## 2.3 WHITE NOISE AND LINEAR TIME SERIES

***White Noise.*** A time series $x_t$ is called a *white noise* if $\{x_t\}$ is a sequence of iid random variables with finite mean and variance. In particular, if $x_t$ is normally distributed with mean 0 and variance $\sigma^2$, the series is called a *Gaussian white noise*. For a white noise series, all the ACFs are 0. In practice, if all sample ACFs are close to 0, then the series is a white noise series. On the basis of Figures 2.7 and 2.6b, the monthly returns of IBM stock are close to white noise, whereas those of the Decile 10 portfolio are not.

In the following text, we discuss some simple statistical models that are useful in modeling the dynamic structure of a time series. The concepts presented are also useful later in modeling volatility of asset returns.

***Linear Time Series.*** A time series $x_t$ is said to be linear if it can be written as

$$x_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}, \tag{2.4}$$

where $\mu$ is the mean of $x_t$, $\psi_0 = 1$, and $\{a_t\}$ is a sequence of iid random variables with mean 0 and a well-defined distribution (i.e., $\{a_t\}$ is a white noise series). It will be seen later that $a_t$ denotes the new information at time $t$ of the time series and is often referred to as the *innovation* or *shock* at time $t$. Thus, a time series is linear if it can be written as a linear combination of past innovations. In this book, we are mainly concerned with the case where the innovation $a_t$ is a continuous random variable. Not all financial time series are linear, but linear models can often provide accurate approximations in real applications.

For a linear time series in Equation (2.4), the dynamic structure of $x_t$ is governed by the coefficients $\psi_i$, which are called the $\psi$-*weights* of $x_t$ in the time series literature. If $x_t$ is weakly stationary, we can obtain its mean and variance easily by using properties of $\{a_t\}$ as

$$E(x_t) = \mu, \qquad \mathrm{Var}(x_t) = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i^2, \tag{2.5}$$

where $\sigma_a^2$ is the variance of $a_t$. Because $\mathrm{Var}(x_t) < \infty, \{\psi_i^2\}$ must be a convergent sequence, implying that $\psi_i^2 \to 0$ as $i \to \infty$. Consequently, for a stationary series, impact of the remote shock $a_{t-i}$ on the return $x_t$ vanishes as $i$ increases.

The lag-$\ell$ autocovariance of $x_t$ is

$$\gamma_\ell = \mathrm{Cov}(x_t, x_{t-\ell}) = E\left[\left(\sum_{i=0}^{\infty} \psi_i a_{t-i}\right)\left(\sum_{j=0}^{\infty} \psi_j a_{t-\ell-j}\right)\right]$$

$$= E\left(\sum_{i,j=0}^{\infty} \psi_i \psi_j a_{t-i} a_{t-\ell-j}\right) = \sum_{j=0}^{\infty} \psi_{j+\ell} \psi_j E(a_{t-\ell-j}^2)$$

$$= \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+\ell}. \tag{2.6}$$

Consequently, the $\psi$-weights are related to the autocorrelations of $x_t$ as follows:

$$\rho_\ell = \frac{\gamma_\ell}{\gamma_0} = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+\ell}}{1 + \sum_{i=1}^{\infty} \psi_i^2}, \quad \ell \geq 0, \tag{2.7}$$

where $\psi_0 = 1$. Linear time series models are econometric and statistical models employed to describe the pattern of the $\psi$-weights of $x_t$. For a weakly stationary time series, $\psi_i \to 0$ as $i \to \infty$ and, hence, $\rho_\ell$ converges to 0 as $\ell$ increases. For asset returns, this means that, as expected, the linear dependence of the current return $x_t$ on the remote past return $x_{t-\ell}$ diminishes for large $\ell$.

## 2.4 SIMPLE AUTOREGRESSIVE MODELS

When $x_t$ has a statistically significant lag-1 autocorrelation, the lagged value $x_{t-1}$ might be useful in predicting $x_t$. A simple model that makes use of such predictive power is

$$x_t = \phi_0 + \phi_1 x_{t-1} + a_t, \tag{2.8}$$

where $\{a_t\}$ is assumed to be a white noise series with mean 0 and variance $\sigma_a^2$. This model is in the same form as the well-known simple linear regression model, in which $x_t$ is the dependent variable and $x_{t-1}$ is the explanatory variable. In the time series literature, model (Eq. 2.8) is referred to as an *AR model* of order 1 or simply an AR(1) model. This simple model is also widely used in stochastic volatility modeling when $x_t$ is replaced by its log volatility (Chapter 4).

The AR(1) model in Equation (2.8) has several properties similar to those of the simple linear regression model. However, there are some significant differences between the two models, which we discuss later. Here, it suffices to note that an AR(1) model implies that, conditional on the past return $x_{t-1}$, we have

$$E(x_t|x_{t-1}) = \phi_0 + \phi_1 x_{t-1}, \qquad \text{Var}(x_t|x_{t-1}) = \text{Var}(a_t) = \sigma_a^2.$$

For asset returns, the above results imply that given the past return $x_{t-1}$, the current return is centered around $\phi_0 + \phi_1 x_{t-1}$ with standard deviation $\sigma_a$. This is a Markov property such that conditional on $x_{t-1}$, the return $x_t$ is not correlated with $x_{t-i}$ for $i > 1$. Obviously, there are situations in which $x_{t-1}$ alone cannot determine the conditional expectation of $x_t$ and a more flexible model must be sought. A straightforward generalization of the AR(1) model is the AR($p$) model

$$x_t = \phi_0 + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + a_t, \tag{2.9}$$

where $p$ is a nonnegative integer and $\{a_t\}$ is defined in Equation (2.8). This model says that, given the past data, the first $p$ lagged variables $x_{t-i} (i = 1, \ldots, p)$ jointly determine the conditional expectation of $x_t$. The AR($p$) model is in the same form as a multiple linear regression model with lagged values serving as the explanatory variables.

### 2.4.1 Properties of AR Models

For effective use of AR models, it pays to study their basic properties. We discuss properties of AR(1) and AR(2) models in detail and give the results for the general AR($p$) model.

***AR(1) Model.*** We begin with the sufficient and necessary condition for weak stationarity of the AR(1) model in Equation (2.8). Assuming that the series is weakly stationary, we have $E(x_t) = \mu$, $\mathrm{Var}(x_t) = \gamma_0$, and $\mathrm{Cov}(x_t, x_{t-j}) = \gamma_j$, where $\mu$ and $\gamma_0$ are constants and $\gamma_j$ is a function of $j$, not $t$. We can easily obtain the mean, variance, and autocorrelations of the series as follows. Taking the expectation of Equation (2.8) and using $E(a_t) = 0$, we obtain

$$E(x_t) = \phi_0 + \phi_1 E(x_{t-1}).$$

Under the stationarity condition, $E(x_t) = E(x_{t-1}) = \mu$ and hence

$$\mu = \phi_0 + \phi_1 \mu \quad \text{or} \quad E(x_t) = \mu = \frac{\phi_0}{1 - \phi_1}.$$

This result has two implications for $x_t$. First, the mean of $x_t$ exists if $\phi_1 \neq 1$. Second, the mean of $x_t$ is 0 if and only if $\phi_0 = 0$. Thus, for a stationary AR(1) process, the constant term $\phi_0$ is related to the mean of $x_t$ via $\phi_0 = (1 - \phi_1)\mu$, and $\phi_0 = 0$ implies that $E(x_t) = 0$.

Next, using $\phi_0 = (1 - \phi_1)\mu$, the AR(1) model can be rewritten as

$$x_t - \mu = \phi_1 (x_{t-1} - \mu) + a_t. \tag{2.10}$$

By repeated substitutions, the prior equation implies that

$$\begin{aligned} x_t - \mu &= a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \cdots \\ &= \sum_{i=0}^{\infty} \phi_1^i a_{t-i}. \end{aligned} \tag{2.11}$$

This equation expresses an AR(1) model in the form of Equation (2.4) with $\psi_i = \phi_1^i$. Thus, $x_t - \mu$ is a linear function of $a_{t-i}$ for $i \geq 0$. Using this property and the independence of the series $\{a_t\}$, we obtain $E[(x_t - \mu)a_{t+1}] = 0$. By the stationarity assumption, we have $\text{Cov}(x_{t-1}, a_t) = E[(x_{t-1} - \mu)a_t] = 0$. This latter result can also be seen from the fact that $x_{t-1}$ occurred before time $t$ and $a_t$, being a shock at time $t$, does not depend on any past information. Taking the square and the expectation of Equation (2.10), we obtain

$$\text{Var}(x_t) = \phi_1^2 \text{Var}(x_{t-1}) + \sigma_a^2,$$

where $\sigma_a^2$ is the variance of $a_t$, and we make use of the fact that the covariance between $x_{t-1}$ and $a_t$ is 0. Under the stationarity assumption, $\text{Var}(x_t) = \text{Var}(x_{t-1})$, so that

$$\text{Var}(x_t) = \frac{\sigma_a^2}{1 - \phi_1^2}$$

provided that $\phi_1^2 < 1$. The requirement of $\phi_1^2 < 1$ results from the fact that the variance of a random variable is nonnegative and $x_t$ is weakly stationary. Consequently, the weak stationarity of an AR(1) model implies that $-1 < \phi_1 < 1$, that is, $|\phi_1| < 1$. Yet if $|\phi_1| < 1$, then by Equation (2.11) and the independence of the $\{a_t\}$ series, we can show that the mean and variance of $x_t$ are finite and time invariant; see Equation (2.5). In addition, by Equation (2.6), all the autocovariances of $x_t$ are finite. Therefore, the AR(1) model is weakly stationary. In summary, the necessary and sufficient condition for the AR(1) model in Equation (2.8) to be weakly stationary is $|\phi_1| < 1$.

Using $\phi_0 = (1 - \phi_1)\mu$, one can rewrite a stationary AR(1) model as

$$x_t = (1 - \phi_1)\mu + \phi_1 x_{t-1} + a_t.$$

This model is often used in the finance literature with $\phi_1$ measuring the persistence of the dynamic dependence of an AR(1) time series.

***Autocorrelation Function of an AR(1) Model.*** Multiplying Equation (2.10) by $a_t$, using the independence between $a_t$ and $x_{t-1}$, and taking expectation, we obtain

$$E[a_t(x_t - \mu)] = \phi_1 E[a_t(x_{t-1} - \mu)] + E(a_t^2) = E(a_t^2) = \sigma_a^2,$$

where $\sigma_a^2$ is the variance of $a_t$. Multiplying Equation (2.10) by $(x_{t-\ell} - \mu)$, taking expectation, and using the prior result, we have

$$\gamma_\ell = \begin{cases} \phi_1 \gamma_1 + \sigma_a^2 & \text{if } \ell = 0 \\ \phi_1 \gamma_{\ell-1} & \text{if } \ell > 0, \end{cases}$$

where we use $\gamma_\ell = \gamma_{-\ell}$. Consequently, for a weakly stationary AR(1) model in Equation (2.8), we have

$$\text{Var}(x_t) = \gamma_0 = \frac{\sigma^2}{1 - \phi_1^2} \quad \text{and} \quad \gamma_\ell = \phi_1 \gamma_{\ell-1}, \quad \text{for} \quad \ell > 0.$$

From the latter equation, the ACF of $x_t$ satisfies

$$\rho_\ell = \phi_1 \rho_{\ell-1}, \quad \text{for} \quad \ell > 0.$$

Because $\rho_0 = 1$, we have $\rho_\ell = \phi_1^\ell$. This result says that the ACF of a weakly stationary AR(1) series decays exponentially with rate $\phi_1$ and starting value $\rho_0 = 1$. For a positive $\phi_1$, the plot of ACF of an AR(1) model shows a nice exponential decay. For a negative $\phi_1$, the plot consists of two alternating exponential decays with rate $\phi_1^2$. Figure 2.8 shows the ACF of two AR(1) models with $\phi_1 = 0.8$ and $\phi_1 = -0.8$.

***AR(2) Model.*** An AR(2) model assumes the form

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + a_t. \tag{2.12}$$

Using the same technique as that of the AR(1) case, we obtain

$$E(x_t) = \mu = \frac{\phi_0}{1 - \phi_1 - \phi_2}$$

provided that $\phi_1 + \phi_2 \neq 1$. Using $\phi_0 = (1 - \phi_1 - \phi_2)\mu$, we can rewrite the AR(2) model as

$$(x_t - \mu) = \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + a_t.$$

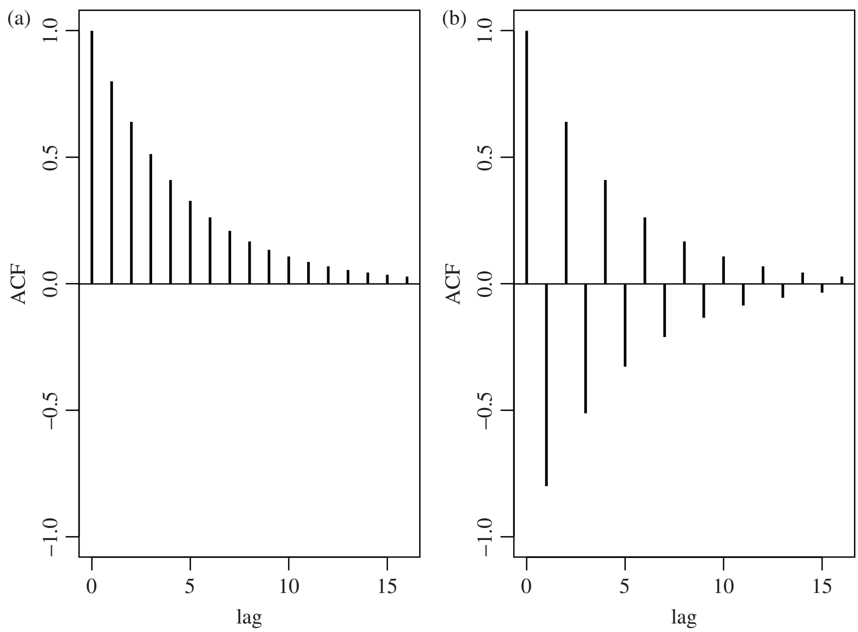


Figure 2.8. The autocorrelation function of an AR(1) model: (a) for $\phi_1 = 0.8$ and (b) for $\phi_1 = -0.8$.

Multiplying the prior equation by $(x_{t-\ell} - \mu)$, we have

$$(x_{t-\ell} - \mu)(x_t - \mu) = \phi_1(x_{t-\ell} - \mu)(x_{t-1} - \mu) + \phi_2(x_{t-\ell} - \mu)(x_{t-2} - \mu) + (x_{t-\ell} - \mu)a_t.$$

Taking expectation and using $E[(x_{t-\ell} - \mu)a_t] = 0$ for $\ell > 0$, we obtain

$$\gamma_\ell = \phi_1\gamma_{\ell-1} + \phi_2\gamma_{\ell-2}, \quad \text{for} \quad \ell > 0.$$

This result is referred to as the *moment equation* of a stationary AR(2) model. Dividing the above equation by $\gamma_0$, we have the property

$$\rho_\ell = \phi_1\rho_{\ell-1} + \phi_2\rho_{\ell-2}, \quad \text{for} \quad \ell > 0, \tag{2.13}$$

for the ACF of $x_t$. In particular, the lag-1 ACF satisfies

$$\rho_1 = \phi_1\rho_0 + \phi_2\rho_{-1} = \phi_1 + \phi_2\rho_1.$$

Therefore, for a stationary AR(2) series $x_t$, we have $\rho_0 = 1$,

$$\rho_1 = \frac{\phi_1}{1 - \phi_2}$$
$$\rho_\ell = \phi_1\rho_{\ell-1} + \phi_2\rho_{\ell-2}, \quad \ell \geq 2.$$

The result of Equation (2.13) says that the ACF of a stationary AR(2) series satisfies the second-order difference equation

$$(1 - \phi_1 B - \phi_2 B^2)\rho_\ell = 0,$$

where $B$ is called the *backshift* operator such that $B\rho_\ell = \rho_{\ell-1}$. This difference equation determines the properties of the ACF of a stationary AR(2) time series. It also determines the behavior of the forecasts of $x_t$. In the time series literature, some people use the notation $L$ instead of $B$ for the backshift operator. Here, $L$ stands for *lag* operator. For instance, $Lx_t = x_{t-1}$ and $L\psi_k = \psi_{k-1}$.

Corresponding to the prior difference equation, there is a second-order polynomial equation

$$1 - \phi_1 z - \phi_2 z^2 = 0. \tag{2.14}$$

Solutions of this equation are

$$z = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}.$$

In the time series literature, inverses of the two solutions are referred to as the *characteristic roots* of the AR(2) model. Denote the two characteristic roots by $\omega_1$ and $\omega_2$. If both $\omega_i$ are real valued, then the second-order difference equation of the model can be factored as $(1 - \omega_1 B)(1 - \omega_2 B)$, and the AR(2) model can be regarded as an AR(1) model operates on top of another AR(1) model. The ACF of $x_t$ is then a mixture of two exponential decays. If $\phi_1^2 + 4\phi_2 < 0$, then $\omega_1$ and $\omega_2$ are complex numbers (called a *complex conjugate pair*), and the plot of ACF of $x_t$ would show a picture of damping sine and cosine waves. In business and economic applications, complex characteristic roots are important. They give rise to the behavior of business cycles. It is then common for economic time series models to have complex-valued characteristic roots. For an AR(2) model in Equation (2.12) with a pair of complex characteristic roots, the *average* length of the stochastic cycles is

$$k = \frac{2\pi}{\cos^{-1}[\phi_1/(2\sqrt{-\phi_2})]},$$

where the cosine inverse is stated in radian. If one writes the complex solutions as $a \pm bi$, where $i = \sqrt{-1}$, then we have $\phi_1 = 2a, \phi_2 = -(a^2 + b^2)$, and

$$k = \frac{2\pi}{\cos^{-1}(a/\sqrt{a^2 + b^2})},$$

where $\sqrt{a^2 + b^2}$ is the absolute value of $a \pm bi$. See Example 2.3 for an illustration.

Figure 2.9 shows the ACF of four stationary AR(2) models. Part (b) is the ACF of the AR(2) model $(1 - 0.6B + 0.4B^2)x_t = a_t$. Because $\phi_1^2 + 4\phi_2 = 0.36 + 4 \times (-0.4) = -1.24 < 0$, this particular AR(2) model contains two complex characteristic roots, and hence its ACF exhibits damping sine and cosine waves. The other three AR(2) models have real-valued characteristic roots. Their ACFs decay exponentially.

**Example 2.3.** As an illustration, consider the quarterly growth rate of US gross national product (GNP), seasonally adjusted, from the second quarter of 1947 to the first quarter of 2010 for 252 observations. The log series of GNP, in billions of dollars, and its growth rate are shown in Figure 2.10. A horizontal line of zero is added to the time plot of the growth rate. The plot clearly shows that most of the growth rates are positive and the largest drop in GNP occurred in the 2008 recession.

On the basis of the model building procedure of the next section, we employ an AR(3) model for the data. The fitted model is

$$(1 - 0.438B - 0.206B^2 + 0.156B^3)(x_t - 0.016) = a_t, \quad \hat{\sigma}_a = 9.55 \times 10^{-5}. \quad (2.15)$$

The standard errors of the estimates are 0.062, 0.067, 0.063, and 0.001, respectively. See the attached R output for further information. Model (2.15) gives rise to a third-order polynomial equation

$$1 - 0.438z - 0.206z^2 + 0.156z^3 = 0,$$

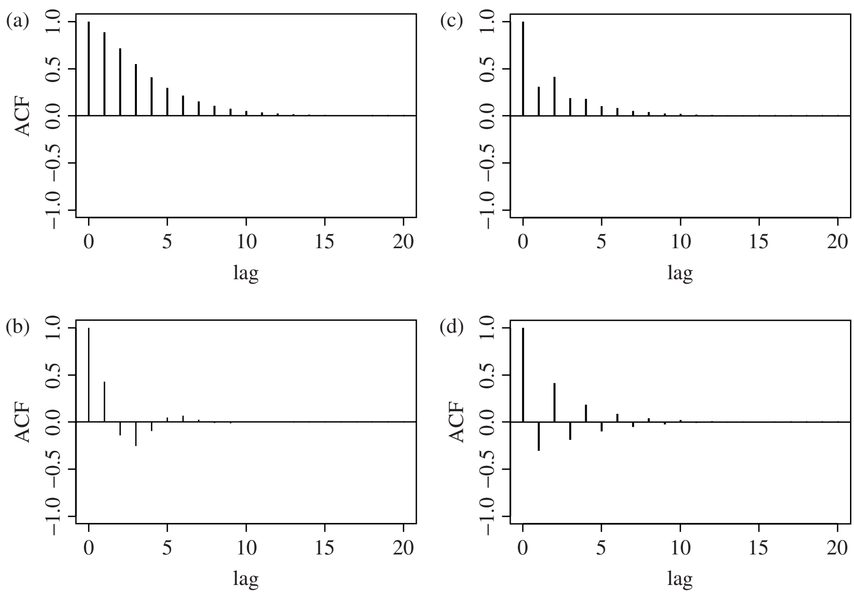


Figure 2.9. The autocorrelation function of an AR(2) model: (a) $\phi_1 = 1.2$ and $\phi_2 = -0.35$, (b) $\phi_1 = 0.6$ and $\phi_2 = -0.4$, (c) $\phi_1 = 0.2$ and $\phi_2 = 0.35$, and (d) $\phi_1 = -0.2$ and $\phi_2 = 0.35$.

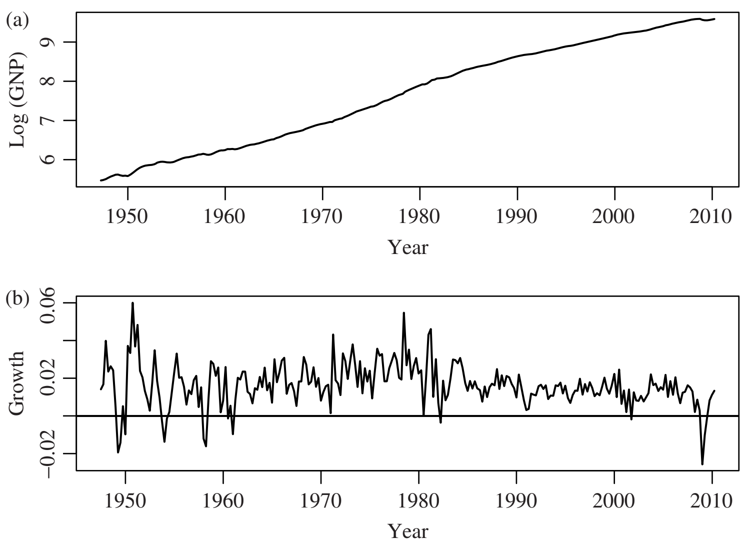


Figure 2.10. Time plots of US quarterly gross national product from 1947.I to 2010.I: (a) Log GNP series and (b) growth rate. The data are seasonally adjusted and in billions of dollars.

which has three solutions, namely, $1.616 + 0.864i, 1.616 - 0.864i$, and $-1.909$. The real solution corresponds to a factor $[1 - (1/-1.909)z] = (1 + 0.524z)$ that shows an exponentially decaying feature of the GNP growth rate. Focusing on the complex conjugate pair $1.616 \pm 0.864i$, we obtain the absolute value $\sqrt{1.616^2 + 0.864^2} = 1.833$ and

$$k = \frac{2\pi}{\cos^{-1}(1.616/1.833)} \approx 12.80.$$

Therefore, the fitted AR(3) model confirms the existence of business cycles in the US economy, and the average length of the cycles is 12.8 quarters, which is about 3 years. This result is reasonable as the US economy went through expansion and contraction and the length of expansion is generally believed to be around 3 years. If one uses a nonlinear model to separate US economy into "expansion" and "contraction" periods, the data show that the average duration of contraction periods is about three quarters and that of expansion periods is about 3 years; see, for instance, the analysis in Tsay (2010, Chapter 4). The average duration of 12.8 quarters is a compromise between the two separate durations. The periodic feature obtained here is common among growth rates of national economies. For example, similar features can be found for many economies in the Organization for Economic Cooperation and Development (OECD) countries.

```
> da=read.table("q-gnp4710.txt",header=T)
> head(da)
  Year Mon Day VALUE
1 1947   1   1 238.1
 ...
6 1948   4   1 268.7
> G=da$VALUE
> LG=log(G)
> gnp=diff(LG)
> dim(da)
[1] 253   4
> tdx=c(1:253)/4+1947 # create the time index
> par(mfcol=c(2,1))
> plot(tdx,LG,xlab='year',ylab='GNP',type='l')
> plot(tdx[2:253],gnp,type='l',xlab='year',ylab='growth')
> acf(gnp,lag=12)
> pacf(gnp,lag=12) # compute PACF
> m1=arima(gnp,order=c(3,0,0))
> m1
Call:
arima(x = gnp, order = c(3, 0, 0))

Coefficients:
         ar1     ar2      ar3  intercept
      0.4386  0.2063  -0.1559     0.0163
s.e.  0.0620  0.0666   0.0626     0.0012
```

```
sigma^2 estimated as 9.549e-05:log likelihood=808.6,aic=-1607.1
> tsdiag(m1,gof=12)  # model checking discussed later
> p1=c(1,-m1$coef[1:3]) # set-up the polynomial
> r1=polyroot(p1) # solve the polynomial equation
> r1
[1]  1.616116+0.864212i -1.909216-0.000000i  1.616116-0.864212i
> Mod(r1)
[1] 1.832674 1.909216 1.832674  # compute absolute values
> k=2*pi/acos(1.616116/1.832674) # compute length of the period
> k
[1] 12.79523
```

□

***Stationarity.*** The stationarity condition of an AR(2) time series is that the absolute values of its two characteristic roots are less than 1, that is, its two characteristic roots are less than 1 in modulus. Equivalently, the two solutions of the characteristic equation are greater than 1 in modulus. Under such a condition, the recursive equation in Equation (2.13) ensures that the ACF of the model converges to 0 as the lag $\ell$ increases. This convergence property is a necessary condition for a stationary time series. In fact, the condition also applies to the AR(1) model, where the polynomial equation is $1 - \phi_1 z = 0$. The characteristic root is $w = 1/z = \phi_1$, which must be less than 1 in modulus for $x_t$ to be stationary. As shown before, $\rho_\ell = \phi_1^\ell$ for a stationary AR(1) model. The condition implies that $\rho_\ell \to 0$ as $\ell \to \infty$.

***AR(p) Model.*** The results of AR(1) and AR(2) models can readily be generalized to the general AR($p$) model in Equation (2.9). The mean of a stationary series is

$$E(x_t) = \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p}$$

provided that the denominator is not 0. The associated characteristic equation of the model is

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0.$$

If all the solutions of this equation are greater than 1 in modulus, then the series $x_t$ is stationary. Again, inverses of the solutions are the *characteristic roots* of the model. Thus, stationarity requires that all characteristic roots are less than 1 in modulus. For a stationary AR($p$) series, the ACF satisfies the difference equation

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)\rho_\ell = 0, \quad \text{for} \quad \ell > 0.$$

The plot of ACF of a stationary AR($p$) model would then show a mixture of damping sine and cosine patterns and exponential decays depending on the nature of its characteristic roots.

### 2.5.1 Properties of MA Models

Again, we focus on the simple MA(1) and MA(2) models. The results of MA($q$) models can easily be obtained by the same techniques.

***Stationarity.*** MA models are always weakly stationary because they are finite linear combinations of a white noise sequence for which the first two moments are time invariant. For example, consider the MA(1) model in Equation (2.20). Taking expectation of the model, we have

$$E(x_t) = c_0,$$

which is time invariant. Taking the variance of Equation (2.20), we have

$$\text{Var}(x_t) = \sigma_a^2 + \theta_1^2\sigma_a^2 = (1 + \theta_1^2)\sigma_a^2,$$

where we use the fact that $a_t$ and $a_{t-1}$ are uncorrelated. Again, Var($x_t$) is time invariant. The prior discussion applies to general MA($q$) models, and we obtain two general properties. First, the constant term of an MA model is the mean of the series (i.e., $E(x_t) = c_0$). Second, the variance of an MA($q$) model is

$$\text{Var}(x_t) = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)\sigma_a^2.$$

***Autocorrelation Function.*** Assume for simplicity that $c_0 = 0$ for an MA(1) model. Multiplying the model by $x_{t-\ell}$, we have

$$x_{t-\ell}x_t = x_{t-\ell}a_t - \theta_1 x_{t-\ell}a_{t-1}.$$

Taking expectation, we obtain

$$\gamma_1 = -\theta_1\sigma_a^2 \quad \text{and} \quad \gamma_\ell = 0, \quad \text{for} \quad \ell > 1.$$

Using the prior result and the fact that Var($x_t$) $= (1 + \theta_1^2)\sigma_a^2$, we have

$$\rho_0 = 1, \quad \rho_1 = \frac{-\theta_1}{1 + \theta_1^2}, \quad \text{and} \quad \rho_\ell = 0, \quad \text{for} \quad \ell > 1.$$

Thus, for an MA(1) model, the lag-1 ACF is not 0, but all higher-order ACFs are 0. In other words, the ACF of an MA(1) model cuts off at lag 1. For the MA(2) model in Equation (2.21), the autocorrelation coefficients are

$$\rho_1 = \frac{-\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}, \quad \rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}, \quad \text{and} \quad \rho_\ell = 0, \quad \text{for} \quad \ell > 2. \tag{2.23}$$

Here, the ACF cuts off at lag 2. This property generalizes to other MA models. For an MA($q$) model, the lag-$q$ ACF is not 0, but $\rho_\ell = 0$ for $\ell > q$. Consequently, an MA($q$) series is only linearly related to its first $q$ lagged values and hence is a "finite memory" model.

***Invertibility.*** Rewriting a zero-mean MA(1) model as $a_t = x_t + \theta_1 a_{t-1}$, one can use repeated substitutions to obtain

$$a_t = x_t + \theta_1 x_{t-1} + \theta_1^2 x_{t-2} + \theta_1^3 x_{t-3} + \cdots.$$

This equation expresses the current shock $a_t$ as a linear combination of the present and past values of $x_t$. Intuitively, $\theta_1^j$ should go to 0 as $j$ increases because the remote return $x_{t-j}$ should have very little impact on the current shock, if any. Consequently, for an MA(1) model to be plausible, we require $|\theta_1| < 1$. Such an MA(1) model is said to be *invertible*. If $|\theta_1| = 1$, then the MA(1) model is noninvertible. See Tsay (2010, Chapter 2) for further discussion on invertibility.

## 2.5.2 Identifying MA Order

The ACF is useful in identifying the order of an MA model. For a time series $x_t$ with ACF $\rho_\ell$, if $\rho_q \neq 0$, but $\rho_\ell = 0$ for $\ell > q$, then $x_t$ follows an MA($q$) model.

Figure 2.14 shows the time plot of monthly simple returns of the CRSP equal-weighted index from January 1926 to December 2008 and the sample ACF of the series. The two dashed lines shown on the ACF plot denote the two standard error

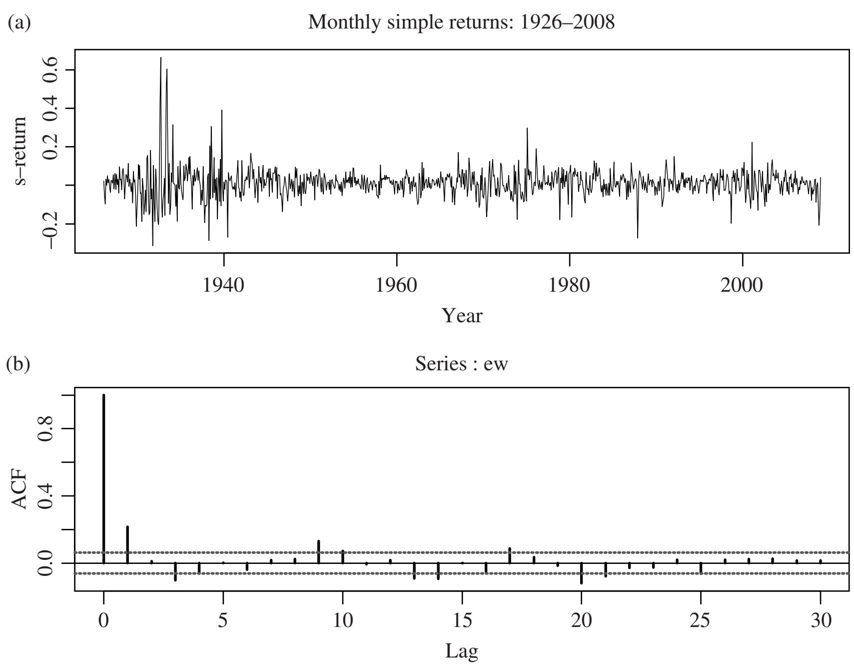


Figure 2.14. Time plot and sample autocorrelation function of monthly simple returns of the CRSP equal-weighted index from January 1926 to December 2008.

limits. It is seen that the series has significant ACF at lags 1, 3, and 9. There are some marginally significant ACF at higher lags, but we do not consider them here. On the basis of the sample ACF, the following MA(9) model

$$x_t = c_0 + a_t - \theta_1 a_{t-1} - \theta_3 a_{t-3} - \theta_9 a_{t-9}$$

is identified for the series. Note that, unlike the sample PACF, sample ACF provides information on the nonzero MA lags of the model. To see this, consider, for example, a simple MA(2) model with $\theta_1 = 0$. The model is $x_t = c_0 + a_t - \theta_2 a_{t-2}$. Using Equation (2.23) or via direct evaluation, the ACF of the model is

$$\rho_0 = 1, \quad \rho_1 = 0, \quad \rho_2 = \frac{-\theta_2}{1+\theta_2^2}, \quad \text{and} \quad \rho_j = 0 \quad \text{for} \quad j > 2.$$

Therefore, for this particular case, ACF provides the exact information on the structure of the model.

### 2.5.3 Estimation

Maximum likelihood estimation is commonly used to estimate MA models. There are two approaches for evaluating the likelihood function of an MA model. The first approach assumes that the initial shocks (i.e., $a_t$ for $t \le 0$) are 0. As such, the shocks needed in likelihood function calculation are obtained recursively from the model, starting with $a_1 = x_1 - c_0$ and $a_2 = x_2 - c_0 + \theta_1 a_1$. This approach is referred to as the *conditional likelihood method* and the resulting estimates the conditional maximum likelihood estimates. The second approach treats the initial shocks $a_t, t \le 0$ as additional parameters of the model and estimate them jointly with other parameters. This approach is referred to as the *exact likelihood method*. The exact likelihood estimates are preferred over the conditional ones, especially when the MA model is close to being noninvertible. The exact method, however, requires more intensive computation. If the sample size is large, then the two types of maximum likelihood estimates are close to each other. For details of conditional and exact likelihood estimates of MA models, readers are referred to Box et al. (1994) or Tsay (2010, Chapter 8).

For illustration, consider the monthly simple return series of the CRSP equal-weighted index and the specified MA(9) model. The conditional maximum likelihood method produces the fitted model

$$x_t = 0.012 + a_t + 0.189 a_{t-1} - 0.121 a_{t-3} + 0.122 a_{t-9}, \quad \hat{\sigma}_a = 0.0714, \qquad (2.24)$$

where standard errors of the coefficient estimates are 0.003, 0.031, 0.031, and 0.031, respectively. The Ljung–Box statistics of the residuals give $Q(12) = 17.5$ with $p$-value 0.041, which is based on an asymptotic chi-squared distribution with 9 degrees of freedom. The model needs some refinements in modeling the linear dynamic