACF and PCF of the process $Y_t = +0.9Y_{t-12} + \varepsilon_t$

Observing the PACF of Figure 24, we can see that the PACF has a significant peak at the 12th lag, where 12 is the span of the season which is in line with the model specification, where the value suggests that the past value at the same season (12 periods ago) has a strong influence on the current value. It also has a smaller peak at lag 6 and after that, it goes to zero and at the same time it does not show a significant spike at lag 1, which would typically be expected for an AR(1) process. On the other hand, ACF decays exponentially as a damped sine with peaks at lags 0,6,12,18... which is expected due to seasonal autoregressive term at lag12 of the model.

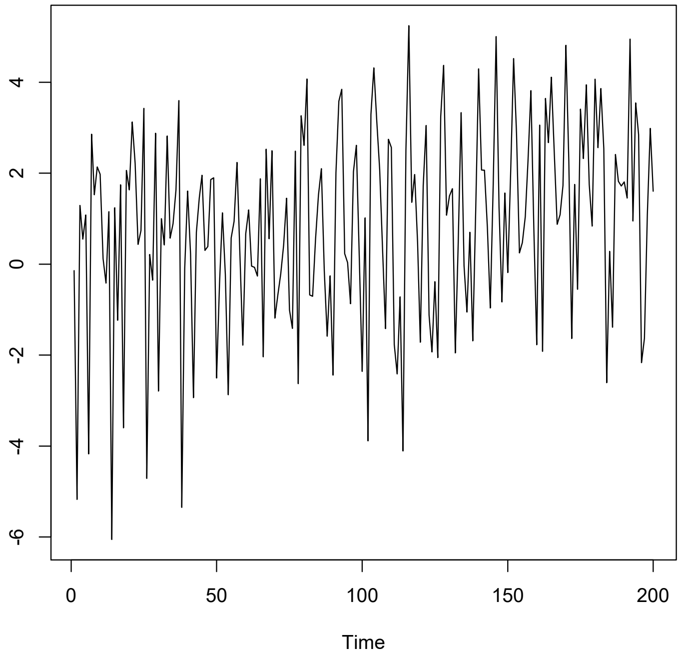

Figure 25: 1 realization with 200 observations of the process $Y_t = +0.9Y_{t-12} + \varepsilon_t$

Looking at Figure 25, the time series appears to show an upwards trend but small to be sure about stationarity from the plot, meaning that its statistical properties, such as mean and variance, might not be constant over time. The series displays fluctuations around a mean of zero, with these oscillations being quite regular due to the influence of the seasonal lag. The amplitude of the fluctuations is not constant but seems to be fairly consistent, without showing any obvious increasing or decreasing trend.

3. The $(1, 0, 0) \times (0, 0, 1)_{12}$ model with $\phi_1 = 0.9$ and $\Theta_1 = -0.7$ can be written as:

$$(1 - \phi_1 B)Y_t = (1 + \Theta_1 B^{12})\varepsilon_t \iff Y_t + 0.9Y_{t-1} = \varepsilon_t - 0.7\varepsilon_{t-12} \tag{10}$$

$$Y_t = -0.9Y_{t-1} - 0.7\varepsilon_{t-12} + \varepsilon_t \tag{11}$$

The process is seasonal because it includes a seasonal MA part of order 1. Additionally, there is a non-seasonal AR component.

To examine if the process is stationary, we should find the roots of the polynomial$\phi(z^{-1})$ and if all the roots are within the unit circle then the process is stationary.

For the given Yt process, the polynomial $\phi(B)$ is the following:

$$\phi(B) = 1 + 0.9B \tag{12}$$

Therefore, the polynomial $\phi(z^{-1})$ is:

$$\phi(z^{-1}) = 1 + 0.9z^{-1} \tag{13}$$

The polynomial $\phi(z^{-1})$ that is described by the equation (36) has root z = 0.9, which is within the unit circle. Therefore, the process is stationary.

Figure 27 displays one realization with 200 observations of the process $Y_t = -0.9Y_{t-1} + \varepsilon_t - 0.7\varepsilon_{t-12}$ and the Figure 28 depicts the ACF and PACF of the process.

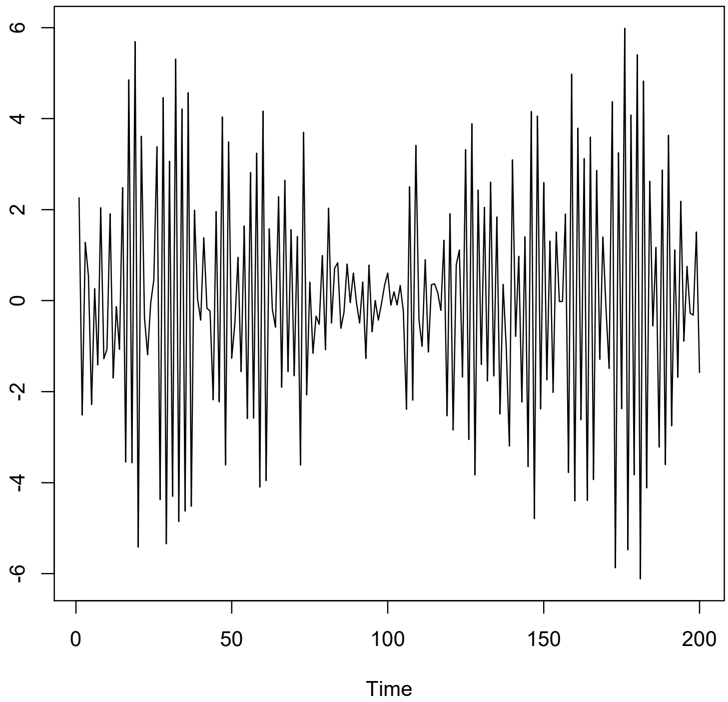

Figure 26: 1 realization with 200 observations of the process $Y_t = -0.9Y_{t-1} + \varepsilon_t - 0.7\varepsilon_{t-12}$

The plot itself displays a series that oscillates and appears to have no clear trend over time. This fluctuation with a consistent range implies that the series is stationary. However, the presence of spikes at regular intervals could suggest the seasonal impact of the moving average term, which occurs every 12 observations.

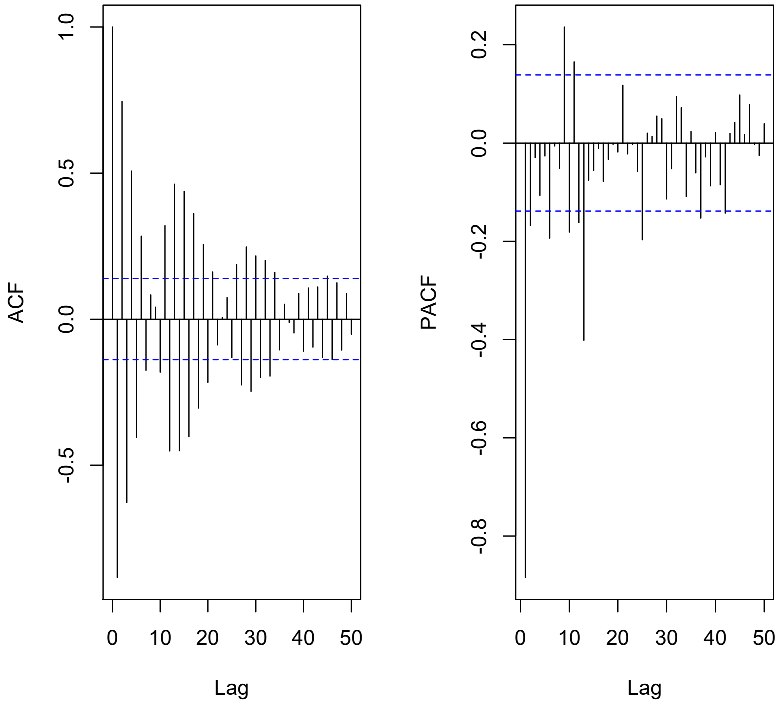

Figure 27: ACF and PACF of the process

The presence of both non-seasonal AR and seasonal MA components in a time series can result in complex autocorrelation structures, as the effects of these components may overlap or mask each other. Observing Figure 27, we can see that PACF presents an exponential decay and exhibits a significant spike at lag 1, which is indicative of an AR(1) process. There is also a significant spike at lag 12 and a smaller one at lag 24, which is evidence of a seasonal effect. ACF showcases a significant autocorrelation at lag 1 and follows an alternating decaying seasonal pattern that fades out after lag 36.

4. The $(1,0,0) \times (1,0,0)_{12}$ model with $\phi_1 = -0.6$ and $\Phi_1 = -0.8$ can be written as:

$$(1 + \phi_1 B)(1 + \Phi_1 B^{12})Y_t = \varepsilon_t \iff (1 - 0.6B)(1 - 0.8B^{12})Y_t = \varepsilon_t \tag{14}$$

$$\iff Y_t = +0.6Y_{t-1} + 0.8Y_{t-12} - 0.48Y_{t-13} + \varepsilon_t \tag{15}$$

The process has a non-seasonal and a seasonal AR part of order 1. Specifically for the given Y_t process, the polynomial $\phi(B)$ is the following:

$$\phi(B) = 1 - 0.6B - 0.8B^{12} + 0.48B^{13} \tag{16}$$

Therefore, the polynomial $\phi(z^{-1})$ is:

$$\phi(z^{-1}) = 1 - 0.6z^{-1} - 0.8z^{-12} + 0.48z^{-13} \tag{17}$$

The polynomial that is described by the equation (39) has roots:

$$\begin{aligned} z = &\, 0.6972238 \pm 0.6818343i, -0.7825597 \pm 0.4514276i, \\ &- 0.4695520 \pm 0.8059856i, 0.9089295 \pm 0.2420431i, \\ &0.3643062 \pm 0.9903491i, -0.8950804 \pm 0i, -0.0308076 \pm 1.003263i \end{aligned} \tag{18}$$

that not all of them lie within the unit circle. Therefore, the process $Y_t = +0.6Y_{t-1} + 0.8Y_{t-12} - 0.48Y_{t-13} + \varepsilon_t$ is not stationary.

In Figure 28 we can see one realization with 200 observations of the process Y_t and we can easily conclude that the process is stationary. In Figure 27, the ACF and PACF of the specific process are displayed.

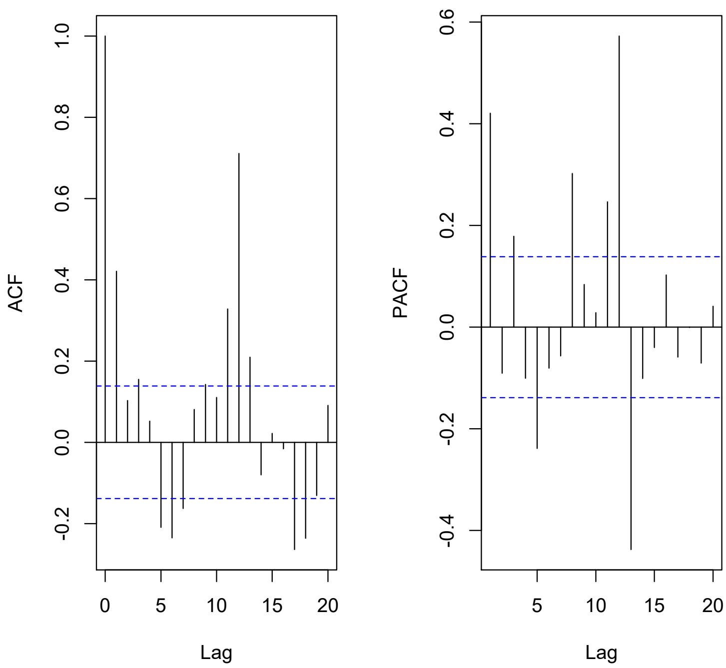

Figure 28: ACF and PACF of the process $Y_t = +0.6Y_{t-1} + 0.8Y_{t-12} - 0.48Y_{t-13} + \varepsilon_t$

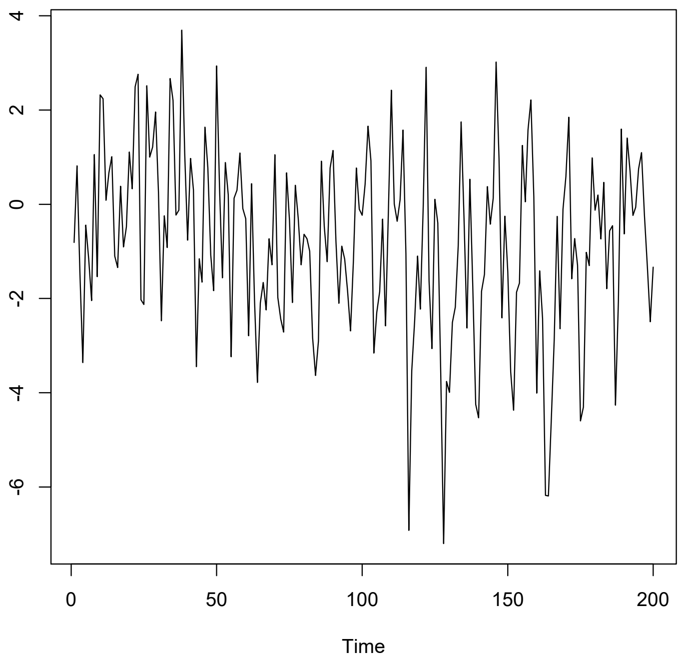

Figure 29: 1 realisation with 200 observations of the process $Y_t = +0.6Y_{t-1} + 0.8Y_{t-12} - 0.48Y_{t-13} + \varepsilon_t$

Looking at the PACF plot (Figure 28), we can observe peaks at the lags 1,3,5,8,12 and 13 and after that, the PACF cuts off. The significant spike at lag 1 confirms the suspected AR(1) component and the negative spike at lag 12 in the PACF plot also supports the presence of a seasonal AR component. The significant negative correlation at this lag suggests that the current value is inversely related to the value from the same period one year ago. The ACF function is damped sine that has some spikes firstly in lags 0,1 then to lags 5,6,7 then to lags 10,11,12 then to lag 17,18 and after that converges to zero. This pattern suggests the presence of an AR(1) component, which in this case is negative, and typically results in a gradual decay in the ACF rather than a sharp cut-off.

5. The $(0,0,1) \times (0,0,1)_{12}$ model with $\theta_1 = 0.4$ and $\Theta_1 = -0.8$ (note the sign change as per your request, but the original quote had $\Theta_1 = -0.8$, adjust accordingly if needed) can be written as:

$$Y_t = (1 + \theta_1 B)(1 + \Theta_1 B^{12})\varepsilon_t \iff Y_t = (1 + 0.4B)(1 - 0.8B^{12})\varepsilon_t \tag{19}$$

$$\iff Y_t = \varepsilon_t + 0.4\varepsilon_{t-1} - 0.8\varepsilon_{t-12} - 0.32\varepsilon_{t-13} \tag{20}$$

The process has a non-seasonal and a seasonal MA component of order 1. It is also stationary by definition.

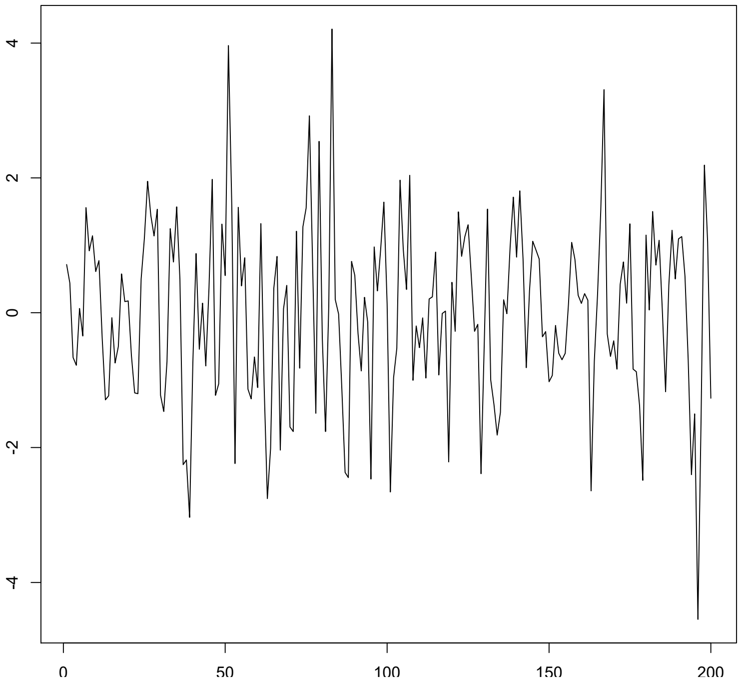

Figure 30: Realization for $Y_t = \varepsilon_t + 0.4\varepsilon_{t-1} - 0.8\varepsilon_{t-12} - 0.32\varepsilon_{t-13}$

Observing the PACF function of Figure 31, we can see dominant peaks at the lags 1,12 and minor peaks at lags 4 and 11 and then PACF is truncated. At the same time, ACF is a damped sine that cuts off after lag 13 with peaks on lag 0,1,5,12 and 13.

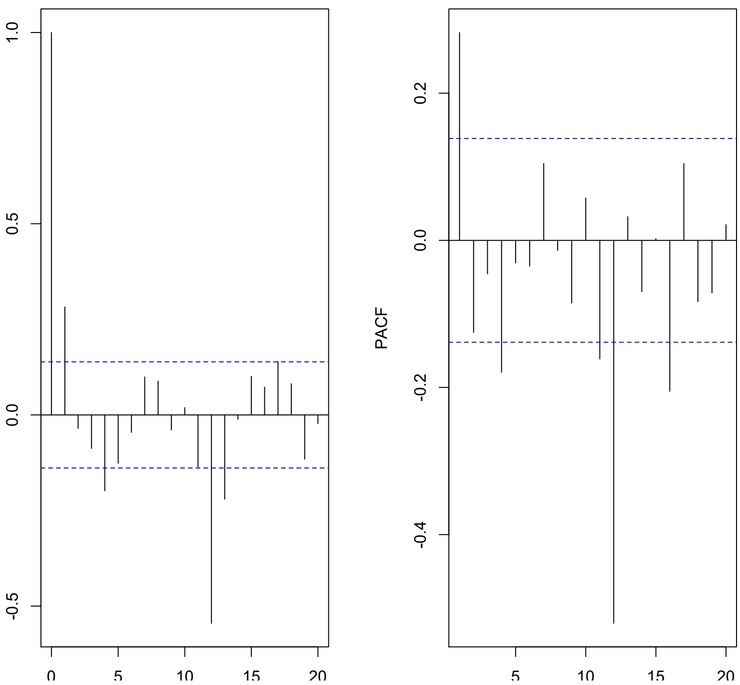

Figure 31: ACF and PACF of the process Yt

6. The $(0,0,1) \times (1,0,0)_{12}$ model with $\theta_1 = -0.4$ and $\Phi_1 = 0.7$ can be written as:

$$(1 + \Phi_1 B^{12})Y_t = (1 + \theta_1 B)\varepsilon_t \iff (1 + 0.7B^{12})Y_t = (1 - 0.4B)\varepsilon_t \tag{21}$$

$$\iff Y_t = -0.7\varepsilon_{t-12} - 0.4\varepsilon_{t-1} + \varepsilon_t \tag{22}$$

This model is an example of a mixed model combining non-seasonal moving average (MA) and seasonal autoregressive (AR) components. The inclusion of a seasonal AR component of order 1 signifies that the model attempts to capture seasonal patterns in the data, while the non-seasonal MA component aims to model the short-term correlations.

The polynomial that is described by the equation (24) has roots

$$z = 0.5150845 \pm 0.8921526i, -0.8921526 \pm 0.5150845i, -0.5150845 \pm 0.8921526i,$$
$$0.8921526 \pm 0.5150845i, 0.0 \pm 1.0301690i, -1.0301690 \pm 0.0i,$$
$$0.0 \pm 1.0301690i, 1.0301690 \pm 0.0i \tag{23}$$

Therefore the process is not stationary since not all the roots are in the unit circle.

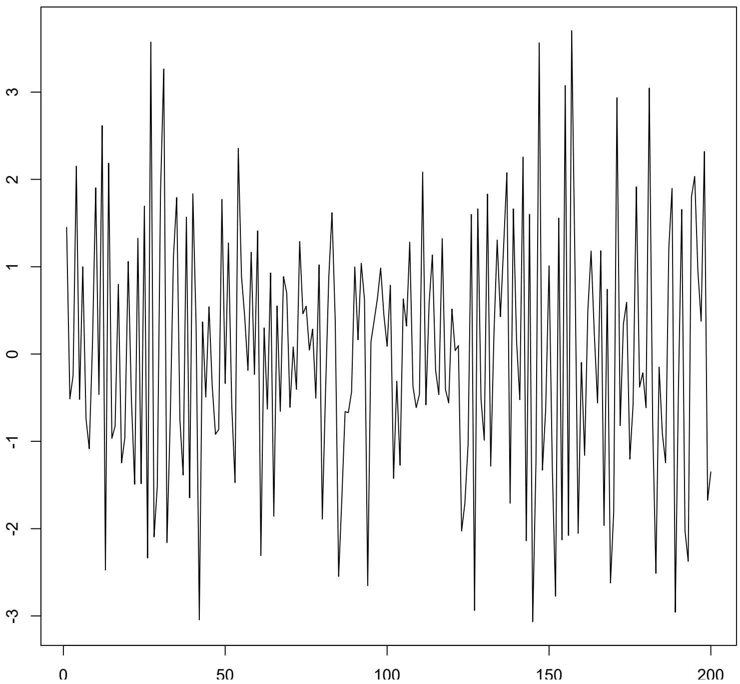

Figure 32: Realization for $Y_t = -0.7\varepsilon_{t-12} - 0.4\varepsilon_{t-1} + \varepsilon_t$

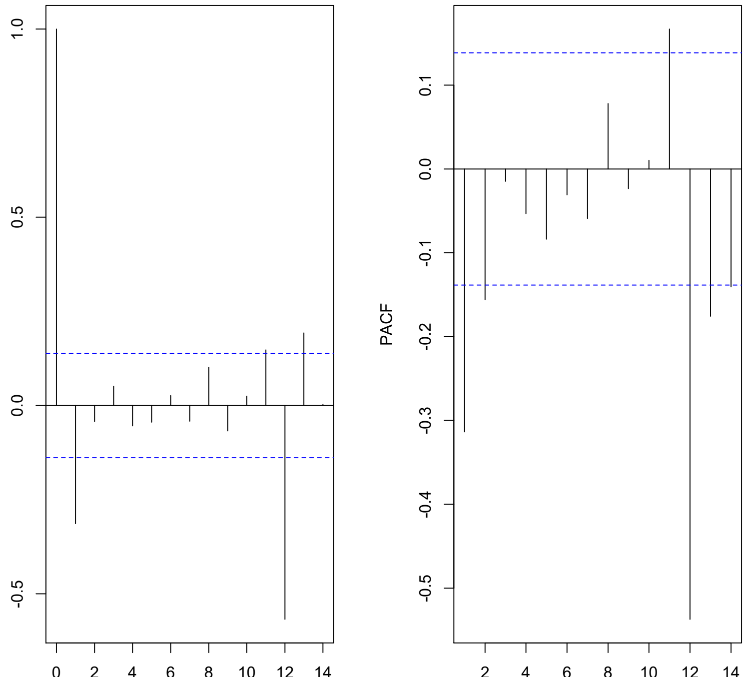

Figure 33: ACF and PACF of the process Yt

7. Each of the processes has been commented on in its corresponding section of the exercise. Additionally, some general conclusions can be made after exploring the processes above.

   It can be noted that a non-seasonal AR of order p, presents spikes on the ACF until lag p and then it cuts off and will often show a more gradual decay. In models with seasonal components, the ACF will typically exhibit significant spikes at multiples of the seasonal lag. For example, in a model with a seasonal period of 12, you might see spikes at lags 12, 24, 36, and so on. The PACF for a seasonal AR process will typically show a sharp cutoff after the seasonal lag. For instance, in a model with a seasonal AR component at lag 12, the PACF will show significance at lag 12 and cut off after that. For a seasonal MA process, the ACF will present spikes but will typically rapidly cut off after a few lags while the PACF will demonstrate a sharp cutoff after the order of the seasonal component similar to the case of non-seasonal processes.

   All in all, ACF and PACF plots need to be examined for significant values that stand out beyond the confidence bounds. These indicate correlations that are statistically significant and suggest potential terms for the ARIMA model. The behavior of both the ACF and PACF together helps in model identification and for pure seasonal AR models, the PACF will generally show a slow decay, while for pure seasonal MA models, the ACF will exhibit this pattern.

4 Identifying ARMA Models

4.1 Guess the ARMA model structure, and give a short reasoning of your guess

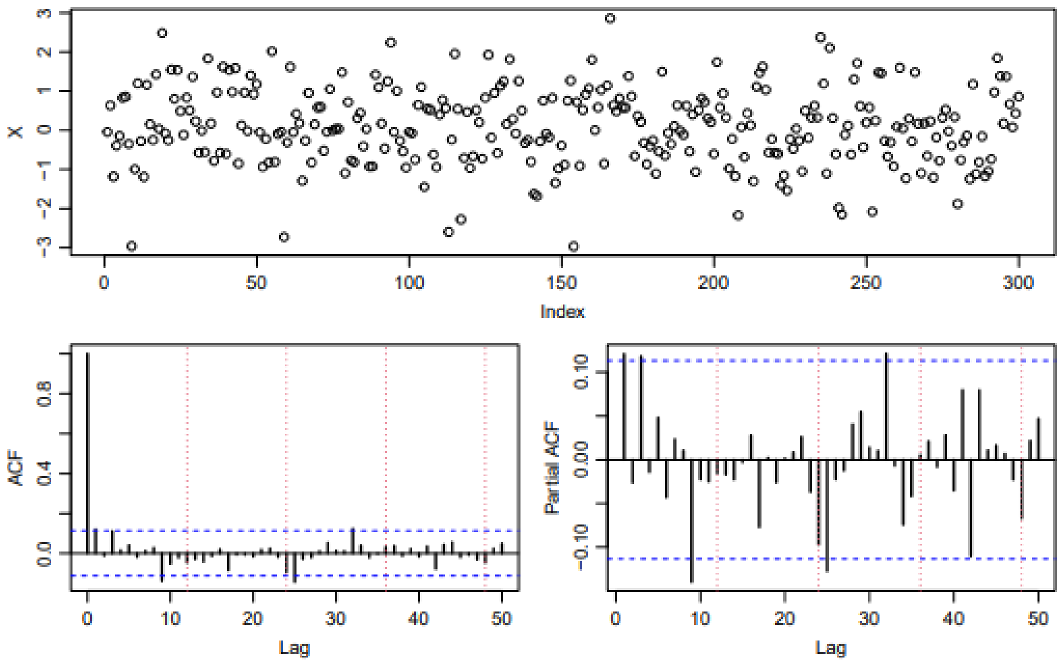

Figure 34

Time Series Plot: The scatter plot of the time series does not show any clear trend or seasonality. The data points seem to be randomly scattered around a constant level.

ACF Plot: The ACF shows significant autocorrelation at lag 0 as expected and then it quickly drops within the bounds of significance. The autocorrelations remain within the confidence interval for all other lags, suggesting that there is no strong autocorrelation in the data at lags greater than zero.

PACF Plot: The PACF shows a few spikes that are significant at the early lags, specifically at lag 1, and then several others beyond that, but it is not clear cut. These could suggest an AR component; however, the significance of these spikes is sporadic and does not show a clear pattern.

The conclusion is that the linear process depicted in this case is such that there are no AR or MA componentes, which implies that it is just white noise i.e:

$$Y_t = \epsilon_t \tag{24}$$

4.2 Guess the ARMA model structure, and give a short reasoning of your guess

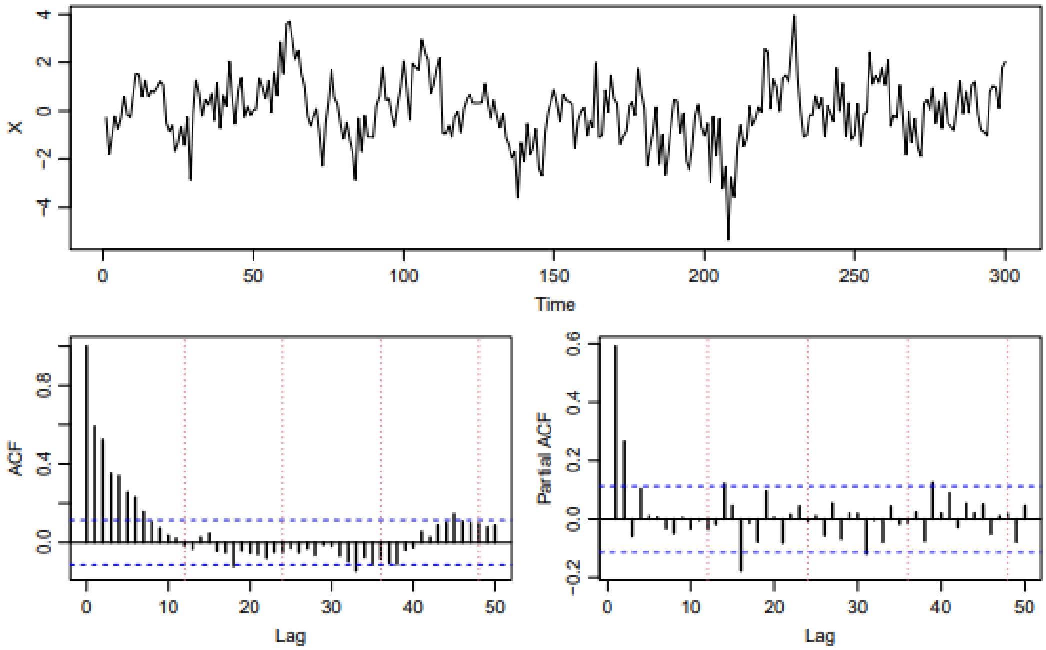

Figure 35

Time Series Plot: The time series data does not exhibit a clear trend or seasonal effect.

ACF Plot: The ACF shows a exponential decay starting at lag one which may indicate an AR or ARMA structure.

PACF Plot: Shows a high correlation at lag 1 that immediately after falls below the confidence interval line which suggests an AR(1) model, as if it was an ARMA the PACF plot would show more spikes that grow smaller as the lag increases.

With the information presented we can conclude that this linear process is indeed AR(1) since the spike in the PACF plot occurs at lag 1. So the linear process would be modeled by:

$$(1 - \phi_1 B)Y_t = \epsilon_t \tag{25}$$

4.3 Guess the ARMA model structure, and give a short reasoning of your guess

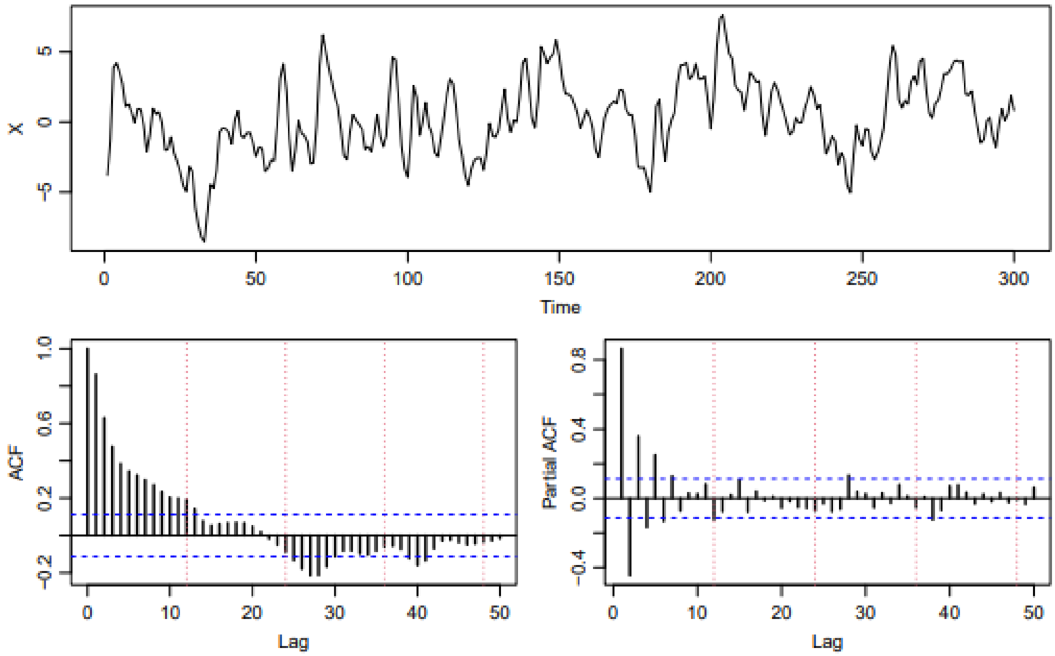

Figure 36: Linear model 3

Time Series Plot: The plot does not exhibit any clear or consistent trend, seasonality, or long-term cyclical structure.
ACF Plot: It gradually decays into, which indicates the presence of an AR component.

PACF Plot: We can see another damped exponential possibly combined with a sine function instead of an abrupt end, which means that there is an MA component. In addition, the is only high correlation at lag 1, meaning that $p = 1$.

Combining these insights, the plots suggest an ARMA model where the AR part is likely of order 1. In this case we can use the fact that the exponential decay in the ACF plot starts at 1, so $1 = p + 1 - q \implies p = q$, so we can conclude that the model is ARMA(1,1).

List of Figures

1 Theoritical Autocorrelation Function AR(2) 2
2 Realizations of AR(2) Process 3
3 ACF for each realization of the process Xt 4
4 Realizations of AR(2) Process 5
5 Theoretical ACF 5
6 ACF for each realization of the process Xt 6
7 Realizations of AR(2) Process 7
8 Theoretical ACF 7
9 ACF for each realization of the process Xt 8
10 Realizations of AR(2) Process 9
11 Theoretical ACF 9
12 ACF for each realization of the process Xt 10
13 Realizations of AR(2) Process 11
14 Theoretical ACf 11
15 Autocorrelation $\gamma(k)$ up to nlag=30 12
16 Our Time Series 14
17 Residuals 17
18 QQplot 17
19 Autocorrelation in residuals 18
20 Historical data and forecasts 19
21 Historical data, forecasts and confidence intervals 21
22 1 realization with 200 observations of the process: $Y_t = -0.6Y_{t-1} + \varepsilon_t$ 24
23 ACF and PCF of the process $Y_t = -0.6Y_{t-1} + \varepsilon_t$ 25
24 ACF and PCF of the process $Y_t = +0.9Y_{t-12} + \varepsilon_t$ 26
25 1 realization with 200 observations of the process $Y_t = +0.9Y_{t-12} + \varepsilon_t$ 27
26 1 realization with 200 observations of the process $Y_t = -0.9Y_{t-1} + \varepsilon_t - 0.7\varepsilon_{t-12}$ 28
27 ACF and PACF of the process 29
28 ACF and PACF of the process $Y_t = +0.6Y_{t-1} + 0.8Y_{t-12} - 0.48Y_{t-13} + \varepsilon_t$ 30
29 1 realisation with 200 observations of the process $Y_t = +0.6Y_{t-1} + 0.8Y_{t-12} - 0.48Y_{t-13} + \varepsilon_t$ 31
30 Realization for $Y_t = \varepsilon_t + 0.4\varepsilon_{t-1} - 0.8\varepsilon_{t-12} - 0.32\varepsilon_{t-13}$ 32
31 ACF and PACF of the process Yt 32
32 Realization for $Y_t = -0.7\varepsilon_{t-12} - 0.4\varepsilon_{t-1} + \varepsilon_t$ 33
33 ACF and PACF of the process Yt 34
34 35
35 36
36 Linear model 3 37

List of Tables

1 Predicted values and residuals for our time series. 16
2 Forecasted values of $\hat{Y}_{36+k|36}$. 19
3 Prediction Intervals 21