

Time Series Analysis: Homework #2

Due on April 20, 2021

Professor Tianwei Yu

Peng Deng

Problem 1

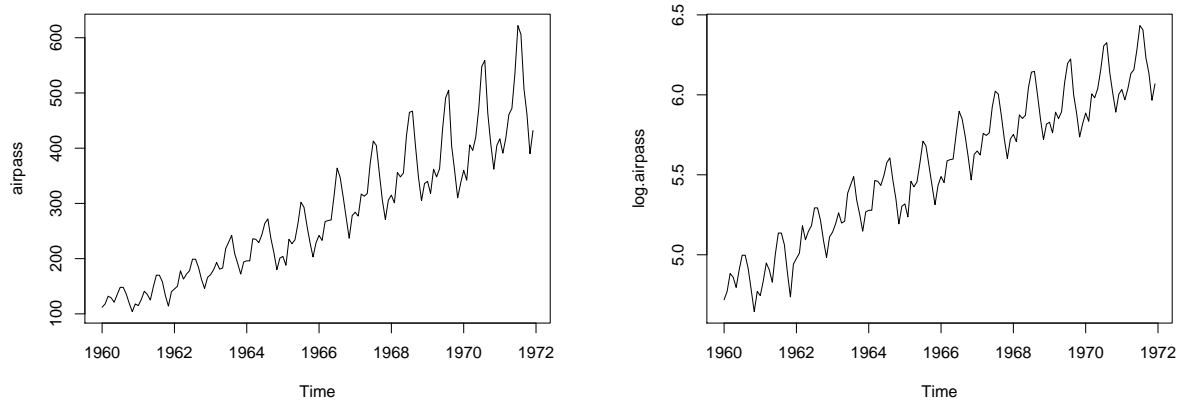
The monthly airline passenger time series, first investigated in Box and Jenkins (1976), is considered a classic time series. The data is in the R library "TSA", named "airpass".

- (a) Display the time series plots of both the original series and the logarithms of the series. Argue that taking logs is an appropriate transformation.
- (b) Display and interpret the time series plots of the first difference of the logged series.
- (c) Display and interpret the time series plot of the seasonal difference of the first difference of the logged series.
- (d) Calculate and interpret the sample ACF of the seasonal difference of the first difference of the logged series.
- (e) Fit the "airline model" (ARIMA(0,1,1)(0,1,1)₁₂) to the logged series.
- (f) Investigate diagnostics for this model, including autocorrelation and normality of the residuals.
- (g) Produce forecasts for this series with a lead time of two years. Note: the TSA package also contains an "arima" command. However it doesn't work well with the "forecast" command we used in the class. Make sure if you want to use "forecast", use the arima in the stats package by specifying "stats::arima(...)".

Solution

Subproblem (a)

Firstly, we can plot the original series and the logarithms of the series as Figure 1. As we can see, in the original series, the seasonal effect increases as the trend increases. However, after taking logarithms, the seasonal effect tends to be the same as the trend increases. Thus, the logarithms will make the series more "linearizable", which can help the following analysis.



(a) The original series

(b) The logarithms of the series

Figure 1 The plot of the time series

Subproblem (b)

As we can see from Figure 2, after taking the first difference of the logged series, we can remove the linear trend of the logged series.

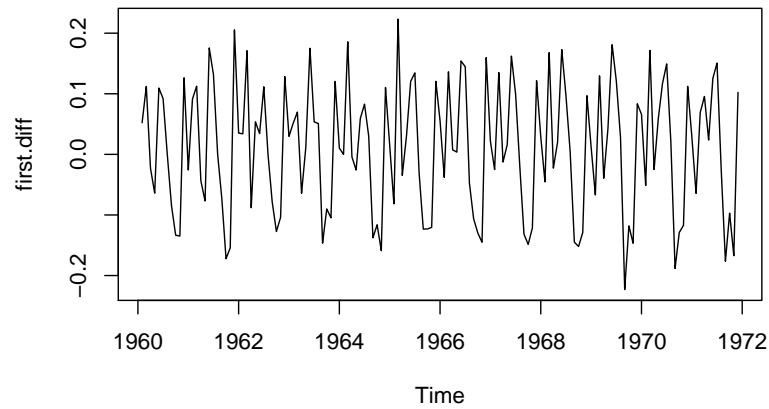


Figure 2 The first difference of the logged series

Subproblem (c)

As we can see from Figure 3, the seasonal difference of the first difference of the logged series seems not stationary, because the variance seems to vary as the trend increases.

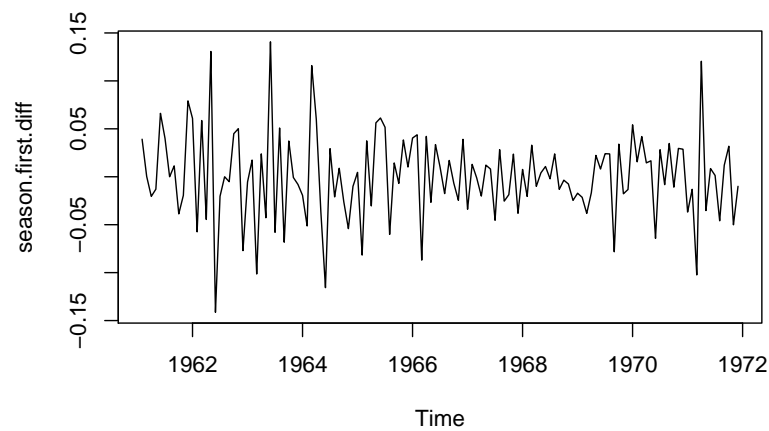


Figure 3 The seasonal difference of the first difference of the logged series

Subproblem (d)

As we can see from Figure 4, there is evident acf of the seasonal difference of the first difference of the logged series.

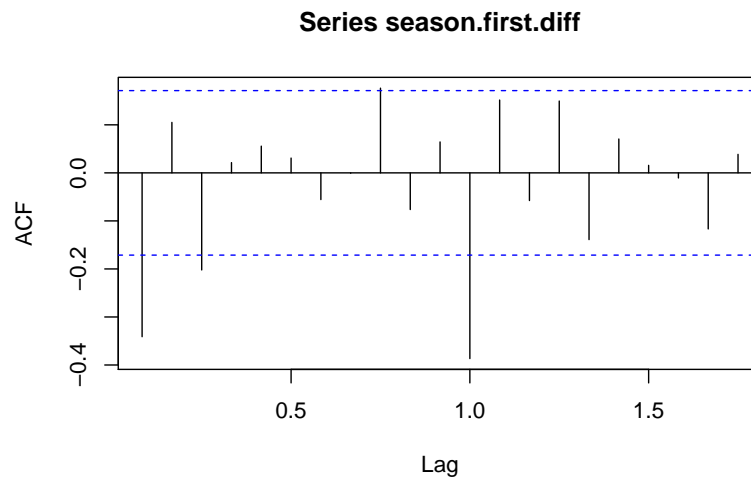


Figure 4 The ACF of the seasonal difference of the first difference of the logged series

Subproblem (e)

We use the following code to fit $(ARIMA(0,1,1)(0,1,1)_{12})$ to the logged series.

```
log.airpass.arima = stats::arima(log.airpass, order=c(0,1,1),
                                seasonal=list(order=c(0,1,1), period=12))
```

The fitted curve is showed as Figure 5. In this figure, the black line is the logged series, and the red dashed line is the fitted value by $(ARIMA(0,1,1)(0,1,1)_{12})$.

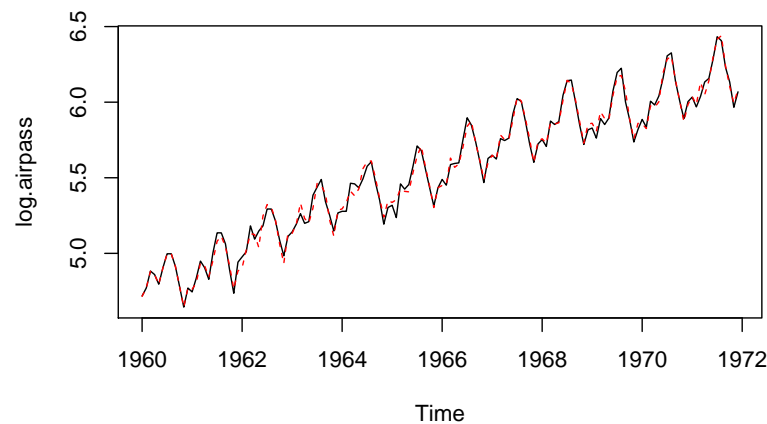


Figure 5 The fitted figure of Seasonal ARIMA model

Subproblem (f)

As we can see from Figure 6, there is significant acf, so that the model is not very good. However, we can do the Ljung-Box test of the residuals to see if the residuals exhibit serial correlation, the result is as follow:

```
Box-Ljung test
data: log.airpass.arima$residuals
X-squared = 8.8097, df = 9, p-value = 0.455
```

As we can see from the Ljung-Box test, the p-value equals 0.455, which is larger than 0.05, so that we should not reject the null hypothesis, which means the residuals does not exhibit serial correlation. Thus, we can believe that the Seasonal ARIMA model is suitable.

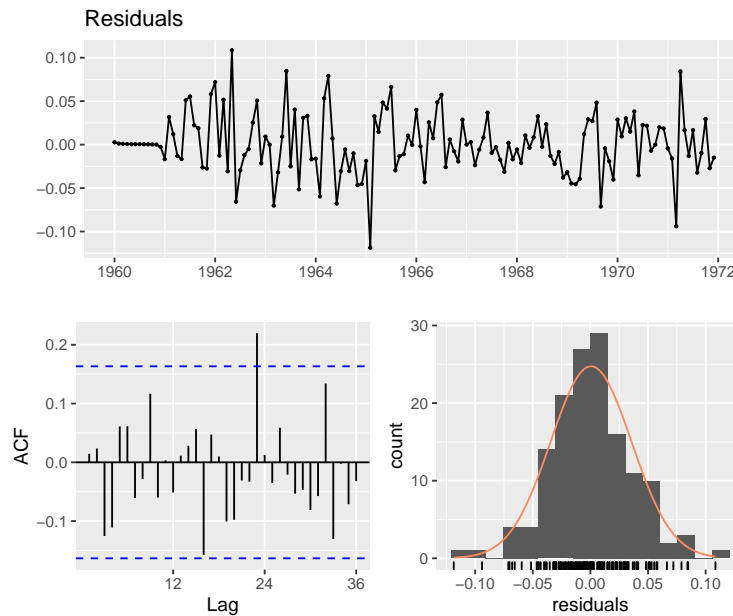


Figure 6 Check the residuals of Seasonal ARIMA

Then, we would like to check the normality of the residuals. We use Shapiro-Wilk normality test to test the normality. The result is as follow.

Shapiro-Wilk normality test

```
data: log.airpass.arma$residuals
W = 0.98637, p-value = 0.1674
```

As we can see from the result of Shapiro-Wilk normality test, p-value equals 0.1674, which is larger than 0.05, so we can conclude that the residuals is subject to normal distribution.

Subproblem (g)

The forecast for this series with a lead time of two years is as Figure 7.

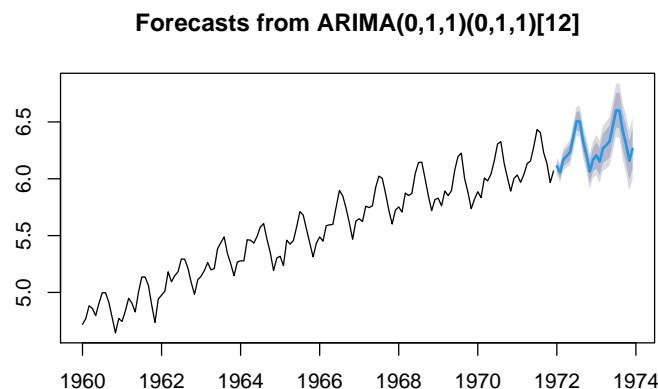


Figure 7 The forecast with Seasonal ARIMA model in two years

Problem 2

Consider the model $Y_t = e_{t-1} - e_{t-2} + 0.5e_{t-3}$.

- (a) Find the autocovariance function for this process.
 (b) Argue this is just an MA process in disguise. What is that process?

Solution

Subproblem (a)

We suppose $e_t \sim IWN(0, \sigma^2)$. Then, we can derive the autocovariance function as follow

$$\begin{aligned} \gamma_k &= \text{Cov}(Y_t, Y_{t+k}) \\ &= \text{Cov}(e_{t-1} - e_{t-2} + 0.5e_{t-3}, e_{t+k-1} - e_{t+k-2} + 0.5e_{t+k-3}) \\ \Rightarrow \gamma_k &= \begin{cases} 2.25\sigma^2 & (k=0) \\ -1.5\sigma^2 & (k=1) \\ 0.5\sigma^2 & (k=2) \\ 0 & (k>2) \end{cases} \end{aligned} \quad (1)$$

Subproblem (b)

Actually, this is just an $MA(2)$ process in disguise. This is because the error term e_t is a random variable and actually the same thing in different time. Thus, we can treat the original series as follow

$$Y_t = e_t - e_{t-1} + 0.5e_{t-2} \quad (2)$$

Thus, this is a $MA(2)$ process with parameters $\phi_2(\mathbf{B}) = 1 - \mathbf{B} + 0.5\mathbf{B}^2$.

Problem 3

Identify as specific $ARIMA$ models, that is, what are p, d , and q and what are the values of the parameters—the ϕ 's and θ 's? Note: For the model to be an $ARIMA$ model, the $ARMA$ part needs to be stationary. You may need to determine how many differences to take.

- (a) $Y_t = Y_{t-1} - 0.25Y_{t-2} + e_t - 0.1e_{t-1}$
 (b) $Y_t = 2Y_{t-1} - Y_{t-2} + e_t$
 (c) $Y_t = 0.5Y_{t-1} - 0.5Y_{t-2} + e_t - 0.5e_{t-1} + 0.25e_{t-2}$

Solution

Subproblem (a)

The characteristic equation of the AR process is

$$\theta_2(\mathbf{B}) = 1 - \mathbf{B} + 0.25\mathbf{B}^2 = 0 \quad (3)$$

Thus, we can derive the roots are $\mathbf{B}_1 = \mathbf{B}_2 = 2$. Thus, the process is stationary, so we can derive p, d, q and the parameters of the ϕ 's and θ 's as follow.

$$\begin{aligned} p &= 2 \\ q &= 1 \\ d &= 0 \\ \theta_2(\mathbf{B}) &= 1 - \mathbf{B} + 0.25\mathbf{B}^2 \\ \phi_1(\mathbf{B}) &= 1 - 0.1\mathbf{B} \end{aligned} \quad (4)$$

Subproblem (b)

The characteristic equation of the AR process is

$$\theta_2(\mathbf{B}) = 1 - 2\mathbf{B} + \mathbf{B}^2 = 0 \quad (5)$$

Thus, we can derive the roots are $\mathbf{B}_1 = \mathbf{B}_2 = 1$. Thus, the process is not stationary. Then can describe the process as follow

$$\begin{aligned} (1 - 2\mathbf{B} + \mathbf{B}^2) Y_t &= e_t \\ \implies (1 - \mathbf{B})^2 Y_t &= e_t \end{aligned} \quad (6)$$

Thus, the second order difference of Y_t is stationary, so we can derive p, d, q and the parameters of the ϕ 's and θ 's as follow.

$$\begin{aligned} p &= 0 \\ q &= 0 \\ d &= 2 \\ \theta_0(\mathbf{B}) &= 1 \\ \phi_0(\mathbf{B}) &= 1 \end{aligned} \quad (7)$$

Subproblem (c)

The characteristic equation of the AR process is

$$\theta_2(\mathbf{B}) = 1 - 0.5\mathbf{B} + 0.5\mathbf{B}^2 = 0 \quad (8)$$

Thus, we can derive the roots are $\mathbf{B}_1 = \frac{1}{2} + \frac{\sqrt{7}}{2}i, \mathbf{B}_2 = \frac{1}{2} - \frac{\sqrt{7}}{2}i$. We know the absolute value of \mathbf{B}_1 and \mathbf{B}_2 are $\sqrt{2}$, which is larger than 1. Thus, the process is stationary, so we can derive p, d, q and the parameters of the ϕ 's and θ 's as follow.

$$\begin{aligned} p &= 2 \\ q &= 2 \\ d &= 0 \\ \theta_2(\mathbf{B}) &= 1 - 0.5\mathbf{B} + 0.5\mathbf{B}^2 \\ \phi_2(\mathbf{B}) &= 1 - 0.5\mathbf{B} + 0.25\mathbf{B}^2 \end{aligned} \quad (9)$$

Problem 4

Describe the important characteristics of the autocorrelation function for the following models: (a) $MA(1)$, (b) $MA(2)$, (c) $AR(1)$, and (d) $ARMA(1, 1)$. Simulate a series for each of them with $T = 10000$, and show the ACF plots.

Solution**Subproblem (a)**

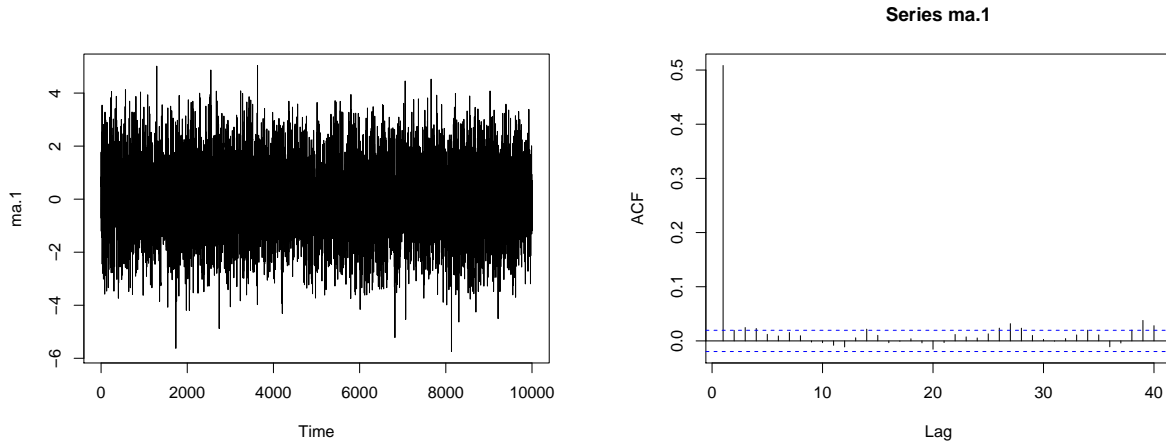
For the $MA(1)$ process, we have

$$Y_t = e_t + \beta e_{t-1} \quad (10)$$

Then, we can derive the autocorrelation function as follow

$$\rho_k = \begin{cases} 1 & (k = 0) \\ \frac{\beta}{1+\beta^2} & (k = 1) \\ 0 & (k > 1) \end{cases} \quad (11)$$

Then, we simulate $MA(1)$ with $\beta = 1$, and show the ACF plot as Figure 8.



(a) The simulated $MA(1)$ series

(b) The acf of simulated $MA(1)$ series

Figure 8 The plot of simulated $MA(1)$ process

Subproblem (b)

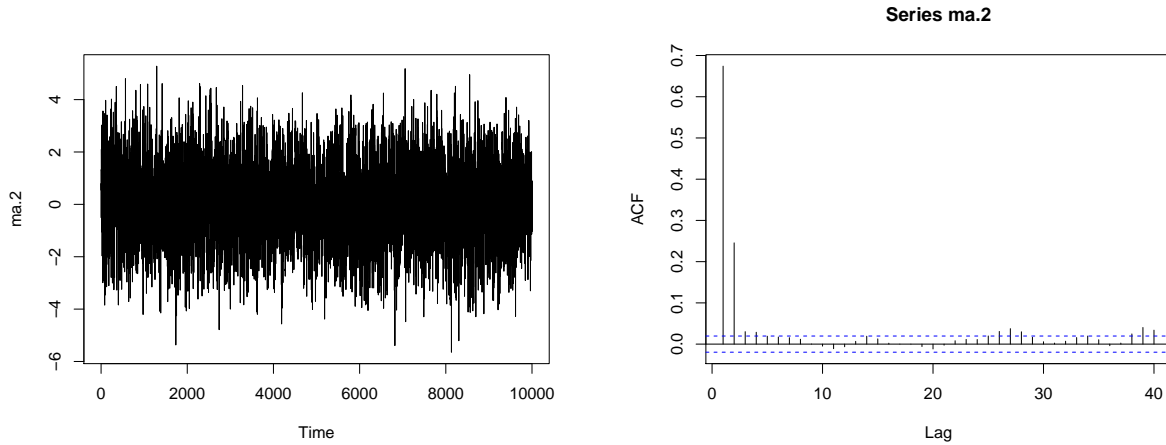
For the $MA(2)$ process, we have

$$Y_t = e_t + \beta_1 e_{t-1} + \beta_2 e_{t-2} \quad (12)$$

Then, we can derive the autocorrelation function as follow

$$\rho_k = \begin{cases} 1 & (k = 0) \\ \frac{\beta_1 + \beta_1 \beta_2}{1 + \beta_1^2 + \beta_2^2} & (k = 1) \\ \frac{\beta_2}{1 + \beta_1^2 + \beta_2^2} & (k = 2) \\ 0 & (k > 2) \end{cases} \quad (13)$$

Then, we simulate $MA(2)$ with $\beta_1 = 1, \beta_2 = 0.5$, and show the ACF plot as Figure 9.



(a) The simulated $MA(2)$ series

(b) The acf of simulated $MA(2)$ series

Figure 9 The plot of simulated $MA(2)$ process

Subproblem (c)

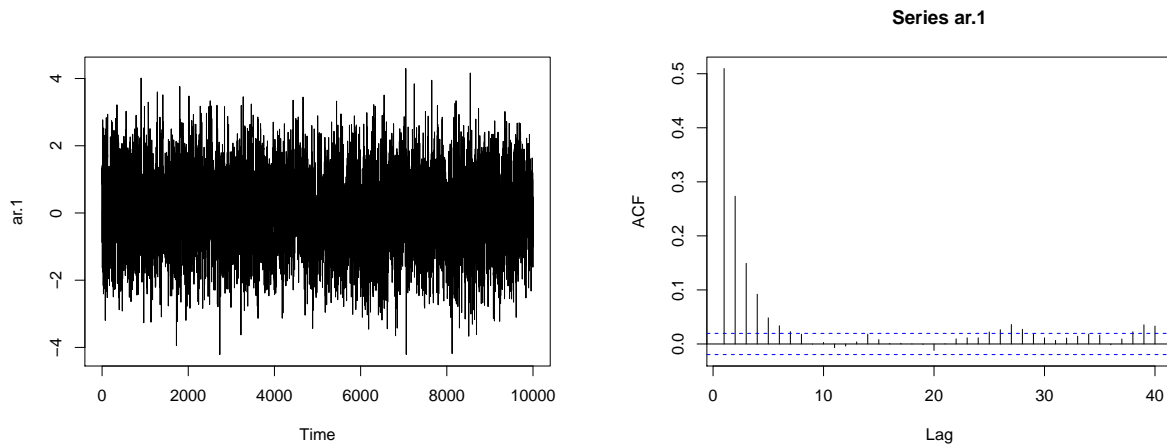
For the $AR(1)$ process, we have

$$Y_t = \alpha Y_{t-1} + e_t \quad (14)$$

Then, we can derive the autocorrelation function as follow

$$\rho_k = \alpha^k \quad (15)$$

Then, we simulate $AR(1)$ with $\alpha = 0.5$, and show the ACF plot as Figure 10.



(a) The simulated $AR(1)$ series

(b) The acf of simulated $AR(1)$ series

Figure 10 The plot of simulated $AR(1)$ process

Subproblem (d)

For the $ARMA(1, 1)$ process, we have

$$Y_t = \alpha Y_{t-1} + e_t + \beta e_{t-1} \quad (16)$$

Then, we can derive the autocorrelation function as follow

$$\rho_k = \begin{cases} 1 & (k = 0) \\ \frac{\alpha^{k-1}(\alpha + \beta)(1 + \alpha\beta)}{1 + 2\alpha\beta + \beta^2} & (k \geq 1) \end{cases} \quad (17)$$

Then, we simulate $ARMA(1, 1)$ with $\alpha = 0.7, \beta = 0.5$, and show the ACF plot as Figure 11.

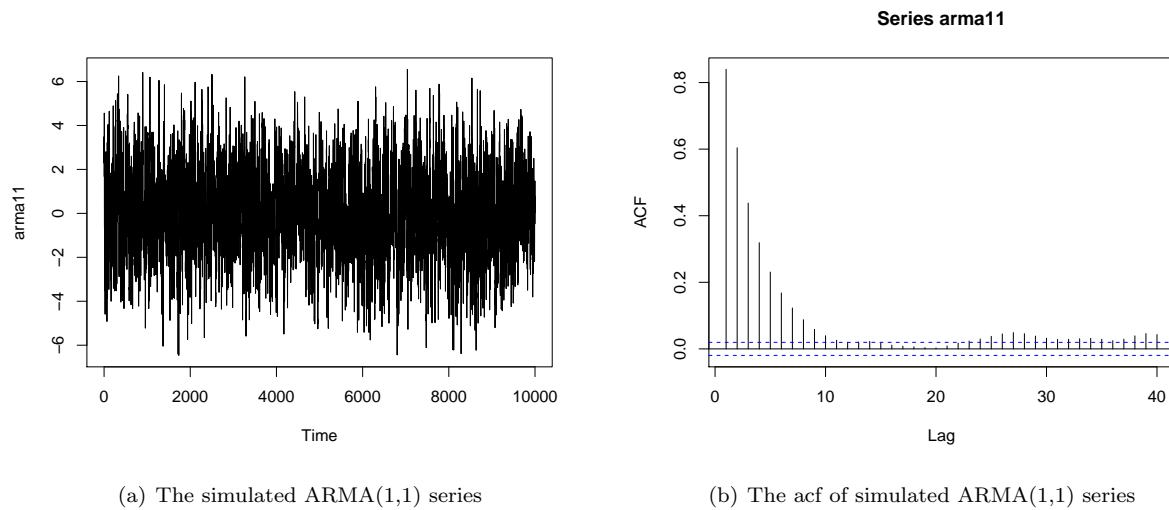


Figure 11 The plot of simulated ARMA(1,1) process

Problem 5

The daily returns of Google stock from August 14, 2004 to September 13, 2006 are stored in `data(google)` in the package *TSA*.

- Display the time sequence plot for the return data and show that the data are essentially uncorrelated over time. Is there changes in variance over time?
- Fit $GARCH(1,1)$ model for the Google daily return data. Draw and comment on the time sequence plot of the estimated conditional variances. Does the model appear to have addressed the conditional heteroskedasticity?

Solution

Subproblem (a)

We plot the google stock daily return time series as Figure 12(a), and the acf of google stock daily return time series as Figure 12(b). As we can see, there is no evident autocorrelation over time. Thus, the data are essentially uncorrelated over time.

As we can see from Figure 12(a), the variance seems not constant over time. Then, we can plot the correlogram of the squared values to detect the volatility as Figure 13. As we can see, the variance is correlated in time, since the series exhibits volatility which can be derived from Figure 13(b).

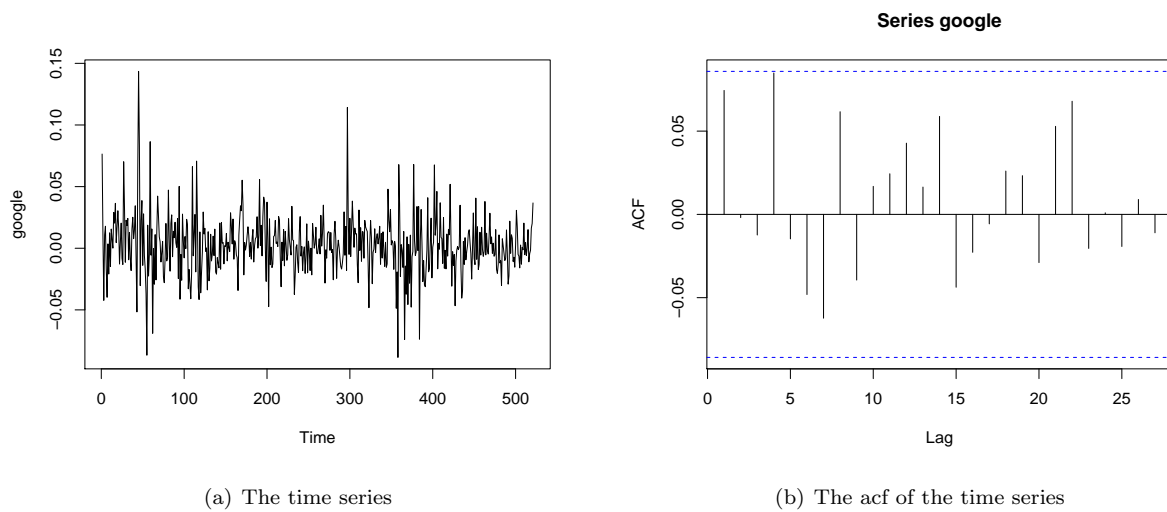


Figure 12 The google stock daily return time series

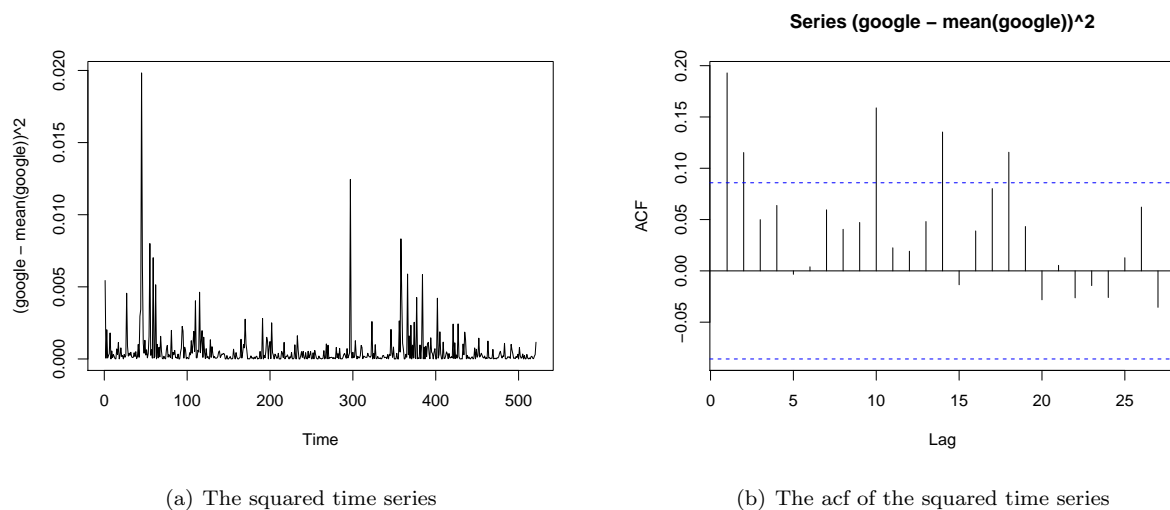


Figure 13 The squared value of google stock daily return time series

Subproblem (b)

We fit the google daily return data with GARCH(1,1) model, the series with fitted standard deviation is as Figure 14. And the time sequence plot of the estimated conditional variance is as Figure 15. Then, we would like to check if the model have addressed the conditional heteroskedasticity. As we know, if all estimated conditional variance captures the conditional variance, the residuals will be white noise with mean 0 and variance 1.

Thus, we would like to check the residuals. The test result is as follow and the test figures are as Figure 16.

Diagnostic Tests:

Jarque Bera Test

data: Residuals

X-squared = 223.48, df = 2, p-value < 2.2e-16

Box-Ljung test

```
data: Squared.Residuals  
X-squared = 0.0004683, df = 1, p-value = 0.9827
```

As we can see from the Jarque Bera Test result, the p -value $< 2.2e^{-16}$, which means the distribution of residuals is not normal. On the other hand, as we can see from the Box-Ljung test result, the p -value equals 0.9827, which means there is no serial autocorrelation in residuals. With the ACF figure of residuals in Figure 16, we can derive that the residuals can be treated as white noise, which means the GARCH(1,1) model has addressed the conditional heteroskedasticity, although the distribution of residuals is not normal.

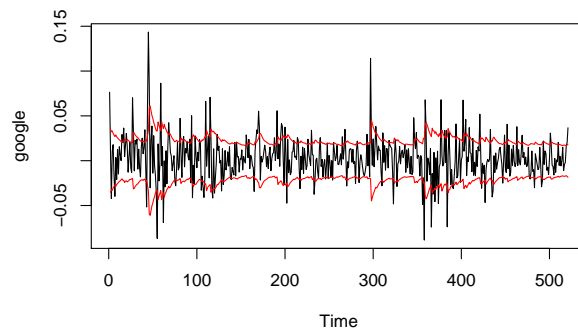


Figure 14 The google daily return data with GARCH(1,1) model

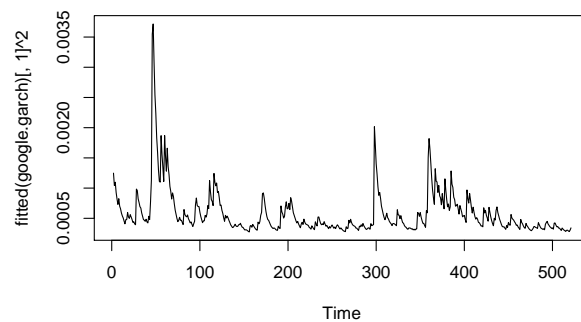


Figure 15 The time sequence plot of the estimated conditional variance

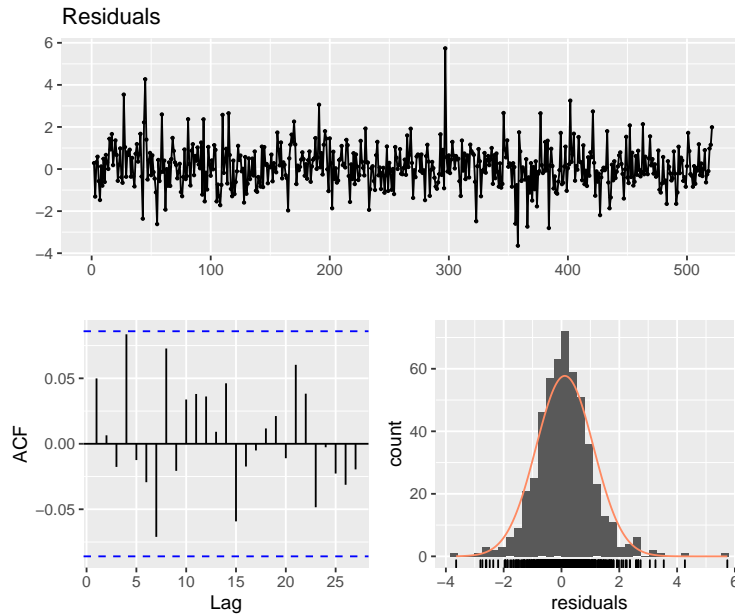


Figure 16 The check of residuals

Problem 6

Consider the $AR(1)$ model

$$Y_t = \phi Y_{t-1} + e_t$$

The theoretical spectral density is

$$S(f) = \frac{\sigma_e^2}{1 + \phi^2 - 2\phi \cos(2\pi f)}$$

- Show that when $\phi > 0$ the spectral density for an $AR(1)$ process is a decreasing function of frequency, while for $\phi < 0$ the spectral density increases.
- Graph the theoretical spectral density for an $AR(1)$ time series with $\phi = 0.7$. Interpret the implications of the shape of the spectrum on the possible plots of the time series values.
- Graph the theoretical spectral density for an $AR(1)$ time series with $\phi = -0.4$. Interpret the implications of the shape of the spectrum on the possible plots of the time series values.

Solution

Subproblem (a)

We can derive the derivative function of $S(f)$ as follow

$$S'(f) = \frac{-4\pi\phi\sigma_e^2 \sin(2\pi f)}{[1 + \phi^2 - 2\phi \cos(2\pi f)]^2} \quad (18)$$

We know that when $f \in [0, \frac{1}{2}]$, $\sin(2\pi f) \geq 0$. Thus, we have

$$\begin{cases} S'(f) \leq 0 & (\phi > 0) \\ S'(f) \geq 0 & (\phi < 0) \end{cases} \quad (19)$$

Thus, we can derive when $\phi > 0$ the spectral density for an $AR(1)$ process is a decreasing function of frequency, while for $\phi < 0$ the spectral density increases.

Subproblem (b)

We plot the simulated $AR(1)$ time series with $\phi = 0.7$ as Figure 17. As we can see from Figure 17, because $\phi = 0.7$ is larger than 0, so there is no very strong variation in the time series. And we can see that the ACF is positive mostly.

We graph the theoretical spectral density for an $AR(1)$ time series with $\phi = 0.7$ as Figure 18. As we can see from Figure 18, the spectral density decreases rapidly at the beginning and then decreases slowly later, which agreed with the conclusion derived in subproblem (a). Besides, the density is much stonger for lower frequencies than for high frequencies, this is because the time series's variation is not very strong, which means the process tends to change slowly from one time instance to the next.

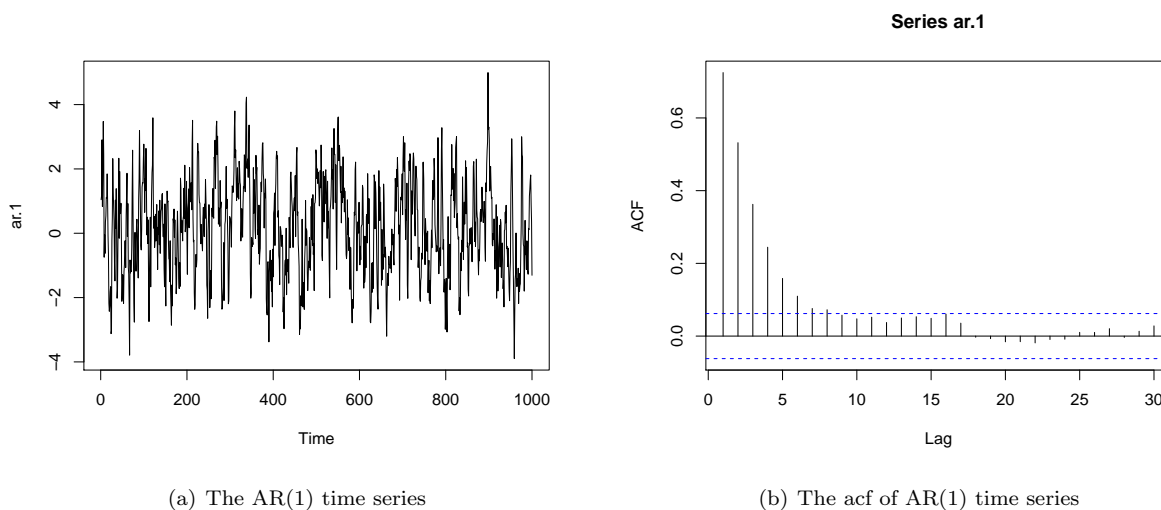


Figure 17 The AR(1) time series with $\phi = 0.7$

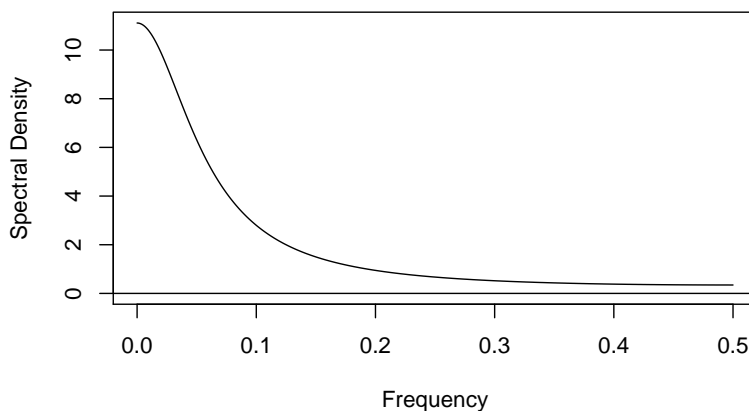
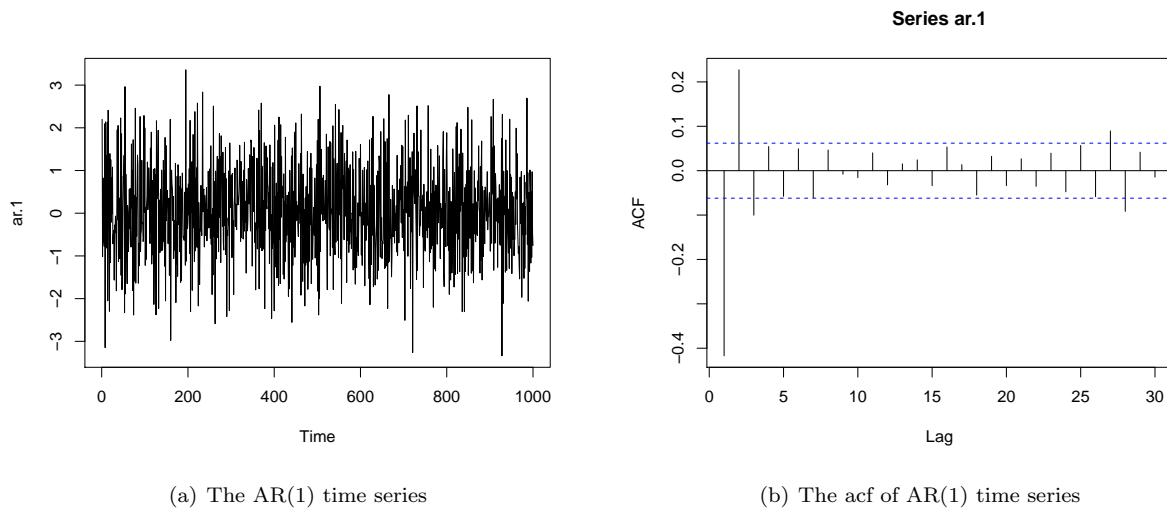
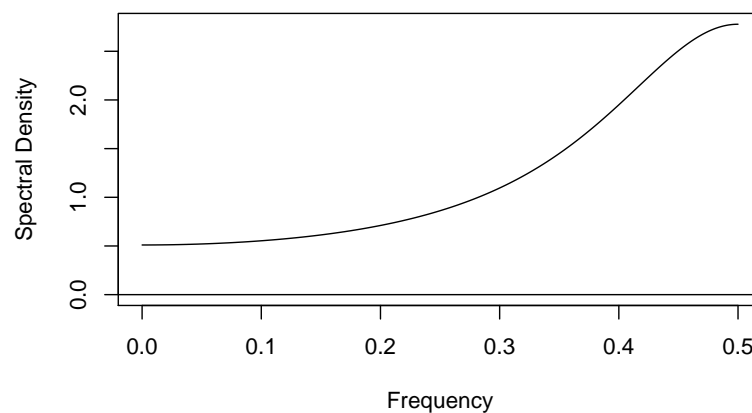


Figure 18 The theoretical spectral density of AR(1) time series with $\phi = 0.7$

Subproblem (c)

We plot the simulated $AR(1)$ time series with $\phi = -0.4$ as Figure 19. As we can see from Figure 19, because $\phi = -0.4$ is less than 0, so there is very strong variation in the time series. And we can see that the ACF is positive and negative alternate.

We graph the theoretical spectral density for an $AR(1)$ time series with $\phi = -0.4$ as Figure 20. As we can see from Figure 20, the spectral density is increasing with the increasing of frequency, which is agreed with the conclusion derived in subproblem (a). Besides, the density is much stronger for higher frequencies than for low frequencies, this is because the time series's variation is very strong, which means the process tends to oscillate back and forth across its mean level quickly.

(a) The $AR(1)$ time series(b) The acf of $AR(1)$ time series**Figure 19 The $AR(1)$ time series with $\phi = -0.4$** **Figure 20 The theoretical spectral density of $AR(1)$ time series with $\phi = -0.4$**

Problem 7

Consider the $MA(1)$ model

$$Y_t = e_t - \theta e_{t-1}$$

The theoretical spectral density is:

$$S(f) = [1 + \theta^2 - 2\theta \cos(2\pi f)] \sigma_e^2$$

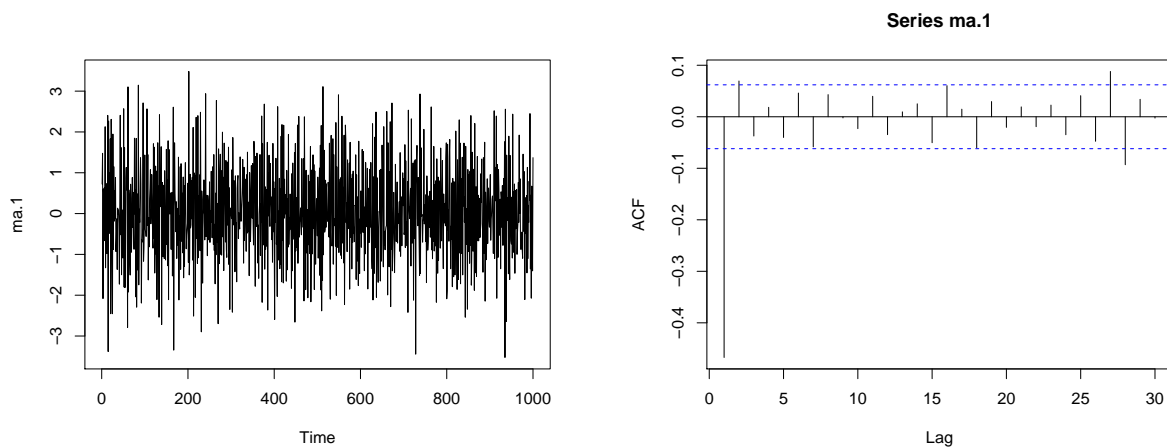
- Graph the theoretical spectral density for an $MA(1)$ process with $\theta = 0.6$. Interpret the implications of the shape of the spectrum on the possible plots of the time series values.
- Graph the theoretical spectral density for an $MA(1)$ process with $\theta = -0.8$. Interpret the implications of the shape of the spectrum on the possible plots of the time series values.

Solution

Subproblem (a)

We plot the simulated $MA(1)$ time series with $\theta = 0.6$ as Figure 21.

We graph the theoretical spectral density for an $MA(1)$ time series with $\theta = 0.6$ as Figure 22. As we can see from Figure 22, the spectral density is increasing with the increasing of frequency. Besides, the density is much stronger for higher frequencies than for low frequencies, this is because the time series's variation is very strong, which means the process tends to oscillate back and forth across its mean level quickly.



(a) The $MA(1)$ time series

(b) The acf of $MA(1)$ time series

Figure 21 The $MA(1)$ time series with $\theta = 0.6$

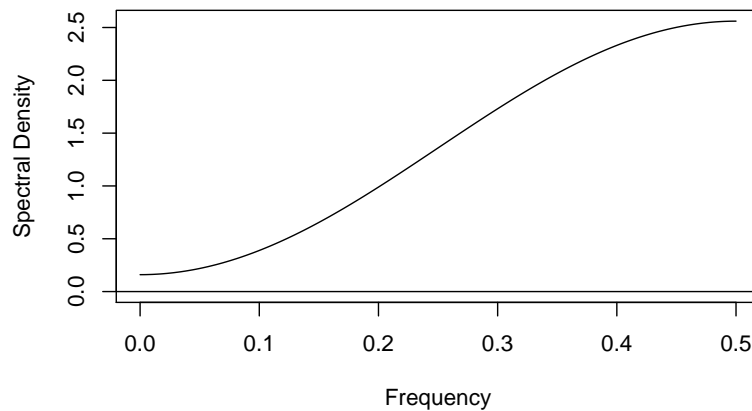
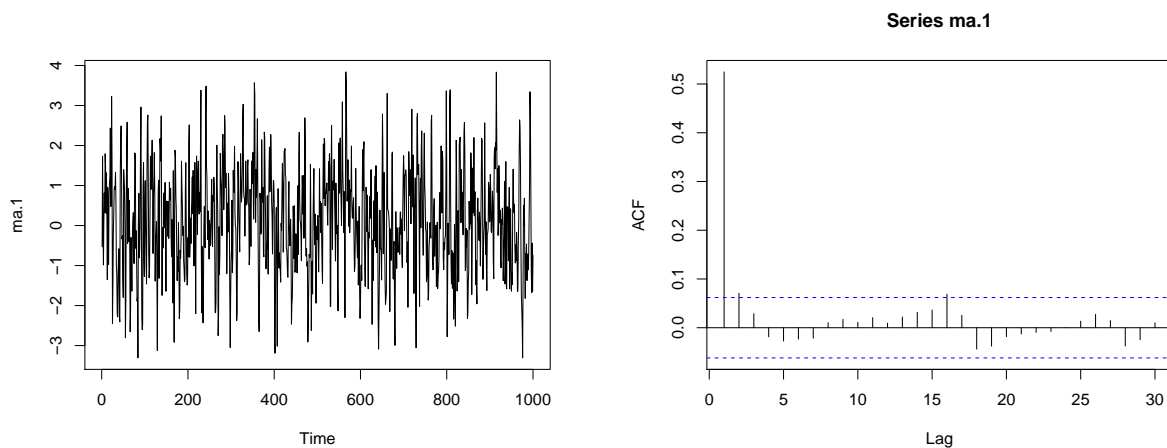


Figure 22 The theoretical spectral density of MA(1) time series with $\theta = 0.6$

Subproblem (b)

We plot the simulated $MA(1)$ time series with $\theta = -0.8$ as Figure 23.

We graph the theoretical spectral density for an $MA(1)$ time series with $\theta = -0.8$ as Figure 24. As we can see from Figure 24, the spectral density is decreasing with the increasing of frequency. Besides, the density is much stronger for lower frequencies than for high frequencies, this is because the time series's variation is not very strong, which means the process tends to change slowly from one time instance to the next.



(a) The MA(1) time series

(b) The acf of MA(1) time series

Figure 23 The MA(1) time series with $\theta = -0.8$

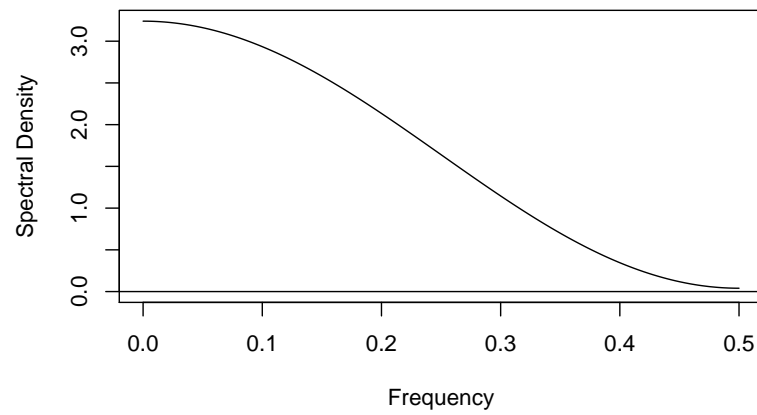


Figure 24 The theoretical spectral density of MA(1) time series with $\theta = -0.8$