2 Chapter 2: Linear Filters

2.1 Introduction

Suppose there are two weakly stationary processes X and Y related by

$$Y_t = \sum_{r=-\infty}^{\infty} c_r X_{t-r}, \quad -\infty < t < \infty, \text{ where } \sum_{r=-\infty}^{\infty} c_r^2 < \infty \tag{2.1}$$

and suppose their spectral densities are $f_X(\lambda)$ and $f_Y(\lambda)$.

We have

$$\begin{aligned} Cov\left(Y_t, Y_{t+k}\right) &= Cov\left(\sum_{r=-\infty}^{\infty} c_r X_{t-r}, \sum_{s=-\infty}^{\infty} c_s X_{t+k-s}\right) \\ &= \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} c_r c_s \gamma_{k+r-s} \\ &= \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} c_r c_s \int_{-\pi}^{\pi} e^{i(k+r-s)\lambda} f_X(\lambda) d\lambda \\ &= \int_{-\pi}^{\pi} e^{ik\lambda} \left| \sum c_r e^{ir\lambda} \right|^2 f_X(\lambda) d\lambda \\ &= \int_{-\pi}^{\pi} e^{ik\lambda} f_Y(\lambda) d\lambda \end{aligned} \tag{2.2}$$

Comparison of the last two equations shows that

$$f_Y(\lambda) = \left| C\left(e^{i\lambda}\right) \right|^2 f_X(\lambda) \tag{2.3}$$

where $C(z) = \sum c_r z^r$ is the generating function of the filter.

Equation 2.3 is the main result of this section. It allows us to calculate the effect of applying any linear filter to a given process $\{X_t\}$.

2.2 Application to AR processes

Define a backshift operator B by

$$BX_t = X_{t-1}, \; B^2 X_t = B(BX_t) = BX_{t-1} = X_{t-2}, \ldots \tag{2.4}$$

including the identity $IX_t = B^0 X_t = X_t$. Using this notation, we may formally write an AR(p) process as

$$\left(I - \sum_{r=1}^{p} \phi_r B^r\right) X_t = \epsilon_t, \tag{2.5}$$

or in even more compact notation as

$$\phi(B)X = \epsilon, \tag{2.6}$$

where $\phi(z)$ is the generating function $1 - \sum \phi_r z^r$.

Applying 2.3 leads to the formula for the spectral density

$$\left|\phi\left(e^{i\lambda}\right)\right|^2 f_X(\lambda) = f_\epsilon(\lambda) = \frac{\sigma_\epsilon^2}{2\pi}, \tag{2.7}$$

and hence

$$f_X(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} \frac{1}{\left|\phi\left(e^{i\lambda}\right)\right|^2}. \tag{2.8}$$

Now we can get all covariances of X by the Taylor expansion of 2.8 in powers of $e^{i\lambda}$ and using $f(\lambda) = \frac{\gamma_0}{2\pi} + \frac{1}{\pi}\sum_{k=1}^{\infty} \gamma_k \cos(\lambda k)$. In order to use the Taylor expansion, we need the following assumption:

(*) *All the zeros of the function $\phi(z)$ lie outside the unit circle in the complex plane.*

To see this, note that if the p complex zeros are at $z_1, \ldots, z_p$, then we can write

$$\phi(z) = \prod_{j=1}^{p} (z - z_j) = \prod_{j=1}^{p} \left\{-z_j\left(1 - \frac{z}{z_j}\right)\right\}, \tag{2.9}$$

and we can expand

$$\left(1 - \frac{z}{z_j}\right)^{-1} = \sum_{s=0}^{\infty} \left(\frac{z}{z_j}\right)^s \tag{2.10}$$

if and only if $|z_j| > 1$.

The relation (*) is called the *stationarity condition* for a causal AR(p) process. It defines exactly what condition is needed on the coefficients $\{\phi_r, \quad r = 1, \ldots, p\}$ to ensure that the process is well-defined and stationary.

For example, for an AR(1) process with $\phi(z) = 1 - \phi_1 z$, we find immediately that $|z_1| = |1/\phi_1| > 1$ is the stationarity condition. Also,

$$\begin{aligned}
\left|\phi\left(e^{i\lambda}\right)\right|^2 &= \left(1 - \phi_1 e^{i\lambda}\right)\left(1 - \phi_1 e^{-i\lambda}\right) \\
&= 1 - \phi_1\left(e^{i\lambda} + e^{-i\lambda}\right) + \phi_1^2 \\
&= 1 - 2\phi_1 \cos\lambda + \phi_1^2,
\end{aligned} \tag{2.11}$$

which leads directly back to the formula which we derived earlier.

2.3 The MA process

The MA(q) process

$$X_t = \epsilon_t + \sum_{s=1}^{q} \theta_s \epsilon_{t-s} \tag{2.12}$$

can similarly be re-written in a polynomial form as

$$X_t = \left(1 + \sum_{s=1}^{q} \theta_s B^s\right) \epsilon_t, \tag{2.13}$$

or more compactly

$$X = \theta(B)\epsilon \tag{2.14}$$

where $\theta(z)$ is the generating function $1 + \sum_1^q \theta_s z^s$. In this case the spectral density function is

$$f_X(\lambda) = \frac{\sigma_\epsilon^2}{2\pi}\left|\theta(e^{i\lambda})\right|^2. \tag{2.15}$$

There is no need for any stationarity condition, since the process is stationary whatever the coefficients $\{b_s\}$, but there is nevertheless a difficulty requiring some restriction on the coefficients. This is most easily seen in the case $q = 1$. In that case we easily calculate the autocovariances to be

$$\gamma_0 = \left(1 + \theta_1^2\right)\sigma_\epsilon^2, \quad \gamma_1 = \theta_1 \sigma_\epsilon^2, \quad \gamma_k = 0 \text{ for } k > 1, \tag{2.16}$$

and hence the autocorrelations

$$\rho_0 = 1, \quad \rho_1 = \frac{\theta_1}{1 + \theta_1^2}, \quad \rho_k = 0 \text{ for } k > 1 \tag{2.17}$$

Now consider the identical process, but with θ_1 replaced by $1/\theta_1$. It is seen from 2.17 that the autocorrelation function is unchanged by this transformation. In other words, the two process defined by θ_1 and $1/\theta_1$ are identical for all practical purposes, so that the two processes cannot be distinguished.

As a resolution of this difficulty, it is customary to impose the following *identifiability condition*:

(**) *All the zeros of the function $\theta(z)$ lie on or outside the unit circle in the complex plane.*

To see why this resolves the problem, suppose we write

$$\theta(z) = \prod_{j=1}^{q} (z - z_j). \tag{2.18}$$

Then

$$\left|\theta\left(e^{i\lambda}\right)\right|^2 = \prod_{j=1}^{q}\left\{\left(e^{i\lambda}-z_j\right)\left(e^{-i\lambda}-z_j\right)\right\}. \tag{2.19}$$

However, the identity

$$\left(e^{i\lambda}-z_j\right)\left(e^{-i\lambda}-z_j\right) = z_j^2\left(e^{i\lambda}-\frac{1}{z_j}\right)\left(e^{-i\lambda}-\frac{1}{z_j}\right) \tag{2.20}$$

shows that there is no change in $|\theta(e^{i\lambda})|$, except for a constant, in replacing any z_j by $1/z_j$. Hence there is no loss of generality in assuming all $|z_j| \geq 1$. The restricted relation (**) that $|z_j| > 1$ is also known as the invertibility relation on the coefficients $\{\theta_j, 1 \leq j \leq q\}$.

2.4 Partial Autocorrelation Coefficients for MA models

We shall derive partial autocorrelation coefficients (pacf) for MA models. We will illustrate it by considering an MA(1) model that can be rewritten in an autoregressive form (rather than in its natural Yule-Wold representation). We have

$$Y_t = \epsilon_t + \theta_1\epsilon_{t-1} \tag{2.21}$$

Also

$$Y_{t-1} = \epsilon_{t-1} + \theta_1\epsilon_{t-2}$$

that after rewriting in terms of ϵ_{t-2} yields

$$\epsilon_{t-1} = Y_{t-1} - \theta_1\epsilon_{t-2}.$$

Substituting this into 2.21 gives

$$Y_t = \epsilon_t + \theta_1(Y_{t-1} - \theta_1\epsilon_{t-2}) = \epsilon_t + \theta_1 Y_{t-1} - \theta_1^2\epsilon_{t-2}. \tag{2.22}$$

Now

$$Y_{t-2} = \epsilon_{t-2} + \theta_1\epsilon_{t-3}$$

and rewriting this in terms of ϵ_{t-2} and substituting it into 2.21 yields

$$\begin{aligned} Y_t &= \epsilon_t + \theta_1 Y_{t-1} - \theta_1^2\left(Y_{t-2} - \theta_1\epsilon_{t-3}\right) \\ &= \epsilon_t + \theta_1 Y_{t-1} - \theta_1^2 Y_{t-2} + \theta_1^3\epsilon_{t-3}. \end{aligned} \tag{2.23}$$

In a similar manner we obtain the **infinite autoregressive expansion**

$$Y_t = \epsilon_t + \theta_1 Y_{t-1} - \theta_1^2 Y_{t-2} + \theta_1^3 Y_{t-3} - \ldots + \theta_1^n Y_{t-n} - \ldots \tag{2.24}$$

From this we see that the pacf of an MA(1) process never cuts off because the dependence on past variables stretches back infinitely far in time.

In general every MA(q) process has an infinite autoregressive expansion and its pacf never cuts off. In fact it can be shown that the pacf of an MA(q) process damps out according to a mixed exponential decay of order q.

2.5 The ARMA process

The general ARMA (p, q) process may, in similar notation, be written in the form

$$\phi(B)X = \theta(B)\epsilon \tag{2.25}$$

where $\phi(\cdot)$ and $\theta(\cdot)$ are the respective generating functions of the autoregressive and moving average operators. By equating the spectral densities of the two sides of (1.21), we see that the spectral density of X is given by

$$f_X(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} \frac{\left|\theta\left(e^{i\lambda}\right)\right|^2}{\left|\phi\left(e^{i\lambda}\right)\right|^2}. \tag{2.26}$$

The conditions now required are:

(a) the stationarity condition on the coefficients $\{\phi_r, 1 \le r \le p\}$,

(b) the identifiability condition on the coefficients $\{\theta_s, 1 \le s \le q\}$,

(c) an additional identifiability condition: *the generating functions $\phi(\cdot)$ and $\theta(\cdot)$ should not have any common zero.*

The reason for condition (c) is that if $\phi(\cdot)$ and $\theta(\cdot)$ have a common zero at z^* say, then it is possible to cancel a common factor $(e^{i\lambda} - z^*)(e^{-i\lambda} - z^*)$ from both the numerator and denominator of 2.26 and so reduce the model to simpler form.

As an example, consider the AR(1) model

$$X_t = 0.8X_{t-1} + \epsilon_t. \tag{2.27}$$

This model also satisfies the equation

$$X_{t-1} = 0.8X_{t-2} + \epsilon_{t-1}, \tag{2.28}$$

so by subtracting, say 0.6 times 2.28 from 2.27, we obtain the model

$$X_t = 1.4X_{t-1} - 0.48X_{t-2} + \epsilon_t - 0.6\epsilon_{t-1}, \tag{2.29}$$

which looks like an ARMA(2,1) model, but of course it is in reality no different from 2.27. In this case, the zeros of $\phi(z) = 1 - 1.4z + 0.48z^2$ are at $1/0.6$ and $1/0.8$, while $\theta(z) = 1 - 0.6z$ is zero at $z = 1/0.6$, i.e. there is a common zero in the two generating functions, and when this is removed the model indeed reduces to 2.27.

2.6 Calculating autocovariances of ARMA models

One application of these formulae is to the calculation of autocovariances of ARMA models. This is useful, e.g. for deciding whether an estimated model provides a good fit to the observed time series, and can also be useful as an initial diagnostic tool.

As an example, consider the ARMA(1,2) process with generating functions

$$\phi(z) = 1 - \phi_1 z,\ \theta(z) = 1 + \theta_1 z + \theta_2 z^2 \tag{2.30}$$

and $|\phi_1| < 1$. In this case, the generating function of the Wold coefficients $\{c_r,\ r \geq 0\}$ is

$$\begin{aligned} C(z) &= \frac{\theta(z)}{\phi(z)} \\ &= \left(1 + \theta_1 z + \theta_2 z^2\right) \sum_{r=0}^{\infty} d_r z^r \\ &= \sum_{r=0}^{\infty} c_r z^r, \end{aligned} \tag{2.31}$$

so by equating coefficients of z^r, we find

$$c_r = \begin{cases} 1 & \text{if } r = 0, \\ \phi_1 + \theta_1 & \text{if } r = 1, \\ \phi_1^r + \theta_1 \phi_1^{r-1} + \theta_2 \phi_1^{r-2} & \text{if } r \geq 2. \end{cases} \tag{2.32}$$

To compute the covariances, we may use the fact that for $k \geq 0$

$$Cov\left(\sum_r c_r \epsilon_{t-r},\ \sum_s c_s \epsilon_{t+k-s}\right) = \left(\sum_{r=0}^{\infty} c_r c_{r+k}\right) \sigma_\epsilon^2, \tag{2.33}$$

so that

$$
\begin{aligned}
\gamma_0 &= \left\{1+(\phi_1+\theta_1)^2+\frac{\left(\phi_1^2+\theta_1\phi_1+\theta_2\right)^2}{1-\phi_1^2}\right\}\sigma_\epsilon^2,\\
\gamma_1 &= \left\{\phi_1+\theta_1+(\phi_1+\theta_1)\left(\phi_1^2+\theta_1\phi_1+\theta_2\right)+\frac{\left(\phi_1^2+\theta_1\phi_1+\theta_2\right)^2\phi_1}{1-\phi_1^2}\right\}\sigma_\epsilon^2,\\
\gamma_k &= \left\{1+(\phi_1+\theta_1)\phi_1+\frac{\left(\phi_1^2+\theta_1\phi_1+\theta_2\right)^2\phi_1^2}{1-\phi_1^2}\right\}\\
&\quad\left(\phi_1^2+\theta_1\phi_1+\theta_2\right)\phi_1^{k-2}\sigma_\epsilon^2,\quad k\geq 2.
\end{aligned} \tag{2.34}
$$

Note that an alternative approach which yields part of the answer is based on the following formula, valid for $k>2$:

$$
Cov\left\{X_t-\phi_1X_{t-1},\ X_{t-k}\right\}=Cov\left\{\epsilon_t+\theta_1\epsilon_{t-1}+\theta_2\epsilon_{t-2},\ X_{t-k}\right\}=0, \tag{2.35}
$$

from which we deduce

$$
\gamma_k=\phi_1\gamma_{k-1},\quad k>2. \tag{2.36}
$$

Although 2.36 does not yield the full solution 2.34, it may be the most useful part for identification purposes, because if it appears from empirical examination of the sample autocorrelations that they are geometrically decaying for $k\geq 2$; that could be taken as providing strong evidence that the process is of ARMA(1,2) form.

2.7 Overview of ARMA models

Since an ARMA(p,q) is a mixture of p autoregressive components and q moving average components, ARMA(p,q) inherited the properties of AR(p) and MA(q). ARMA(p,q) can be expressed as

$$
\phi^p(B)Y_t=\theta^q(B)\epsilon_t, \tag{2.37}
$$

where ϵ_t is white noise, $\phi^p(\cdot)$ and $\theta^q(\cdot)$ are the polynomials of degree p and q respectively. Examples of ARMA models and their expansions:

- **ARMA(1,2)**: short hand notation $\phi^1(B)Y_t=\theta^2(B)\epsilon_t$. Therefore

$$
(1-\phi_1B)\,Y_t=\left(1+\theta_1B+\theta_2B^2\right)\epsilon_t
$$

giving

$$
Y_t-\phi_1Y_{t-1}=\epsilon_t+\theta_1\epsilon_{t-1}+\theta_2\epsilon_{t-2}.
$$

Isolating Y_t on the left side of the equation gives the final form of the expansion

$$Y_t = \phi_1 Y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}.$$

- **ARMA(3,1)**: short hand notation $\phi^3(B)Y_t = \theta^1(B)\epsilon_t$. Therefore

$$\left(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3\right) Y_t = (1 + \theta_1 B)\, \epsilon_t$$

  or

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \phi_3 Y_{t-3} = \epsilon_t + \theta_1 \epsilon_{t-1}.$$

  Isolating Y_t gives the final form

$$Y_t = \epsilon_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \epsilon_t + \theta_1 \epsilon_{t-1}.$$

In a mixed model ARMA(p, q) process (that is, $p > 0,\ q > 0$), neither the acf nor the pacf cuts off abruptly. Both the acf and pacf exhibit mixed exponential decay. This occurs because the AR component introduces mixed exponential decay into the acf, while the MA component introduces mixed exponential decay into the pacf.

2.8 Summary of ARMA(p, q) Models

- AR(p) models: $\phi^p(B)Y_t = \epsilon_t$

    acf: p initial spikes, then damps out as a mixed exponential decay of order p (never 0)

    pacf: p initial spikes, then cuts off; pacf $= 0$ for $k \geq p + 1$

- MA(q) models: $Y_t = \theta^q(B)\epsilon_t$

    acf: q initial spikes, then cuts off; acf $= 0$ for $k \geq q + 1$

    pacf: q initial spikes, then damps out as a mixed exponential decay of order q (never 0)

- ARMA(p, q) models (mixed): $\phi^p(B)Y_t = \theta^q(B)\epsilon_t$

acf: has max (p, q) initial spikes, then damps out as a mixed exponential decay driven by the AR(p) component (never 0).

pacf: has max(p, q) initial spikes, then damps out as a mixed exponential decay driven by the MA(q) component (never 0).

```
###### we read USD-CAD data into R #######
> data<-read.table("C:\\CAD.txt", header=TRUE)

##### produce time series (scatter) plot of the data ######
> ts.plot(ts(data$VALUE, frequency=365, start = c(2000,1), end=c(2004,3)),
ylab="Currency exchange rate between USD and CAD")

##### produce the acf plot of the data ######
> acf(data$VALUE)

##### produce the acf plot of the data ######
> pacf(data$VALUE)

##### example how to simulate time series in R ######
>y<-arima.sim(n = 1000, list(ar = c(0.4, 0.2), ma = c(0.5)))
```

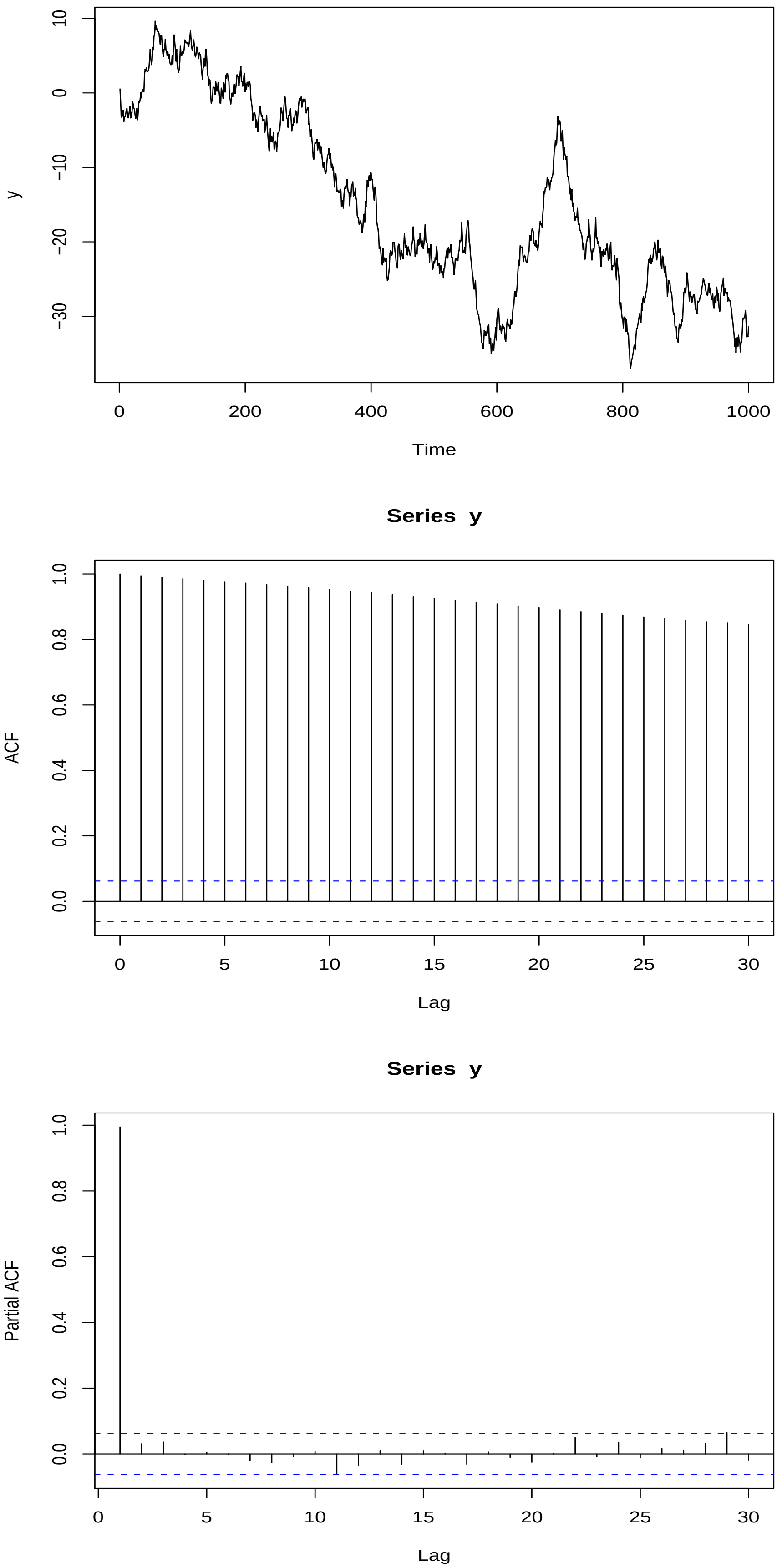

Figure 2.1: The scatter, acf and pacf plots of the simulated AR(1) process with $\phi_1 = 1$, i.e. random walk, 1000 observations.

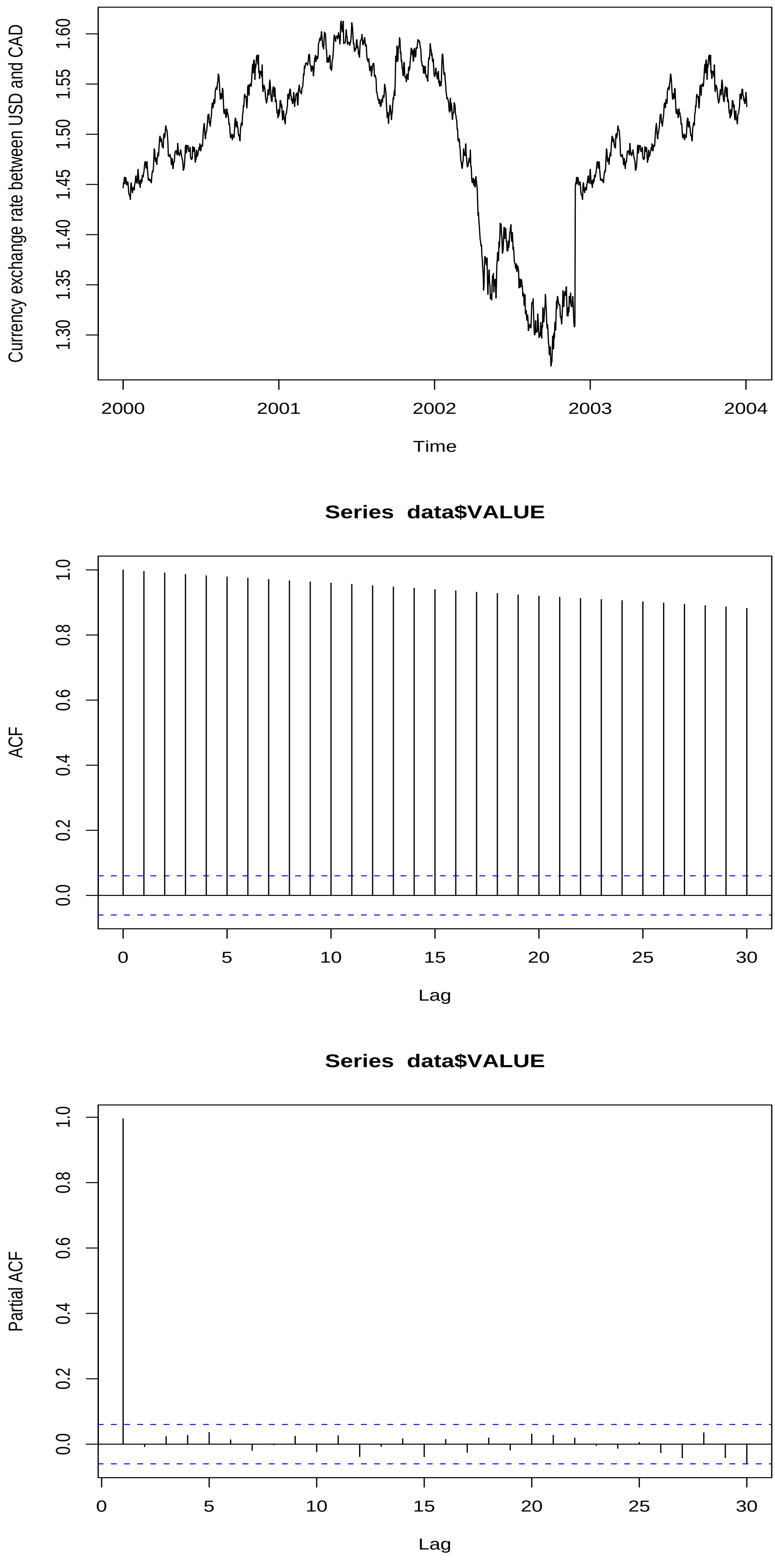

Figure 2.2: The scatter, acf and pacf plots of the currency exchange rate between USD and CAD from January 3, 2000 to March 31, 2004.

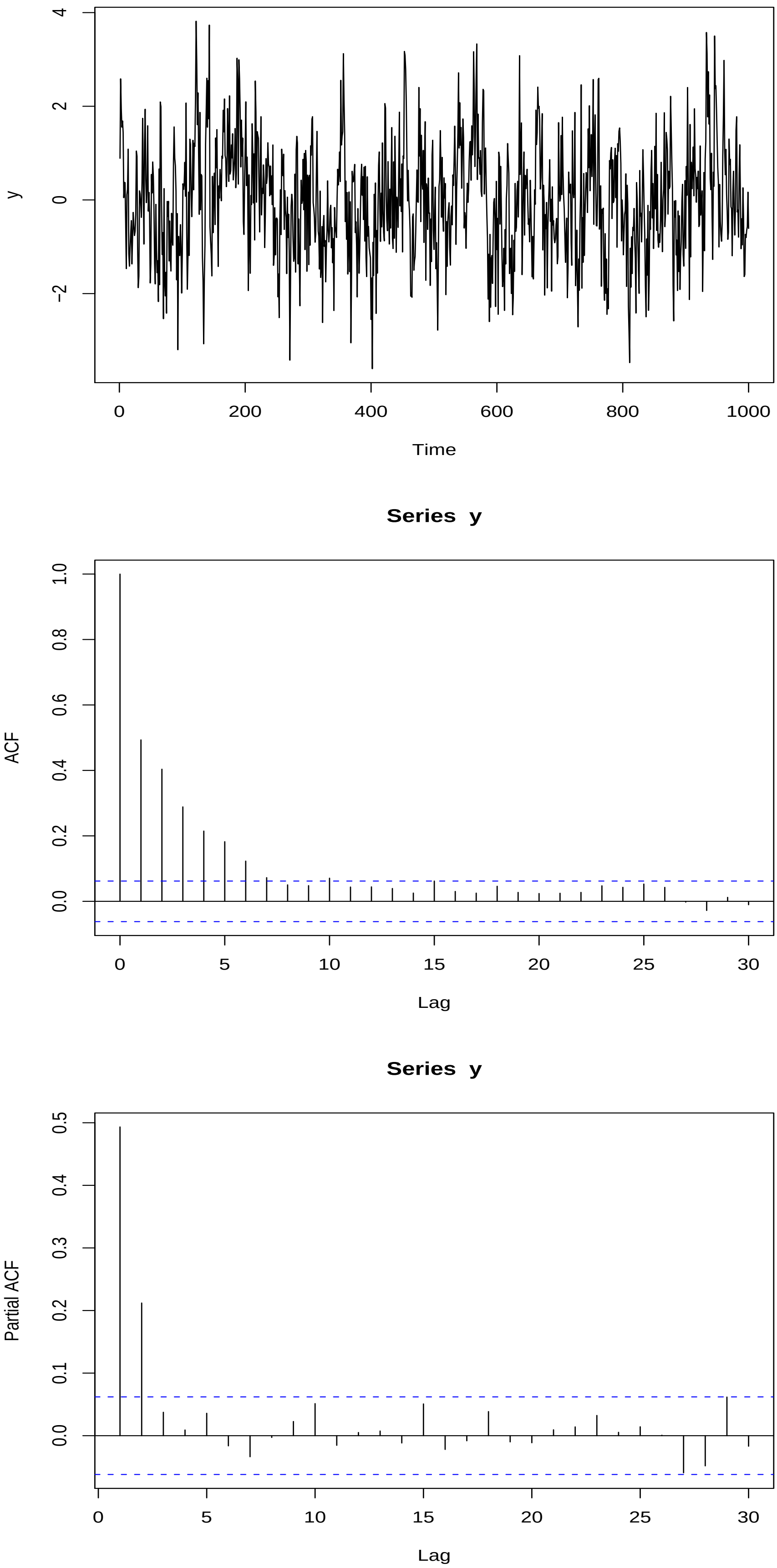

Figure 2.3: The scatter, acf and pacf plots of the simulated AR(2) process with $\phi_1 = 0.4$ and $\phi_2 = 0.2$, 1000 observations.

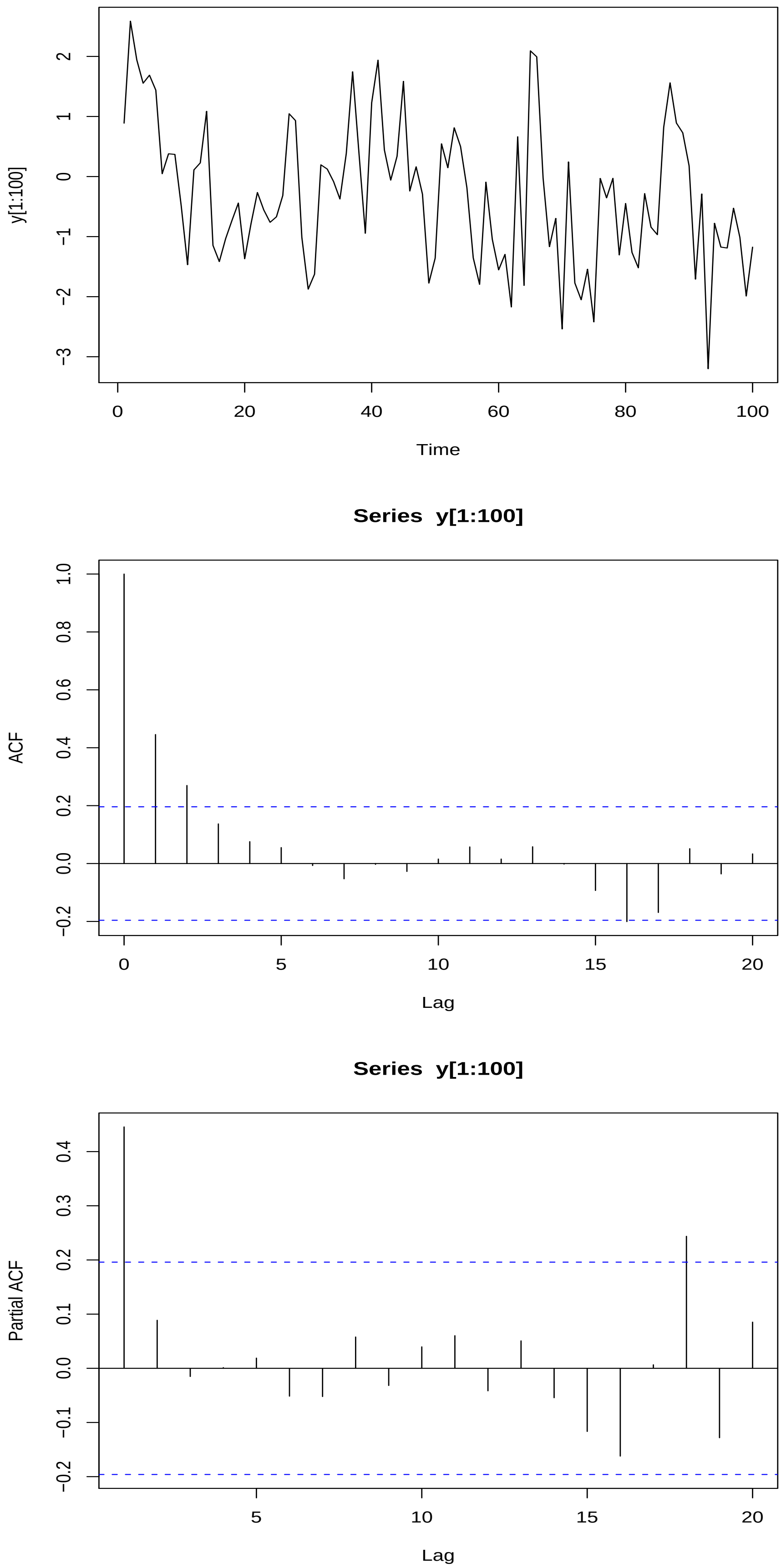

Figure 2.4: The scatter, acf and pacf plots of the subset of 100 observations from the simulated AR(2) process with $\phi_1 = 0.4$ and $\phi_2 = 0.2$.

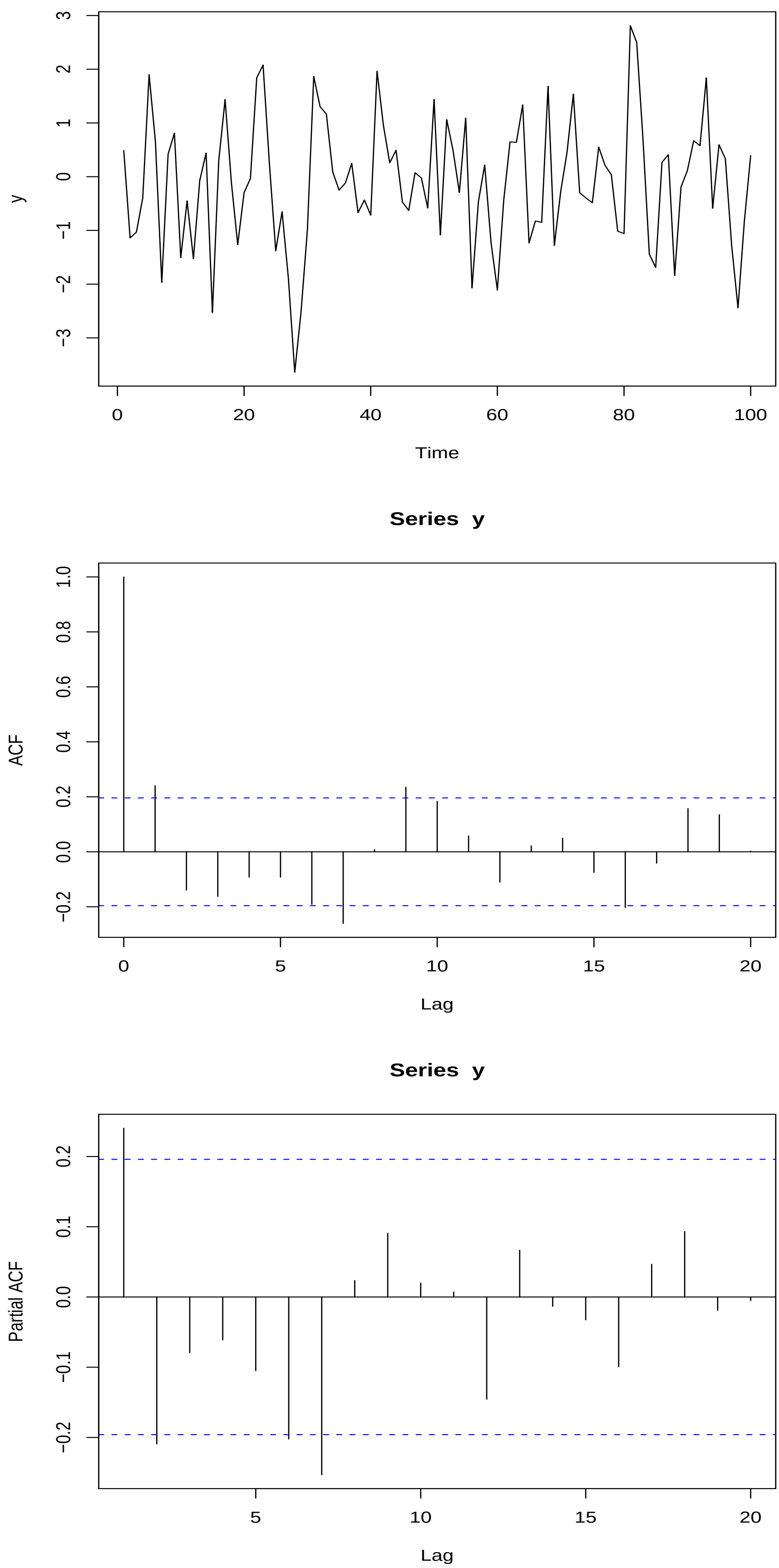

Figure 2.5: The scatter, acf and pacf plots of the simulated AR(2) process with $\phi_1 = 0.4$ and $\phi_2 = -0.2$, 100 observations.

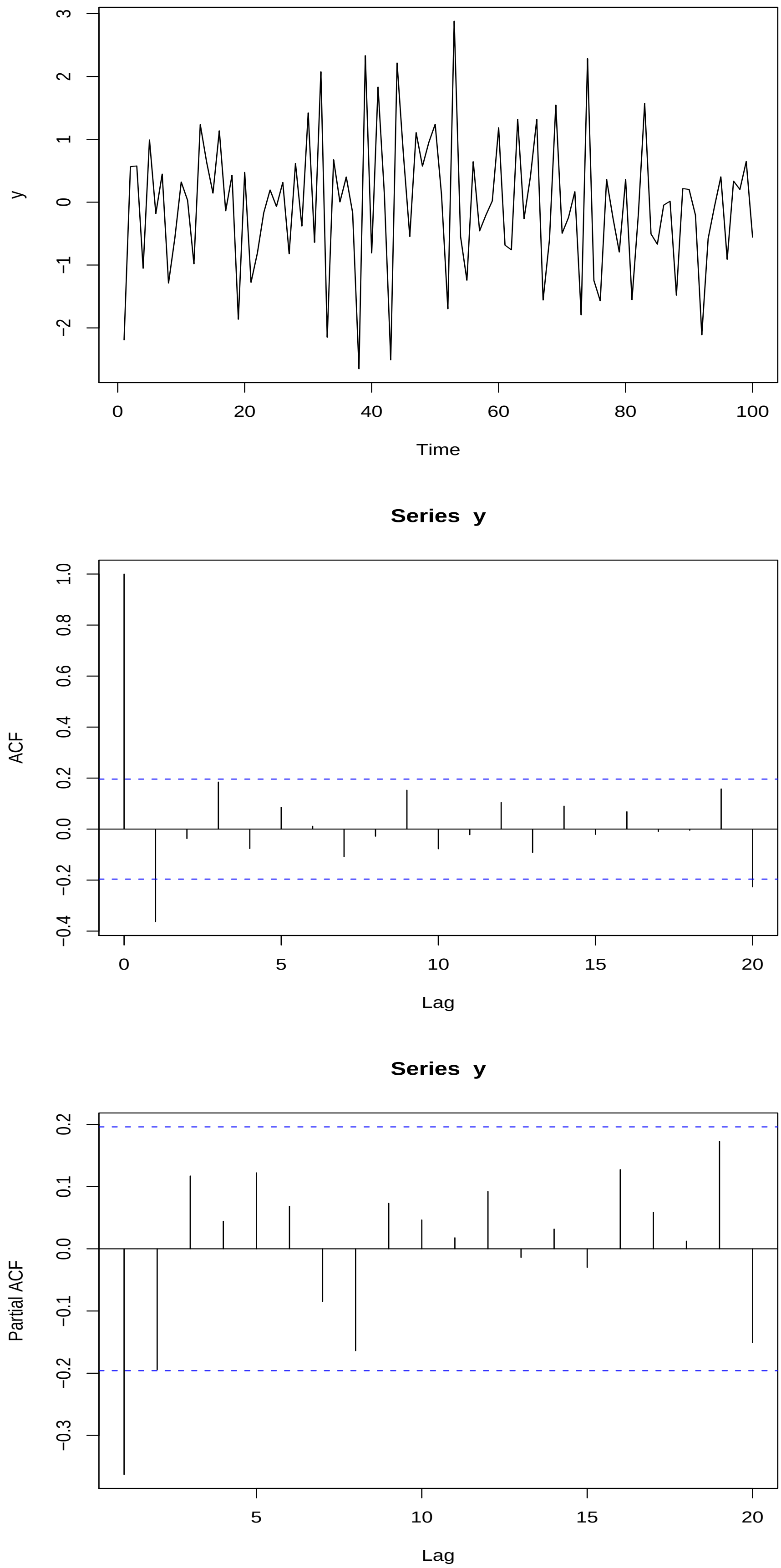

Figure 2.6: The scatter, acf and pacf plots of the simulated AR(2) process with $\phi_1 = -0.4$ and $\phi_2 = -0.2$, 100 observations.

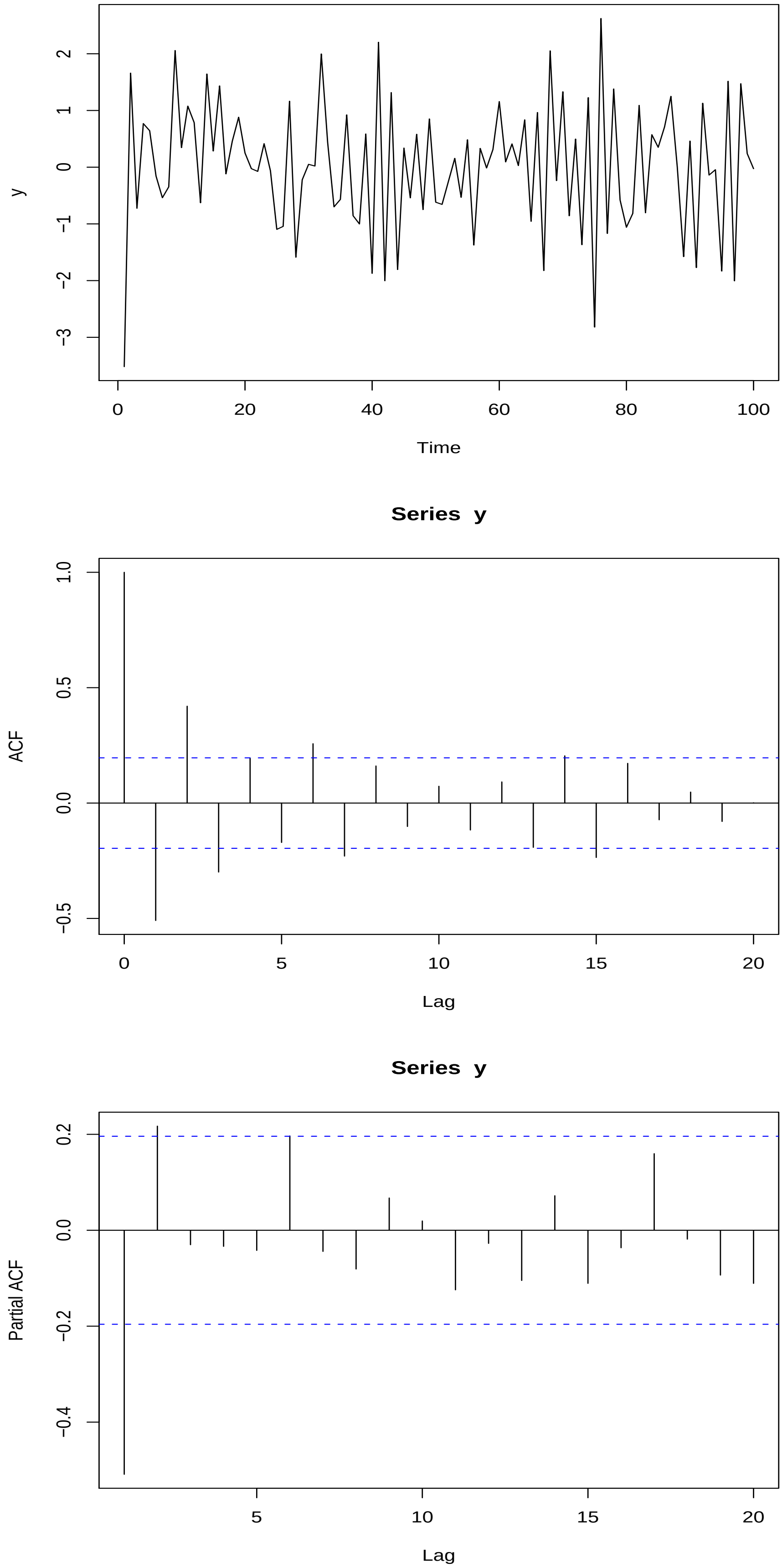

Figure 2.7: The scatter, acf and pacf plots of the simulated AR(2) process with $\phi_1 = -0.4$ and $\phi_2 = 0.2$, 100 observations.

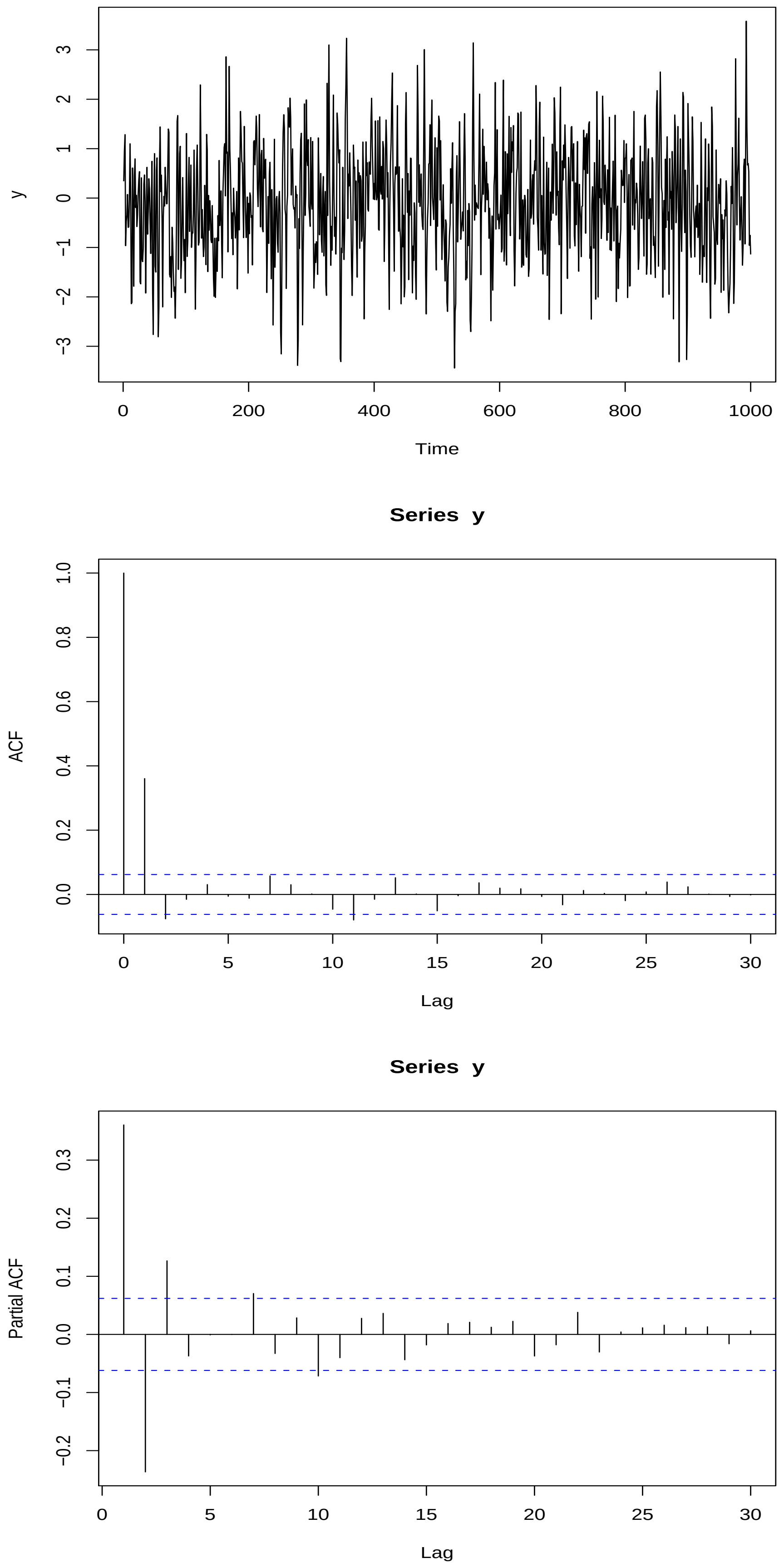

Figure 2.8: The scatter, acf and pacf plots of the simulated MA(1) process with $\theta_1 = 0.5$, 1000 observations.

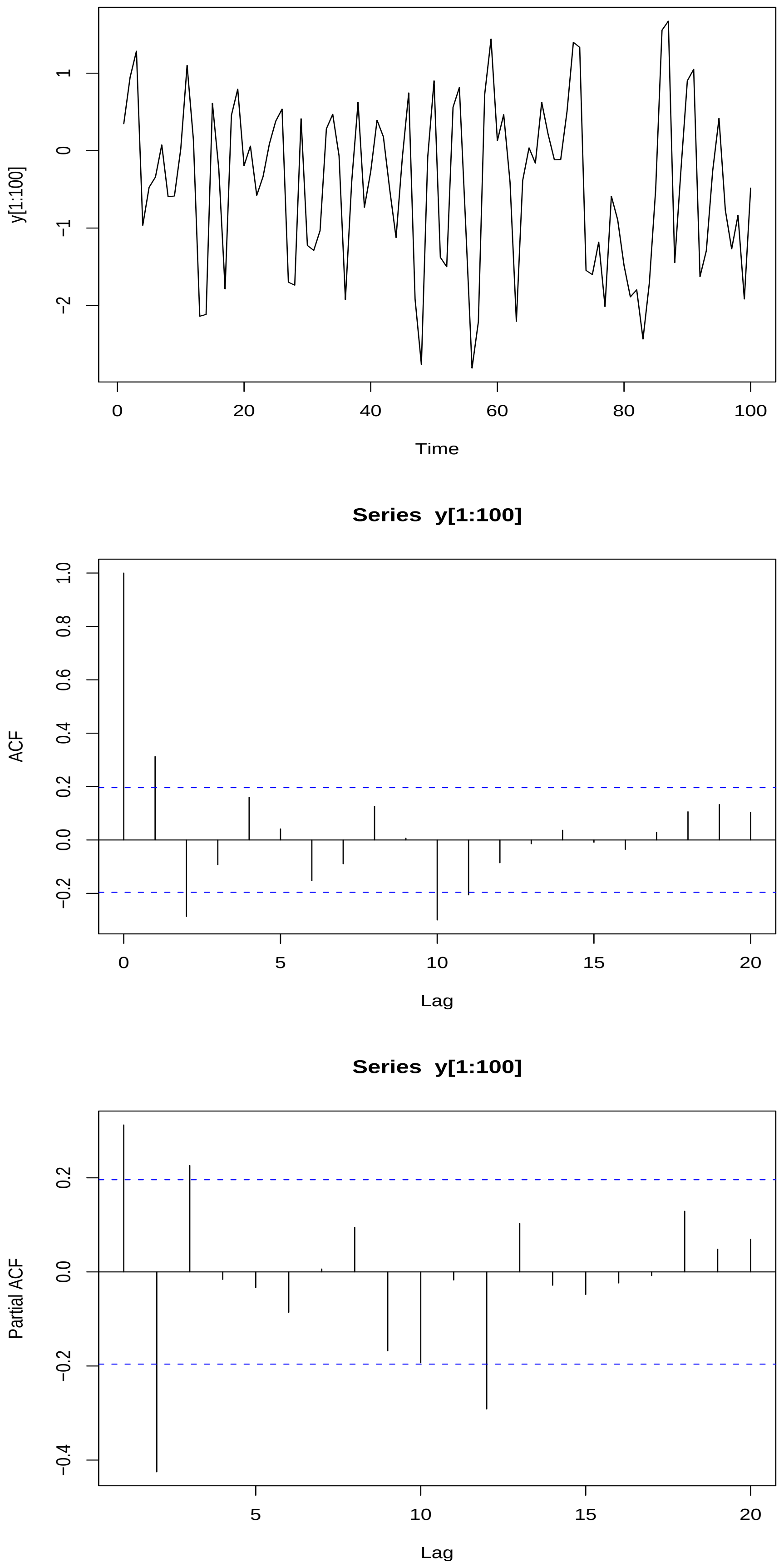

Figure 2.9: The scatter, acf and pacf plots of the subset of 100 observations form the simulated MA(1) process with $\theta_1 = 0.5$.

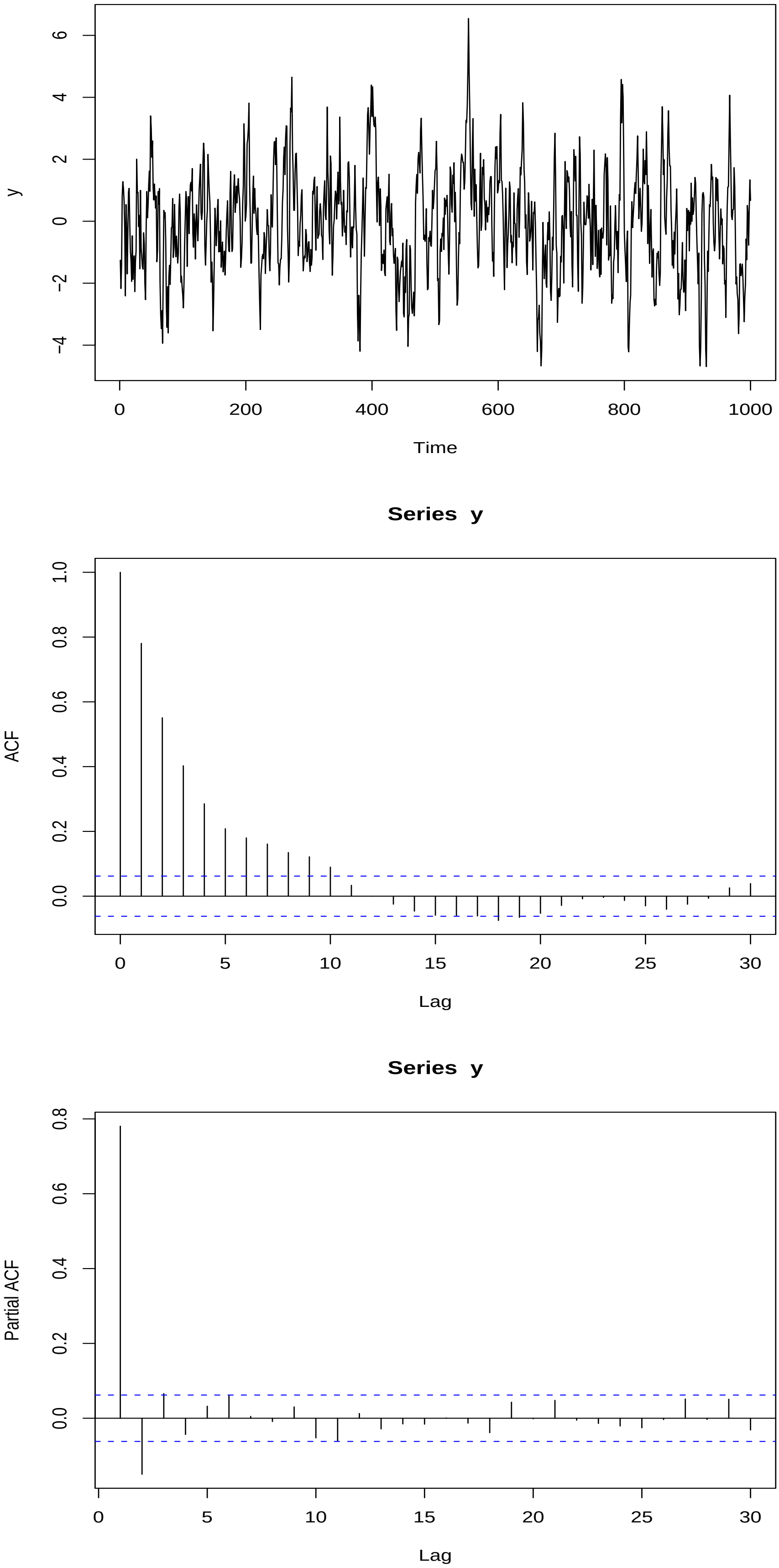

Figure 2.10: The scatter, acf and pacf plots of the simulated ARMA(1) process with $\phi_1 = 0.4$, $\phi_2 = 0.2$ and $\theta_1 = 0.5$, 1000 observations.

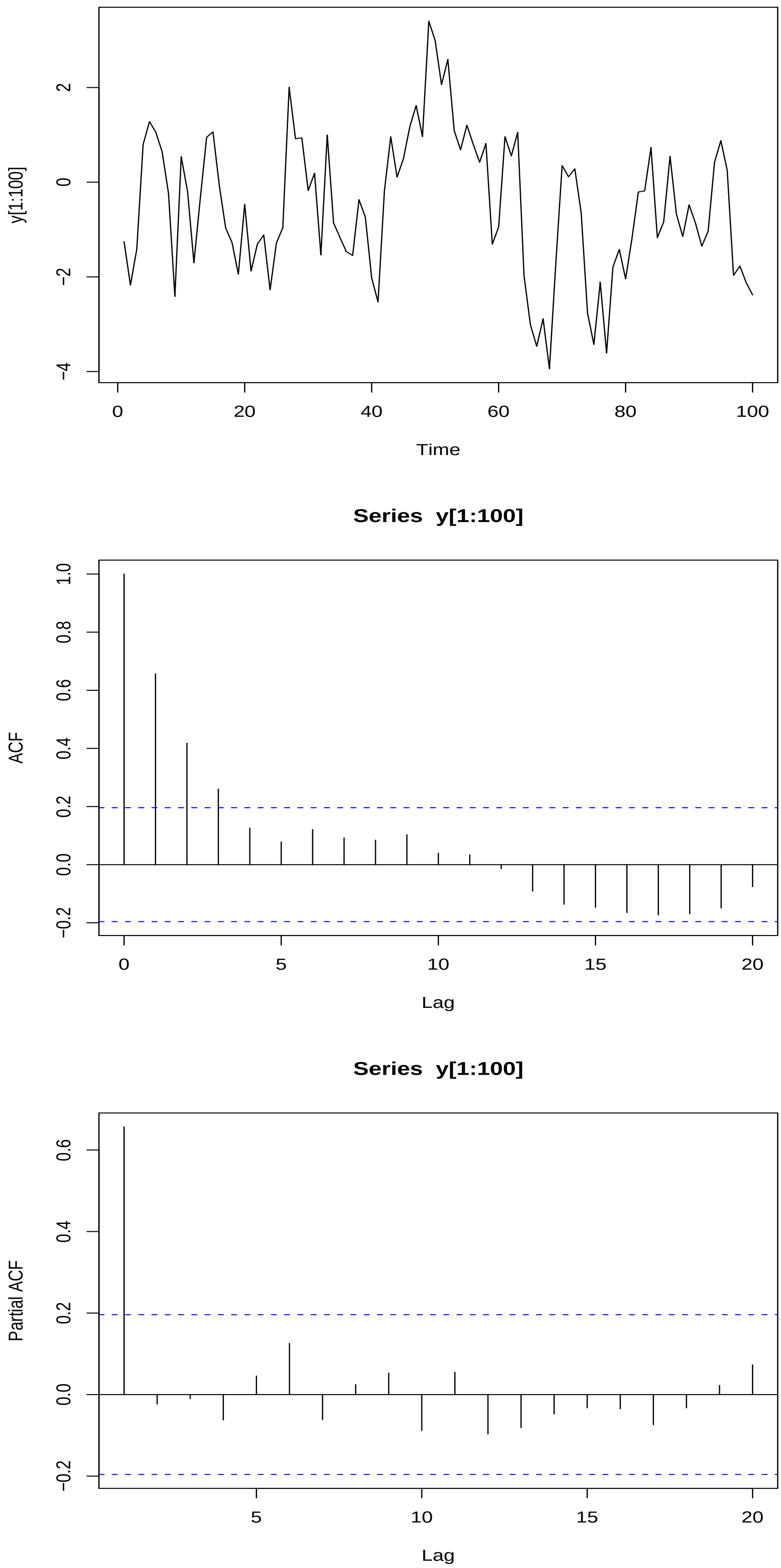

Figure 2.11: The scatter, acf and pacf plots of the subset of 100 observations from the simulated ARMA(2,1) process with $\phi_1 = 0.4$, $\phi_2 = 0.2$ and $\theta_1 = 0.5$.