CHAPTER

22 Time Series Analysis

William W.S. Wei

Abstract

This chapter deals with time domain statistical models and methods on analyzing time series and their use in applications. It covers fundamental concepts, stationary and nonstationary models, nonseasonal and seasonal models, intervention and outlier models, transfer function models, regression time series models, vector time series models, and their applications. We discuss the process of time series analysis including model identification, parameter estimation, diagnostic checks, forecasting, and inference. We also discuss autoregressive conditional heteroscedasticity model, generalized autoregressive conditional heteroscedasticity model, and unit roots and cointegration in vector time series processes.

Key Words: Autoregressive model, moving average model, autoregressive moving average model, autoregressive integrated moving average model, intervention, outlier, transfer function model, autoregressive conditional heteroscedasticity model, generalized autoregressive conditional heteroscedasticity model, vector autoregressive model, vector moving average model, vector autoregressive moving average model

Introduction

In studying a phenomenon, we often encounter a data set where the observations are taken according to the order of time. This time-ordered sequence of observations is called a *time series*. Examples of such data sets are numerous, such as daily closing stock prices, monthly unemployment figures, quarterly crime rates, and annual birth rates. The fundamental characteristic of a time series is that its observations are correlated. Most standard statistical methods based on random samples are not applicable, and different methods are needed. The body of statistical methods for analyzing time series is referred to as *time series analysis*. Some of these methods are descriptive, emphasizing mainly the description of a time series based on non-stochastic methods. The other approach—that is, the stochastic approach—is to treat a time series as a realization of a stochastic time series process or model, and the main purpose of analyzing time series in this approach is to construct a possible underlying process and use it for forecasting, inference, and control. It is on this approach that we will focus our attention in the following discussion.

Time series analysis includes time domain approach and frequency domain approach. In the time domain approach, we use time functions like the *autocorrelation function* (*ACF*) and the *partial autocorrelation function* (*PACF*) to describe the characteristics of a time series process whose evolution is represented through various time-lag relationships. In the frequency domain approach, we try to use a spectral function to study how the variation of a time series may be accounted for by the mixture of sines and cosines at various frequencies. Because of space restrictions, we will concentrate our discussion

on the time domain approach. For the frequency domain approach, we refer readers to the introductory chapter, Spectral Analysis, given by Wei (2008).

After introducing some fundamental concepts, we will start with univariate time series and introduce some commonly used stationary and nonstationary time series models, including seasonal time series models. We will describe a systematic model-building process that has been found useful in constructing a time series model from a given time series data set. We then extend the method to study the relationship of several time series variables. Examples will be used throughout the discussion to illustrate the concepts and procedures.

Some Fundamental Concepts

Strictly and Weakly Stationary Processes

A time series is a realization of a *stochastic process*, which is a family of time-indexed random variables. Let us use Z_t to denote a time series process, where for convenience and with no loss of generality we assume that the time index set is the set of all integers. The process is characterized by the joint probability distribution of these variables. We call the process *strictly stationary* if its joint distribution is invariant with respect to a change of time origin. That is, for a strictly stationary process, we have

$$F_{Z_{t_1},\ldots,Z_{t_n}}(x_1,\ldots,x_n) = F_{Z_{t_1+k},\ldots,Z_{t_n+k}}(x_1,\ldots,x_n) \quad (1)$$

for any n-tuple $(t_1,\ldots,t_n)$ and any k of integers, where $F_{Z_{t_1},\ldots,Z_{t_n}}(x_1,\ldots,x_n)$ is the joint distribution function defined by $F_{Z_{t_1},\ldots,Z_{t_n}}(x_1,\ldots,x_n) = P\{Z_{t_1} \le x_1,\ldots,Z_{t_n} \le x_n\}$. The terms *strongly stationary* and *completely stationary* are also used to denote a strictly stationary process. Unfortunately this assumption is very difficult or impossible to check. For most practical purpose, to identify the underlying model, it is often sufficient to know the first few moments of the time series process. Thus, we will consider the concept of a weakly stationary process.

A process is said to be *second order* or *weakly stationary* if its first two moments are time invariant. That is, if the *mean function* of the process,

$$\mu_t = E(Z_t) = \mu, \quad (2)$$

and the *variance function* of the process,

$$\sigma_t^2 = Var(Z_t) = E(Z_t-\mu)^2 = \sigma^2, \quad (3)$$

are constant, and the *covariance function* between Z_s and Z_t,

$$\gamma(s,t) = Cov(Z_s,Z_t) = E(Z_s-\mu)(Z_t-\mu), \quad (4)$$

is only a function of the time difference, $(t-s)$. Thus, we can simply write the covariance function between Z_t and Z_{t+k} of a weakly stationary process as

$$\gamma(t,t+k) = E(Z_t-\mu)(Z_{t+k}-\mu) = \gamma_k. \quad (5)$$

Because, in practice, it is this class of second-order weakly stationary processes with which we often work, henceforth, when we say that the process is stationary it is understood that we are referring to a second-order weakly stationary process.

The Autocorrelation Function

In time series analysis, the covariance function at lag k is often called the *autocovariance function* at lag k because it represents the covariance between Z_t and Z_{t+k} from the same process, separated by k time lags. Hence, the ACF between Z_t and Z_{t+k} is simply the standardized autocovariance function,

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{Cov(Z_t,Z_{t+k})}{\sqrt{Var(Z_t)}\sqrt{Var(Z_{t+k})}}, \quad (6)$$

where we note that for a stationary process, $Var(Z_t) = Var(Z_{t+k}) = \gamma_0$. It is easy to see that for a stationary process, the ACF has the properties: (1) $\rho_0 = 1$; (2) $|\rho_k| \le 1$; (3) $\rho_{-k} = \rho_k$, which follows from the fact that the time difference between Z_t and Z_{t+k} and between Z_t and Z_{t-k} are the same; and (4) ρ_k is positive semidefinite—that is, $\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\rho_{|t_i-t_j|} \ge 0$, for any set of time-points, $t_1,\ldots,$ and t_n and any real numbers $\alpha_1,\ldots,$ and α_n. Using property 3, we plot an ACF only for the non-negative lags, which is also called a *correlogram*.

The Partial Autocorrelation Function

Other than the autocorrelation between Z_t and Z_{t+k}, we may also want to study the correlation between Z_t and Z_{t+k} after their mutual linear dependency on the intervening variables $Z_{t+1},\ldots,$ and Z_{t+k-1} has been removed. This conditional correlation,

$$\varphi_{kk} = Corr\,(Z_t, Z_{t+k} \,|\, Z_{t+1},\ldots,Z_{t+k-1}), \quad (7)$$

is referred to as the *partial autocorrelation*, and it equals

$$\varphi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & 1 \end{vmatrix}}. \tag{8}$$

White Noise and Gaussian Processes

A process is called a *white noise process*, to be denoted by $\{a_t\}$, if it is a sequence of uncorrelated random variables from a fixed distribution with a constant mean, usually assumed to be 0, and constant variance. Thus it has the ACF

$$\rho_k = \begin{cases} 1, & k = 0, \\ 0, & k \neq 0, \end{cases} \tag{9}$$

and the PACF

$$\varphi_{kk} = \begin{cases} 1, & k = 0, \\ 0, & k \neq 0. \end{cases} . \tag{10}$$

This process plays an important role as a basic building block in the construction of time series models.

A time series process is said to be a *Gaussian* or *normal process* if its joint distribution is normal. Like most other areas in statistics, most results in time series analysis are established for Gaussian processes. Because a normal distribution is uniquely characterized by its first two moments, strictly stationary and weakly stationary are equivalent for a Gaussian process. As a result, mean, variance, ACF, and PACF also become fundamental measures used in the identification of time series models.

Estimation of the Mean, the Variance, the Autocorrelation Function, and the Partial Autocorrelation Function

Given a time series, $Z_1, \ldots, Z_n$, of n observations from a stationary process, we will use the *sample mean*,

$$\overline{Z} = \frac{1}{n}\sum_{t=1}^{n} Z_t, \tag{11}$$

to estimate the mean, μ, the *sample variance*,

$$\hat{\gamma}_0 = \frac{1}{n}\sum_{t=1}^{n}\left(Z_t - \overline{Z}\right)^2, \tag{12}$$

to estimate the variance, $\gamma_0 = \sigma_Z^2$, and the *sample autocovariance function*,

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}\left(Z_t - \overline{Z}\right)\left(Z_{t+k} - \overline{Z}\right), \tag{13}$$

to estimate the autocovariance function, γ_k.

Similarly, we will use the *sample ACF*,

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k}\left(Z_t - \overline{Z}\right)\left(Z_{t+k} - \overline{Z}\right)}{\sum_{t=1}^{n}\left(Z_t - \overline{Z}\right)^2}, \tag{14}$$

and the *sample PACF*,

$$\hat{\varphi}_{kk} = \frac{\begin{vmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{k-2} & \hat{\rho}_1 \\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{k-3} & \hat{\rho}_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \cdots & \hat{\rho}_1 & \hat{\rho}_k \end{vmatrix}}{\begin{vmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{k-2} & \hat{\rho}_{k-1} \\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{k-3} & \hat{\rho}_{k-2} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \cdots & \hat{\rho}_1 & 1 \end{vmatrix}}, \tag{15}$$

to estimate the ACF, ρ_k, and the PACF, φ_{kk}, respectively.

For a stationary Gaussian process that has the autocorrelations $\rho_k = 0$ for $k > m$, the large-lag standard error of $\hat{\rho}_k$ from Bartlett (1946) is

$$S_{\hat{\rho}_k} = \sqrt{\frac{1}{n}(1 + 2\hat{\rho}_1^2 + \cdots + 2\hat{\rho}_m^2)}. \tag{16}$$

Thus, to test a white noise process from a sample series of n observations, we use

$$S_{\hat{\rho}_k} = \sqrt{\frac{1}{n}}, \tag{17}$$

and

$$S_{\hat{\varphi}_{kk}} = \sqrt{\frac{1}{n}}. \tag{18}$$

Moving Average and Autoregressive Representations of Time Series Processes

In time series analysis, we often write a process Z_t as a linear combination of a sequence of white noise random variables, known as the *moving average (MA) representation*,

$$Z_t = \mu + \sum_{j=0}^{\infty}\psi_j a_{t-j}, \tag{19}$$

where $\psi_0 = 1$. Defining the backshift operator $B^j a_t = a_{t-j}$, we can write Equation 19 as

$$Z_t = \mu + \psi(B)a_t, \tag{20}$$

where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$. For a stationary process, we require that the ψ_j is absolutely summable—that is, $\sum_{j=0}^{\infty} |\psi_j| < \infty$, and μ is the mean of the process.

Another useful form known as *autoregressive (AR) representation* is to regress the value of Z_t on its past values plus a random shock,

$$Z_t = \theta_0 + \sum_{j=0}^{\infty} \pi_j Z_{t-j} + a_t, \tag{21}$$

or

$$\pi(B)Z_t = \theta_0 + a_t, \tag{22}$$

where $\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j$. We call the process Z_t invertible if it can be written in Equation 21 such that the π_j is absolutely summable—that is, $\sum_{j=1}^{\infty} |\pi_j| < \infty$.

Univariate Time Series Models

Although the AR and MA representations are useful, they are not the model forms that we will construct from a given time series of n observations because they contain an infinite number of parameters, which cannot be estimated from a finite number of available observations. We need to consider models with a finite number of parameters. For a given n observations, it is known that the more parameters in a model, the less efficient the estimation of the parameters will be. Thus, an important rule in model construction is the principle of parsimony, where all other things being equal, we will, in general, choose a simpler model.

Stationary Time Series Models

In the AR representation, if only a finite number of π_j weights are non-zero—that is, $\pi_1 = \varphi_1, \ldots, \pi_p = \varphi_p$, and $\pi_j = 0$, for $j > p$, the resulting model is said to be an *AR model* (*process*) *of order p*, to be denoted as $AR(p)$,

$$Z_t = \theta_0 + \varphi_1 Z_{t-1} + \cdots + \varphi_p Z_{t-p} + a_t, \tag{23}$$

or

$$\varphi_p(B)Z_t = \theta_0 + a_t, \tag{24}$$

where $\varphi_p(B) = (1 - \varphi_1 B - \cdots - \varphi_p B^p)$. Because $\sum_{j=1}^{\infty} |\pi_j| = \sum_{j=1}^{p} |\varphi_j| < \infty$, the process is always invertible. To be stationary, the roots of $\varphi_p(B) = 0$ must be outside of the unit circle so that its moving average representation in Equation 19 exists. The $\varphi_j, j = 1, \ldots, p$, are often referred to as the AR coefficients or parameters.

For $p = 1$, we have the first-order $AR(1)$ model,

$$Z_t = \theta_0 + \varphi_1 Z_{t-1} + a_t, \tag{25}$$

or

$$(1 - \varphi_1 B)Z_t = \theta_0 + a_t. \tag{26}$$

For a stationary $AR(1)$ process, the root of $(1 - \varphi_1 B) = 0$ must be outside of the unit circle and hence $|\varphi_1| < 1$. It can be easily seen that the mean μ of the process is related to the constant term θ_0 by $\theta_0 = (1 - \varphi_1)\mu$. Let $\dot{Z}_t = Z_t - \mu$. Equation 25 becomes $\dot{Z}_t = \varphi_1 \dot{Z}_{t-1} + a_t$, and we have

$$\gamma_k = E(\dot{Z}_{t-k}\dot{Z}_t) = \varphi_1 E(\dot{Z}_{t-k}\dot{Z}_{t-1}) + E(\dot{Z}_{t-k}a_t)$$
$$= \varphi_1 \gamma_{k-1},\ k \geq 1,$$

and

$$\rho_k = \varphi_1 \rho_{k-1} = \varphi_1^k,\ k \geq 1, \tag{27}$$

where we note that $\rho_0 = 1$. The PACF of the process from Equation 8 is

$$\varphi_{kk} = \begin{cases} \rho_1 = \varphi_1, & k = 1, \\ 0, & \text{for } k \geq 2. \end{cases} \tag{28}$$

Thus, for a stationary $AR(1)$ model, its ACF decays exponentially and its PACF cuts off after lag 1. More generally, for a stationary $AR(p)$ model, its ACF decays exponentially and its PACF cuts off after lag p. The fundamental characteristic of a stationary $AR(p)$ is that its PACF cuts off after lag p.

In the MA representation, if only a finite number of ψ_j weights are non-zero (i.e., $\psi_1 = \theta_1, \ldots, \psi_q = \theta_q$, and $\psi_j = 0$, for $j > q$) then the resulting model is said to be a *moving average model* (*process*) *of order q,* to be denoted as $MA(q)$,

$$Z_t = \mu + a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}, \tag{29}$$

or

$$\dot{Z}_t = \theta_q(B)a_t, \tag{30}$$

where $\theta_q(B) = (1 - \theta_1 B - \cdots - \theta_q B^q)$. Because $\sum_{j=0}^{\infty} |\psi_j| = 1 + \sum_{j=1}^{q} |\theta_j| < \infty$, the process is always stationary. To be invertible, the roots of $\theta_q(B) = 0$ must be outside of the unit circle so that its AR representation in Equation 21 exists. The $\theta_j, j = 1, \ldots, q$, are often referred to as the MA coefficients or parameters.

For $q = 1$, we have the first-order moving average $MA(1)$ model,

$$Z_t = \mu + a_t - \theta_1 a_{t-1}, \tag{31}$$

or

$$\dot{Z}_t = (1 - \theta_1 B)a_t. \tag{32}$$

For a stationary $MA(1)$ process, the root of $(1 - \theta_1 B) = 0$ must be outside of the unit circle and hence $|\theta_1| < 1$. It can be easily seen that

$$\rho_k = \begin{cases} \dfrac{-\theta_1}{1 + \theta_1^2}, & k = 1, \\ 0, & k > 1. \end{cases} \tag{33}$$

The PACF of the process from Equation 8 is

$$\varphi_{kk} = \frac{-\theta_1^k(1 - \theta_1^2)}{1 - \theta_1^{2(k+1)}}, \text{ for } k \geq 1. \tag{34}$$

Thus, for a stationary $MA(1)$ model, its ACF cuts off after lag 1 and its PACF decays exponentially. More generally, for a stationary $MA(q)$ model, its ACF cuts off after lag q and its PACF decays exponentially. The fundamental characteristic of a stationary $MA(q)$ is that its ACF cuts off after lag q.

Naturally, a model may contain both AR and MA parameters, and we have the $ARMA(p, q)$ *models*,

$$\varphi_p(B)Z_t = \theta_0 + \theta_q(B)a_t, \tag{35}$$

where

$$\varphi_p(B) = (1 - \varphi_1 B - \cdots - \varphi_p B^p),$$

and

$$\theta_q(B) = (1 - \theta_1 B - \cdots - \theta_q B^q).$$

For the process to be stationary, the roots of $\varphi_p(B) = 0$ must be outside of the unit circle. To be invertible, the roots of $\theta_q(B) = 0$ must be outside of the unit circle.

For $p = 1$ and $q = 1$, we have the $ARMA(1, 1)$ model,

$$Z_t = \theta_0 + \varphi_1 Z_{t-1} + a_t - \theta_1 a_{t-1}, \tag{36}$$

or

$$(1 - \varphi_1 B)Z_t = \theta_0 + (1 - \theta_1 B)a_t. \tag{37}$$

For stationarity, we require that $|\varphi_1| < 1$, and for invertibility, we require that $|\theta_1| < 1$. Note that the mean μ of Z_t and θ_0 are related by $\theta_0 = (1-\varphi_1)\mu = (1-\varphi_1 B)\mu$. The AR representation of the invertible $ARMA(1, 1)$ model is

$$\pi(B)\dot{Z}_t = a_t,$$

where

$$\pi(B) = (1 + \pi_1 B + \pi_2 B^2 + \cdots) = \frac{(1 - \varphi_1 B)}{(1 - \theta_1 B)},$$

and hence

$$\pi_j = \theta_1^{j-1}(\varphi_1 - \theta_1), \text{ for } j \geq 1. \tag{38}$$

The MA representation of the stationary $ARMA(1, 1)$ model is

$$\dot{Z}_t = \psi(B)a_t,$$

where

$$\psi(B) = (1 + \psi_1 B + \psi_2 B^2 + \cdots) = \frac{(1 - \theta_1 B)}{(1 - \varphi_1 B)},$$

and

$$\psi_j = \varphi_1^{j-1}(\varphi_1 - \theta_1), \text{ for } j \geq 1. \tag{39}$$

Because the $ARMA(p, q)$ process contains $AR(p)$ model and $MA(q)$ model as its special cases, both of its ACF and PACF decays exponentially.

Figure 22.1 illustrates the fundamental properties of some simple AR, MA, and ARMA models in terms of their ACF and PACF.

Nonstationary Time Series Models

In the Stationary Time Series Models section, we introduced stationary time series models. However, in practice, there are many time series that clearly show nonstationary phenomena such as a nonconstant mean and/or nonconstant variance as shown in Figure 22.2 for the yearly U.S. tobacco production between 1871 and 1984.

From earlier discussions, we see that a stationary process is very well characterized by its mean, variance, ACF, and PACF. Because we are using ACF and PACF as our identification tools, we need to use transformations to remove nonstationary phenomena before employing these tools to identify the underlying model. To remove a nonconstant mean, we often use a differencing operator,

$$W_t = (1 - B)^d Z_t, \tag{40}$$

where d is a positive integer. The most commonly used d is 1. For example, Z_t is nonstationary but the series of its changes, $(1-B)Z_t = Z_t - Z_{t-1}$, is often stationary. To remove a nonconstant variance, we can use various variance stabilizing transformations such as Box and Cox's (1964) power transformation,

$$T(Z_t) = \frac{Z_t^\lambda - 1}{\lambda}, \tag{41}$$

which includes logarithmic and many other transformations as special cases. Because the power transformation is defined only for positive series, we may need to add a constant to a series before taking the transformation, which will not affect the correlation structure of the series. With this

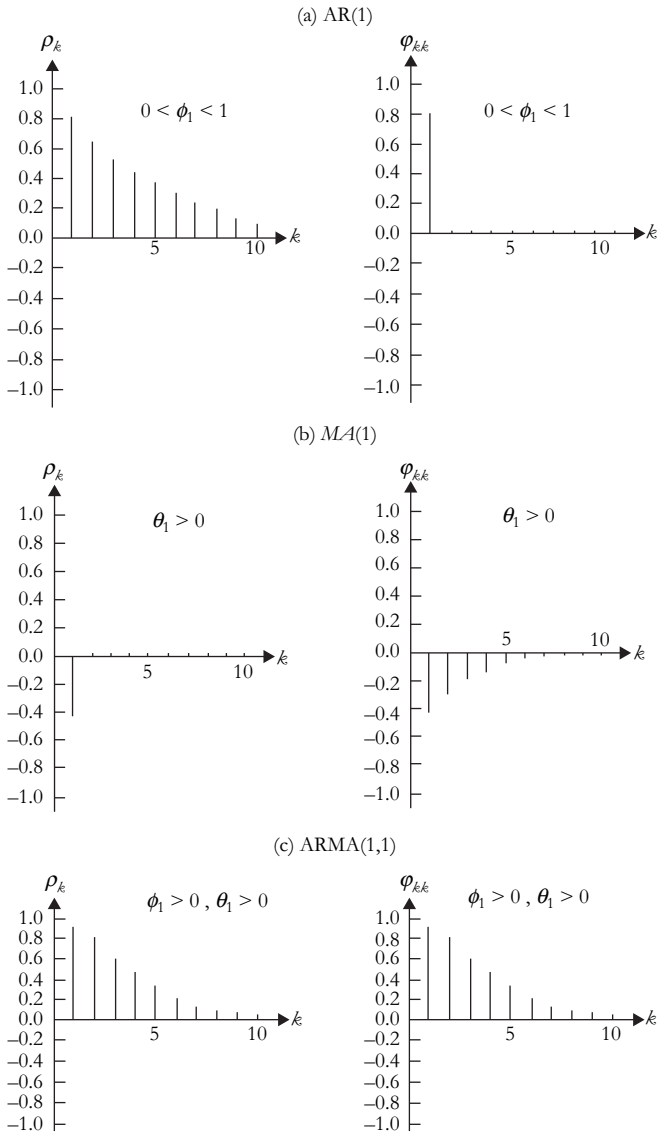

Figure 22.1 ACF and PACF for various time series models.

consideration, we should also apply the power transformation first before taking any differencing. Thus, for a given nonstationary time series, let Z_t be the resulting series from some proper variance stabilizing transformation if necessary. We will extend $ARMA(p, q)$ models to the following *autoregressive integrated moving average* $ARIMA(p, d, q)$ *models*,

$$\varphi_p(B)(1 - B)^d Z_t = \theta_0 + \theta_q(B)a_t, \qquad (42)$$

where the stationary AR polynomial, $\varphi_p(B) = (1 - \varphi_1 B - \cdots - \varphi_p B^p)$, and the invertible MA polynomial, $\theta_q(B) = (1 - \theta_1 B - \cdots - \theta_q B^q)$, are assumed to have no common roots. It should be noted that the parameter θ_0 plays very different roles for $d = 0$ and $d > 0$. When $d = 0$, the process is stationary, and θ_0 is related to the mean of the process—that is, $\theta_0 = (1-\varphi_1 - \cdots - \varphi_p)\mu$. When $d > 0, \theta_0$ is actually equal to the coefficient α_d of t^d from applying the difference operator $(1 - B)^d$ on a deterministic trend $(\alpha_0 + \alpha_1 t + \cdots + \alpha_d t^d)$. For a process without a deterministic trend, the differenced series will have a zero mean. Hence, in general, when $d > 0$, we assume $\theta_0 = 0$ unless the series clearly contains a deterministic component.

Seasonal Time Series Models

Many time series contain a seasonal phenomenon that repeats itself after a regular period of time. The smallest time period for this repetitive phenomenon is called the *seasonal period*. For example, in the U.S.

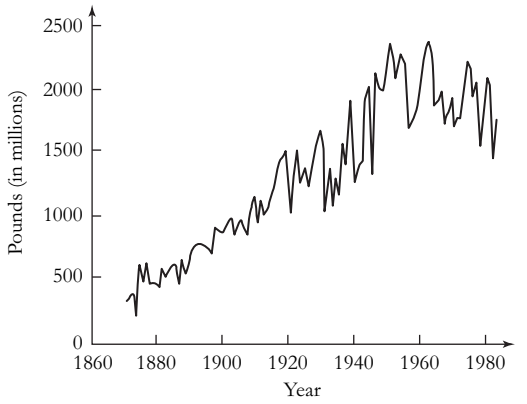

Figure 22.2 Example of a nonstationary time series.

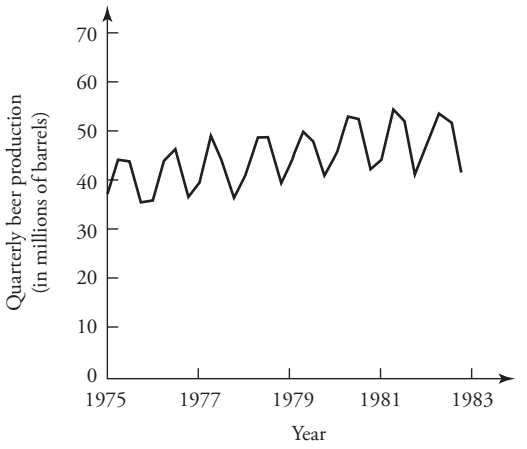

Figure 22.3 Example of a seasonal time series.

quarterly beer production, as shown in Figure 22.3, the beer production is higher during the summer, and the phenomenon repeats itself each year, giving a seasonal period of four.

Suppose we have a seasonal time series with a seasonal period s but fit it with a nonseasonal ARIMA model,

$$\varphi_p(B)(1-B)^d Z_t = \theta_q(B) b_t, \tag{43}$$

where the series b_t will not be white noise because it contains unexplained seasonal correlations. Let

$$\rho_{j(s)} = \frac{E(b_{t-j(s)} - \mu_b)(b_t - \mu_b)}{\sigma_b^2}, j = 1, 2, 3, \ldots$$

be the ACF representing the unexplained seasonal relationship. We can use the following ARIMA model to represent the relation

$$\Phi_P(B^s)(1-B^s)^D b_t = \Theta_Q(B^s) a_t, \tag{44}$$

where $\Phi_P(B^s) = (1 - \Phi_1 B^s - \cdots - \Phi_P B^{Ps})$, $\Theta_Q(B^s) = (1 - \Theta_1 B^s - \cdots - \Theta_Q B^{Qs})$, and a_t is a white noises series with mean *0* and variance σ_a^2. Combining Equations 43 and 44, we obtain the following *seasonal ARIMA model*,

$$\Phi_P(B^s)\varphi_p(B)(1-B^s)^D(1-B)^d Z_t = \theta_0 + \Theta_Q(B^s)\theta_q(B)a_t, \tag{45}$$

which is often denoted as $ARIMA(p, d, q) \times (P, D, Q)_s$, where the index s refers to the seasonal period. For convenience, we often call $(1-B)^d$ and $(1-B^s)^D$ the regular and seasonal difference operators, $\varphi_p(B)$ and $\Phi_P(B^s)$ the regular and seasonal AR polynomials, and $\theta_q(B)$ and $\Theta_Q(B^s)$ the regular and seasonal MA polynomials, respectively.

Time Series Model Building

Model Identification

Given a time series, the first important task is to use the following steps to identify the possible underlying time series model.

Step 1. Plot the time series and if necessary, choose the proper transformation.

Through careful examination of the plot, we usually get a good idea about whether the underlying model is either stationary or nonstationary and seasonal or nonseasonal.

Step 2. Compute and examine the sample ACF and sample PACF of the original series to decide whether differencing is necessary.

Given a series of n observations, we normally compute *n/4* sample ACF and PACF. If the sample ACF decays slowly and the sample PACF cuts off after lag 1, then it indicates that differencing is needed. The process can be repeated and used to find the order, d, of differencing. Sometimes a decision based on visual inspection may be difficult. More rigorously, we can develop a test statistic to determine whether a series is nonstationary and needs differencing. Let us consider the following process,

$$Z_t = \varphi Z_{t-1} + a_t, \tag{46}$$

where a_t is Gaussian $N(0, \sigma_a^2)$ white noise. Given $Z_1, Z_2, \ldots,$ and Z_n, we know that the least square estimator given by,

$$\hat{\varphi} = \frac{\sum_{t=2}^{n} Z_{t-1} Z_t}{\sum_{t=2}^{n} Z_{t-1}^2}, \tag{47}$$

is the best linear unbiased estimator and $\tau = (\hat{\varphi} - \varphi)/S_{\hat{\varphi}}$ follows a $t-$distribution when $|\varphi| < 1$. When $\varphi = 1$, the process becomes nonstationary. It is tempting to use the same test statistic, τ, and

the standard t-distribution to test the hypothesis, $H_0 : \varphi = 1$. However, it has been shown by Dickey and Fuller (1979) and Chan and Wei (1988) that it is hardly a t-distribution. In fact,

$$\tau = \frac{\hat{\varphi} - 1}{S_{\hat{\varphi}}} \to \frac{\frac{1}{2}\left\{[W(1)]^2 - 1\right\}}{\left\{\int_0^1 [W(x)]^2\, dx\right\}^{1/2}}, \qquad (48)$$

where $W(x)$ is a standard Brownian motion process. The percentiles of the distribution were computed by Dickey and Fuller and the critical values are much less than the values from the standard $t-$distributions. Therefore, the test rejects $H_0 : \varphi = 1$ when τ is "too negative." The use of a similar test statistic, $\mathrm{R} = n(\hat{\varphi} - 1)$, was also studied by Dickey and Fuller (1979), who constructed tables of critical values for both τ and R. They extend the above result to the $AR(1)$ with non-zero mean, $Z_t = \alpha + \varphi Z_{t-1} + a_t$, and the $AR(1)$ with a linear time trend, $Z_t = \alpha + \delta t + \varphi Z_{t-1} + a_t$. They also generalized the result to a general $AR(p)$ process, where testing for a unit root is equivalent to testing $\varphi = 1$ in the following model

$$Z_t = \varphi Z_{t-1} + \sum_{j=1}^{p-1} \phi_j \Delta Z_{t-j} + a_t, \qquad (49)$$

where $\Delta Z_{t-j} = (Z_{t-j} - Z_{t-j-1})$. The test is commonly known as the *unit root test*, *Dickey-Fuller test*, or *augmented Dickey-Fuller test*. The test can be repeatedly used on differenced series to determine the order of required differencing and is available in many statistical packages including R (2012), SAS (Statistical Analysis System) (2009), SCA (Scientific Computing Associates) (2008), and SPSS (Statistical Package for the Social Sciences) (2009).

Similarly, for a seasonal time series with the seasonal period s, if the sample ACF decays slowly at multiple lags of s and the sample PACF cuts off after lag s, then it indicates that a seasonal differencing, $(1 - B^s)^D$, is needed for some D. To help identification, we often print these sample ACF and PACF with s of them per line. The seasonal unit root test can also be used to determine the order of required seasonal differencing using the table developed by Dickey, Hasza, and Fuller (1984).

Step 3. Compute and examine the sample ACF and PACF of the properly transformed and differenced series to identify p and q for a regular $ARIMA(p, d, q)$ model or p, q, P, and Q for a seasonal $ARIMA(p, d, q) \times (P, D, Q)_s$ model.

Table 22.1. Characteristics of Theoretical ACF and PACF for Stationary Processes

Process	ACF	PACF
$AR(p)$	Tails off as exponential decay or damped sine wave	Cuts off after lag p
$MA(q)$	Cuts off after lag q	Tails off as exponential decay or damped sine wave
$ARMA(p, q)$	Tails off	Tails off

Again, because variance-stabilizing transformations such as the power transformation are defined only for positive series, they should be performed before any other transformation such as differencing. If necessary, a constant can be added to produce a positive valued series without changing the pattern of the series.

The summary in Table 22.1 should be helpful.

The P and Q for seasonal processes with the seasonal period s are determined similarly based on the ACF and PACF patterns at the lags of multiple s. For example, when $s = 12$, the ACF decays exponentially at $s = 12, 24, 36, \ldots$, and the PACF cuts off after lag 12, implying a seasonal $AR(1)$ model, $(1 - \Phi_1 B^{12})Z_t = a_t$.

One useful procedure to identify the orders of a mixed ARMA model is the use of the *extended sample autocorrelation function* (*ESACF*) or the *smallest canonical correlation* (*SCAN*) introduced by Tsay and Tiao (1984, 1985). They showed that using indicator symbols, with X referring to values greater than or less than ± 2 standard deviations and 0 (zero) for values within ± 2 standard deviations, the orders of an $ARMA(p, q)$ model are determined by the vertex of the triangle formed by these zeros in the ESACF table or the upper-left vertex of the rectangle formed by these zeros in the SCAN table. For example, the ESACF and SCAN tables in Table 22.2 correspond to an $ARMA(1, 1)$ model.

Step 4. Test the deterministic trend term θ_0 when $d > 0$.

One can test for its inclusion by comparing the sample mean $\overline{W}$ of the differenced series $W_t = (1 - B)^d Z_t$, with its approximate standard error

$$S_{\overline{W}} = \left[\frac{\hat{\gamma}_0}{n}(1 + 2\hat{\rho}_1^2 + \cdots + 2\hat{\rho}_k^2)\right]^{1/2} \qquad (50)$$

Table 22.2. ESACF and SCAN Tables for an *ARMA*(1, 1) Model

ESCAF						
AR\MA	0	1	2	3	4	⋯
0	X	X	X	X	X	⋯
1	X	0	0	0	0	⋯
2	X	X	0	0	0	⋯
3	X	X	X	0	0	⋯
4	X	X	X	X	0	⋯
⋮	⋮	⋮	⋮	⋮	⋮	⋱
SCAN						
AR\MA	0	1	2	3	4	⋯
0	X	X	X	X	X	⋯
1	X	0	0	0	0	⋯
	X	0	0	0	0	⋯
3	X	0	0	0	0	⋯
4	X	0	0	0	0	⋯
⋮	⋮	⋮	⋮	⋮	⋮	⋱

Under the null hypothesis $\rho_k = 0$ for $k \geq 1$, it reduces to

$$S_{\overline{W}} = \sqrt{\hat{\gamma}_0/n}. \tag{51}$$

Parameter Estimation

After identifying the orders of a tentative model in Equation 45, we will estimate the parameters in the model. The following are some of the most commonly used estimation methods adopted in various software such as R, SAS, SCA, and SPSS.

1. The method of moments: We express the parameters as functions of moments such as the mean, the variance, the autocovariances, or autocorrelations and then replace these moments by their sample estimates.

2. The maximum likelihood method: We assume a_t following a certain underlying distribution such as a normal distribution, express it as a function of parameters, and then maximize the resultant likelihood function. With a given fixed number of observations, because the expression of a_t involves some unavailable observations and depending on whether any assumption of these unavailable observations is used and how it is used, we have the conditional maximum likelihood estimation, unconditional maximum likelihood estimation, and the exact maximum likelihood estimation.

3. Nonlinear estimation method: Because the expression of a_t as a function of parameters is mostly nonlinear, we can use the nonlinear least squares procedure to find their nonlinear least squares estimates.

Diagnostic Checking

After parameter estimation, we can assess model adequacy by examining whether the model assumption about a_t being white noise is satisfied through various residual analyses such as examining the sample ACF and sample PACF of the residuals. One can also use the portmanteau lack of fit test to test the joint assumption, $H_0: \rho_1 = \cdots = \rho_K = 0$ with the test statistic

$$Q = n(n+2)\sum_{j=1}^{K} \hat{\rho}_j^2/(n-j), \tag{52}$$

where $\hat{\rho}_j^2$ is the residual sample ACF. Under the null hypothesis of model adequacy, the Q statistic was shown by Ljung and Box (1978) to follow approximately a $\chi^2(K-m)$ distribution, where m is the number of AR and MA parameters in the model.

Once we have an adequate model, we can use the model for forecasting, inference, and control. It is important to note that model building is an iterative process as summarized in Figure 22.4.

Model Selection

In data analysis, several models may adequately represent a given data set of n observations. Thus, some criteria have been introduced to help with model selection.

AKAIKE'S *AIC*

Akaike (1974) introduced the following information criterion,

$$AIC(M) = \ln(\hat{\sigma}_a^2) + \frac{2M}{n}, \tag{53}$$

where $\hat{\sigma}_a^2$ is the maximum likelihood estimate of σ_a^2, n is the number of observations, and M is the number of parameters in the model. The optimal order of the model is chosen by the value of M so that $AIC(M)$ is minimum.

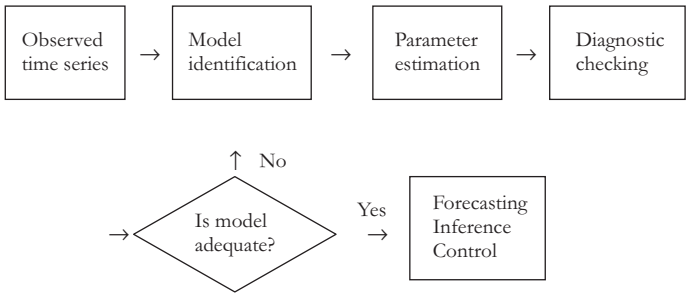

Figure 22.4 Iterative model-building process.

SCHWARTZ'S *SBC*

Schwartz (1978) suggested the following Bayesian criterion of model selection:

$$SBC(M) = \ln(\hat{\sigma}_a^2) + \frac{M \ln(n)}{n}. \qquad (54)$$

Again, the model is chosen by the value of M so that $SBC(M)$ is minimum.

PARZEN'S *CAT*

Because a stationary process can always be approximated by an $AR(p)$ model, Parzen (1977) suggested using AR approximations and computed the following

$$CAT(p) = \begin{cases} -(1 + \frac{1}{n}), & p = 0, \\ \frac{1}{n}\sum_{j=1}^{p} \frac{1}{\hat{\sigma}_j^2} - \frac{1}{\hat{\sigma}_p^2}, & p = 1, 2, \ldots, \end{cases} \qquad (55)$$

where $\hat{\sigma}_j^2$ is the unbiased estimate of σ_a^2 when an $AR(j)$ model is fitted to the series. The optimal order of p is chosen so that $CAT(p)$ is minimum.

As an aid to model selection—especially in selecting orders p and q in an ARMA(p, q) model—SAS has implemented these criteria through its procedure, MINIC. However, it should be noted that either MINIC or ESACF and SCAN methods that were introduced earlier are recommended for suggesting orders p and q of a regular ARMA but not for the orders P and Q of a seasonal ARMA model.

An Illustrative Example of Model Building

The crime index rates are vital statistics that concern many citizens and governments. Figure 22.5 shows a time series Z_t of 49 observations, which is the yearly aggravated assault rate per 100,000 inhabitants of Pennsylvania between 1960 and 2008 obtained from the U.S. FBI Uniform Crime Reports. The series is clearly nonstationary.

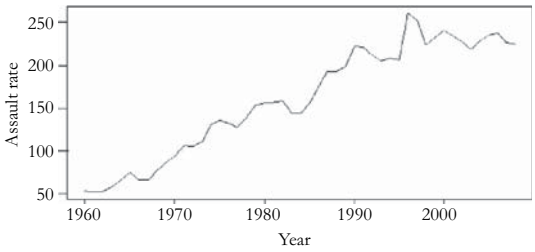

Figure 22.5 The yearly Pennsylvania aggravated assault rate between 1960 and 2008.

Table 22.3. Results of the Power Transformation on Aggravated Assault Rates

λ	Residual mean square error
1.0	309.133
0.5	312.584
0.0	333.326
−0.5	323.860
−1.0	329.683

We first apply Box-Cox power transformation analysis to the series with the result given in Table 22.3. The residual mean square error is lowest when $\lambda = 1$. Thus, no variance stabilizing transformation is needed.

Another nonstationary phenomenon of the series is its increasing trend. This nonstationarity is also shown by the slowly decaying ACF and a single large PACF at lag 1 in Table 22.4 and Figure 22.6. More rigorously, we can apply the Dickey-Fuller unit root tests to the series as shown in Table 22.5. The large p-values for both R and τ statistics for all possible cases clearly indicate that the underlying process for the series contains a unit root.

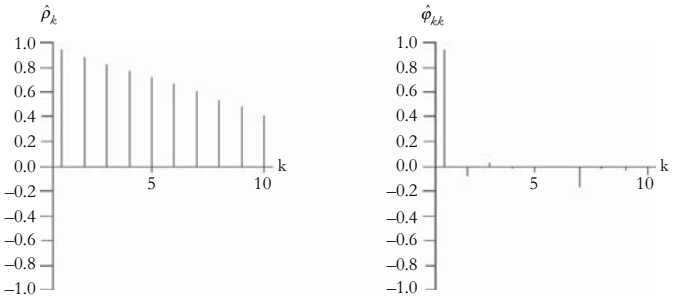

Figure 22.6 Sample ACF and sample PACF of the Pennsylvania aggravated assault rates between 1960 and 2008.

Table 22.4. Sample ACF and Sample PACF of the Pennsylvania Aggravated Assault Rates between 1960 and 2008

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	0.94	0.88	0.82	0.77	0.72	0.67	0.60	0.54	0.48	0.41
St.E.	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14
$\hat{\varphi}_{kk}$	0.94	−.08	0.03	−.01	−.02	−0.00	−0.17	−0.01	−0.03	−.07
St.E.	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14

Table 22.5. Dickey-Fuller Unit Root Tests on the Aggravated Assault Rates

Type	Lags	R	Pr < R (p-value)	τ	Pr < τ (p-value)
Zero mean	0	0.6747	0.8423	1.33	0.9521
	1	0.6446	0.8352	1.14	0.9324
Single mean	0	−1.8037	0.7953	−1.37	0.5910
	1	−2.0912	0.7602	−1.49	0.5306
Trend	0	−9.8459	0.4039	−1.93	0.6264
	1	−15.4288	0.1260	−2.23	0.4606

Because theoretical ACF and PACF do not exist for a nonstationary series, we cannot use Table 22.4 and Figure 22.6 to identify its underlying model. Thus, we compute the sample ACF and PACF of its differenced series, $W_t = (1 - B)Z_t$, which are reported in Table 22.6 with their plots in Figure 22.7. Although the significant PACF cutting off after lag 2 suggests a possible $AR(2)$ model, the significant ACF cutting off after lag 2 suggests an alternative $MA(2)$ model. These selections also agree with the suggested models from ESACF on Table 22.7. The t-ratio, $\overline{W}/S_{\overline{W}} = 3.5667/.3123 = 11.42072$, suggests adding a deterministic trend term. Hence, the following $ARIMA(2, 1, 0)$ and $ARIMA(0, 1, 2)$ models will be entertained:

$$(1 - \varphi_1 B - \varphi_2 B)(1 - B)Z_t = \theta_0 + a_t, \quad (56)$$

and

$$(1 - B)Z_t = \theta_0 + (1 - \theta_1 B - \theta_2 B^2)a_t. \quad (57)$$

The estimation of $AR(2)$ model gives

$$(1 \underset{(0.137)}{-0.05B} + \underset{(0.139)}{0.39}\, B^2)(1 - B)Z_t = \underset{(1.248)}{5.008} + a_t, \quad (58)$$

with $\hat{\sigma}_a^2 = 132.7446$ and AIC $= 374.094$. The estimation of $MA(2)$ model gives

$$(1 - B)Z_t = \underset{(1.051)}{3.79} + (1 \underset{(0.141)}{-0.02B} \underset{(0.146)}{-0.37}\, B^2)a_t, \quad (59)$$

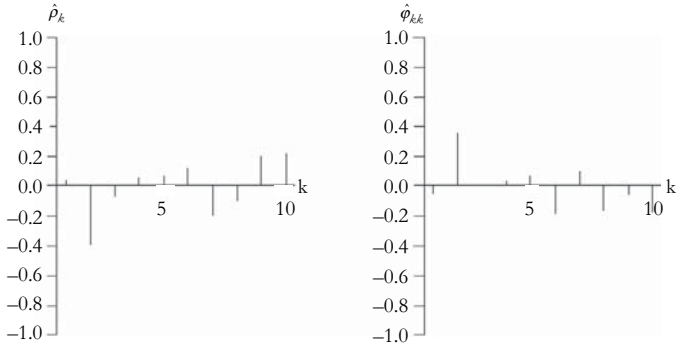

Figure 22.7 Sample ACF and sample PACF for $W_t = (1 - B)Z_t$, where Z_t is the yearly Pennsylvania aggravated assault rate between 1960 and 2008.

Table 22.6. Sample ACF and Sample PACF for $W_t = (1 - B)Z_t$ Where Z_t is the Yearly Pennsylvania Aggravated Assault Rate Between 1960 and 2008. $\overline{W} = 3.5667$, $S_W = .3123$.

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	**0.04**	**−0.39**	**−0.07**	**0.06**	**0.07**	**0.12**	**−0.19**	**−0.09**	**0.20**	**0.22**
St.E.	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14
$\hat{\varphi}_{kk}$	−0.05	0.36	0.01	0.03	0.07	−0.18	0.10	−0.16	−0.05	−0.16
St.E.	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14

Table 22.7. The ESACF for the Differenced Aggravated Assault Rates

AR\MA	0	1	2	3	4	5
0	0	X	0	0	0	0
1	0	X	0	0	0	0
2	0	0	0	0	0	0
3	X	0	0	0	0	0
4	0	X	0	0	0	0
5	0	0	0	0	0	0

with $\hat{\sigma}_a^2 = 134.0952$ and AIC $= 374.5403$. Both models are adequate and pass diagnostic checks. Based on model selection criteria discussed in the Model Selection section, because $\hat{\sigma}_a^2$ and AIC are smaller for the $AR(2)$ model, we will select the $AR(2)$ model as the possible underlying model for the series. However, before we use it for forecasting, inference, and control, we will drop the insignificant parameter φ_1 and re-estimate the model. The final result is

$$(1 + \underset{(0.137)}{0.39}B^2)(1 - B)Z_t = \underset{(0.197)}{5.177} + a_t, \quad (60)$$

with $\hat{\sigma}_a^2 = 130.1907$ and AIC $= 372.2134$, which are smaller than those in the full model in Equation 58.

Time Series Forecasting

One of the most important objectives in the analysis of a time series is to forecast its future values. Let us consider the time series Z_t from the general $ARIMA(p, d, q)$ process

$$\varphi_p(B)(1 - B)^d Z_t = \theta_0 + \theta_q(B)a_t, \quad (61)$$

where θ_0 is normally 0 if $d \neq 0$ and is related to the mean μ of the series when $d = 0, \varphi_p(B) = (1 - \varphi_1 B - \cdots - \varphi_p B^p), \theta_q(B) = (1 - \theta_1 B - \cdots - \theta_q B^q), \varphi_p(B) = 0$ and $\theta_q(B) = 0$ share no common roots that lie outside of the unit circle, and the series a_t is a Gaussian $N(0, \sigma_a^2)$ white noise process.

Minimum Mean Square Error Forecasts and Forecast Limits

The general ARIMA process in Equation 61 can be written as

$$(1 - \phi_1 B - \cdots - \phi_{p+d} B^{p+d}) Z_t = \theta_0 + (1 - \theta_1 B - \cdots - \theta_q B^q) a_t, \quad (62)$$

where $(1 - \phi_1 B - \cdots - \phi_{p+d} B^{p+d}) = \varphi(B)(1 - B)^d$, or equivalently,

$$Z_t = \theta_0 + \phi_1 Z_{t-1} + \cdots + \phi_{p+d} Z_{t-p-d} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}. \quad (63)$$

Suppose that at time $t = n$ we want to forecast the value of $Z_{n+\ell}$. The minimum mean square error forecast $\hat{Z}_n(\ell)$ of $Z_{n+\ell}$ is given by the following conditional expectation

$$\begin{aligned} \hat{Z}_n(\ell) &= E(Z_{n+\ell} \,|Z_t, t \le n) \\ &= \theta_0 + \phi_1 \hat{Z}_n(\ell - 1) + \cdots + \phi_{p+d} \\ &\quad \hat{Z}_n(l - p - d) + \hat{a}_n(\ell) - \theta_1 \hat{a}_n(\ell - 1) \\ &\quad - \cdots - \theta_q \hat{a}_n(\ell - q), \end{aligned} \quad (64)$$

where

$$\begin{aligned} &\hat{Z}_n(j) = E(\hat{Z}_{n+j} \,|Z_t, t \le n), && j \ge 1, \\ &\hat{Z}_n(j) = Z_{n+j}, && j \le 0, \\ &\hat{a}_n(j) = 0, && j \ge 1, \end{aligned}$$

and

$$\hat{a}_n(j) = Z_{n+j} - \hat{Z}_{n+j-1}(1) = a_{n+j}, j \le 0.$$

By rewriting Equation 62 as

$$Z_t = \alpha + \psi(B) a_t = \alpha + \sum_{j=0}^{\infty} \psi_j a_{t-j}, \quad (65)$$

where $(1 - \phi_1 B - \cdots - \phi_{p+d} B^{p+d})(1 + \psi_1 B + \psi_2 B^2 + \cdots) = (1 - \theta_1 B - \cdots - \theta_q B^q)$, α is normally 0 when $d \neq 0$, and $\alpha = \theta_0/(1 - \varphi_1 - \cdots - \varphi_p) = \mu$ when $d = 0$, we can see that the ℓ − step ahead forecast error is

$$e_n(\ell) = Z_{n+\ell} - \hat{Z}_n(\ell) = \sum_{j=0}^{\ell-1} \psi_j a_{n+\ell-j}, \quad (66)$$

where $\psi_0 = 1$. Because $E(e_n(\ell)) = 0$, the forecast is unbiased with the error variance

$$Var(e_n(\ell)) = \left(\sum_{j=0}^{\ell-1} \psi_j^2 \right) \sigma_a^2. \quad (67)$$

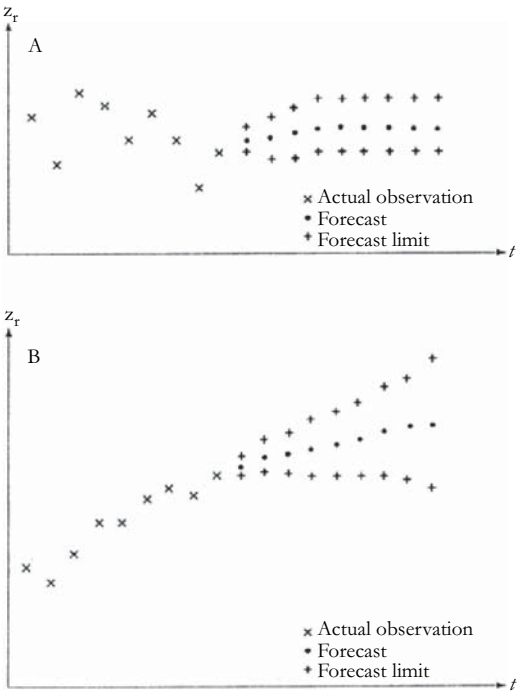

Figure 22.8 Forecasts for (**a**) stationary processes and (**b**) nonstationary processes.

For a normal process, the $(1 - \alpha)100\%$ forecast limits are

$$\hat{Z}_n(\ell) \pm N_{\alpha/2} \left[\sum_{j=0}^{\ell-1} \psi_j^2 \right]^{1/2} \sigma_a, \quad (68)$$

where $N_{\alpha/2}$ is the standard normal deviate such that $P(N > N_{\alpha/2}) = \alpha/2$.

As shown in Figure 22.8, for a stationary process, $\lim_{\ell \to \infty} \sum_{j=0}^{\ell-1} \psi_j^2$ exists and its eventual forecast limits approach to two horizontal lines. For a nonstationary process, because $\sum_{j=0}^{\ell-1} \psi_j^2$ increases as ℓ increases, the forecast limits become wider and wider. The result simply implies that in a nonstationary case, the forecaster becomes less certain about the result as the forecast lead time gets larger.

Updating Forecasts

Note that from Equation 66, we have

$$\begin{aligned} e_n(\ell + 1) &= Z_{n+\ell+1} - \hat{Z}_n(\ell + 1) \\ &= \sum\nolimits_{j=0}^{\ell} \psi_j a_{n+\ell+1-j} \\ &= e_{n+1}(\ell) + \psi_\ell a_{n+1} \\ &= Z_{n+\ell+1} - \hat{Z}_{n+1}(\ell) + \psi_\ell a_{n+1}. \end{aligned}$$

Table 22.8. Yearly Forecasts for the Aggravated Assault Rate per 100,000 Inhabitants in Pennsylvania

Year	Forecast	Std Error	95% Confidence Limits	
2009	234.345	11.4101	211.9579	256.6848
2010	240.146	16.1363	208.4921	271.7453
2011	241.601	17.5829	207.1401	276.0637
2012	244.516	18.9191	207.4489	281.6104

Hence, we obtain the following equation for updating forecasts,

$$\hat{Z}_{n+1}(\ell) = \hat{Z}_n(\ell+1) + \psi_\ell[Z_{n+1} - \hat{Z}_n(1)]. \quad (69)$$

Forecasting Example

From the example in the section An Illustrative Example of Model Building, we have the following model for the series of aggravated assault rate per 100,000 inhabitants in Pennsylvania

$$(1 + \underset{(0.137)}{0.39}B^2)(1-B)Z_t = \underset{(0.197)}{5.177} + a_t, \quad (70)$$

where a_t is Gaussian white noise with mean 0 and variance 130.1907. Given the 49 values of the series from 1960 to 2008, for example, $Z_{47} = 237.6$ for 2006, $Z_{48} = 226.4$ for 2007, and $Z_{49} = 224.82$ for 2008, we can now use the model to forecast future values as follows:

$$\begin{aligned}\hat{Z}_{49}(1) &= 5.177 + Z_{49} - 0.39Z_{48} + 0.39Z_{47}\\ &= 5.177 + 224.8 - 0.39(226.4)\\ &\quad + 0.39(237.6) = 234.345\end{aligned}$$

$$\begin{aligned}\hat{Z}_{49}(2) &= 5.177 + \hat{Z}_{49}(1) - 0.39Z_{49} + 0.39Z_{48}\\ &= 5.177 + 234.345 - 0.39(224.8)\\ &\quad + 0.39(226.4) = 240.146\end{aligned}$$

$$\begin{aligned}\hat{Z}_{49}(3) &= 5.177 + \hat{Z}_{49}(2) - 0.39\hat{Z}_{49}(1) + 0.39Z_{49}\\ &= 5.177 + 240.146 - 0.39(234.345)\\ &\quad + 0.39(224.8) = 241.601,\end{aligned}$$

and for $\ell \geq 4$, we simply use the following forecast equation from the model

$$\begin{aligned}\hat{Z}_{49}(\ell) &= 5.177 + \hat{Z}_{49}(\ell-1) - 0.39\hat{Z}_{49}(\ell-2)\\ &\quad + 0.39\hat{Z}_{49}(\ell-3).\end{aligned}$$

Using Equation 68, we can also compute their 95% forecast limits together with the forecast values given in Table 22.8, which become wider and wider as the forecast lead time gets larger because the model is nonstationary.

Intervention and Outlier Analysis

Time series are often affected by external events such as new treatments, sales promotions, strikes, outbreaks of war, and policy changes. We call these external events interventions and the method of evaluating the effect of the dynamic change for these external events *intervention analysis*.

There are many types of intervention. Some interventions occur at time T and thereafter, which can be represented by

$$I_t = \begin{cases} 1, & t \geq T,\\ 0, & t < T. \end{cases} \quad (71)$$

Some interventions occur only at one time period T, as represented by

$$I_t = \begin{cases} 1, & t = T,\\ 0, & t \neq T. \end{cases} \quad (72)$$

Obviously, there are some interventions that occur at multiple time periods, such as applying a new treatment at time $T_1, T_2, \cdots$, and T_k, and we can represent this type of intervention as

$$I_t = \begin{cases} 1, & t = T_1, T_2, \cdots, T_k,\\ 0, & t \neq= T_1, T_2, \cdots, T_k. \end{cases} \quad (73)$$

There are many possible responses to an intervention. It can be a fixed unknown response after b periods,

$$\omega B^b I_t, \quad (74)$$

a gradual response after b periods,

$$\frac{\omega B^b}{(1-\delta B)} I_t, \quad (75)$$

or more generally, a response that can be described by a rational function,

$$\frac{\omega(B)B^b}{\delta(B)} I_t, \quad (76)$$

where $\omega(B) = \omega_0 - \omega_1 B - \cdots - \omega_s B^s$ and $\delta(B) = 1 - \delta_1 B - \cdots - \delta_r B^r$ are polynomials in B, b represents the time delay for the intervention effect, and the weights ω_j's in the polynomial $\omega(B)$ often represent the expected initial effects of the intervention. The polynomial $\delta(B)$, on the other hand, measures the behavior of the permanent effect of the intervention. The roots of $\delta(B) = 0$ are assumed to be on or outside the unit circle. The unit root represents an impact that increases linearly, and the root outside the unit circle represents a phenomenon that has a gradual response.

Clearly, Equation 76 contains Equations 74 and 75 as special cases. Thus, in general with multiple interventions, we can represent the phenomenon with the following *intervention model*,

$$Z_t = \theta_0 + \sum_{j=1}^{K} \frac{\omega_j(B)B^{b_j}}{\delta_j(B)} I_{jt} + \frac{\theta(B)}{\phi(B)} a_t, \quad (77)$$

where $I_{jt}, j = 1, \cdots, K$ are intervention variables. The form $\omega_j(B)B^{b_j}/\delta_j(B)$ for the jth intervention is postulated based on the expected form of the response given knowledge of the intervention. Because the main purpose of intervention models is to measure the effect of interventions, Box and Tiao (1975), who introduced the intervention model, called the time series free of interventions, represented by $[\theta(B)/\phi(B)]a_t$, where $\phi(B) = \varphi_p(B)(1-B)^d$, as the noise model. The noise model is usually identified using the time series Z_t before the intervention date. For a nonstationary process, the model in Equation 77 normally does not contain a constant term θ_0.

Time series are sometimes affected by interruptive events. The consequences of these interruptions create spurious observations that are inconsistent with the rest of the series. Such observations are usually referred to as *outliers*. When the timing and causes of interruptions are known, their effects can be accounted for by using the intervention model. However, the timing and causes of interruptions are often unknown. Because outliers are known to wreak havoc in data analysis, making the resultant inference unreliable or invalid, it is important to have procedures that will detect and/or remove such outlier effects. There are many types of time series outliers including *additive outliers* (*AO*), *innovational outliers* (*IO*), *level shift* (*LS*), and *transitory change* (*TC*).

Let Z_t be the observed series and X_t be the outlier-free series. Assume that X_t follows a general $ARMA(p, q)$ model, $\varphi_p(B)X_t = \theta_q(B)a_t$. An AO is the outlier that affects only the Tth observation, Z_T, and so

$$Z_t = \omega I_t^{(T)} + X_t = \omega I_t^{(T)} + \frac{\theta_q(B)}{\varphi_p(B)} a_t, \quad (78)$$

where

$$I_t^{(T)} = \begin{cases} 1, & t = T, \\ 0, & t \neq T, \end{cases}$$

is the indicator variable representing the presence or absence of an outlier at time T. An IO is the outlier that affects all observations beyond T through the memory of the system described by $\theta_q(B)/\varphi_p(B)$, and so

$$Z_t = \frac{\theta_q(B)}{\varphi_p(B)} \omega I_t^{(T)} + X_t = \frac{\theta_q(B)}{\varphi_p(B)} \left(\omega I_t^{(T)} + a_t \right). \quad (79)$$

A LS outlier is the outlier that corresponds to a shift of the level of the process starting from time T and continues afterward,

$$Z_t = \frac{1}{(1-B)} \omega I_t^{(T)} + X_t = \frac{1}{(1-B)} \omega I_t^{(T)} + \frac{\theta_q(B)}{\varphi_p(B)} a_t, \quad (80)$$

which is equivalent to a sequence of additive outliers of the same size occurring at time T and afterward. A TC outlier is a level shift that produces an initial impact but the impact decays exponentially as $1/(1 - \delta B)$, so

$$Z_t = \frac{1}{(1-\delta B)} \omega I_t^{(T)} + X_t = \frac{1}{(1-\delta B)} \omega I_t^{(T)} + \frac{\theta_q(B)}{\varphi_p(B)} a_t. \quad (81)$$

The detection of time series outliers was first introduced by Fox (1972). Other references include Chang, Tiao, and Chen (1988), Tsay (1988), Chen and Liu (1991), and Lee and Wei (1995). These procedures have been implemented in many time series software such as SAS, SCA, and SPSS.

In searching for the causes of an outlier, one may find the nature of the disturbance. Some outliers may turn out to be important intervention variables that the analyst overlooked during the preliminary stages of the analysis. We can obviously have a combined *intervention-outlier model*, as illustrated in the following example.

Example of Outlier and Intervention Analysis

As an example, let us consider the monthly airline passengers in the United States from January 1995 to March 2002 plotted in Figure 22.9.

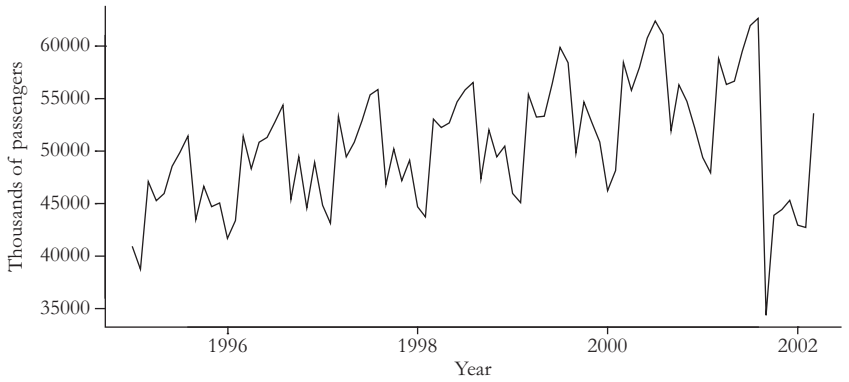

Figure 22.9 The monthly airline passengers in the United States from January 1995 to March 2002.

Without looking at the plot and blindly applying the outlier detection method introduced above with SCA, we obtain the following result:

Detected outliers	
Time	Type
81	TC
82	TC

If we use a significance level less than 0.01, the only outlier found is the observation at time 81 that corresponds to September 2001, the month of the World Trade Center tragedy in New York City. The incident clearly is an intervention event. The outlier procedure not only detects the event but also suggests the form of the intervention.

The standard time series modeling on the subseries from January 1995 to August 2001 suggests the $ARIMA(2,0,0) \times (0,1,0)_{12}$ seasonal model:

$$(1 - \varphi_1 B - \varphi_2 B^2)(1 - B^{12})Z_t = a_t. \qquad (82)$$

Thus, we will combine Model 82 and the information about the observation at time 81 in the following intervention model:

$$Z_t = \frac{\omega}{(1 - \delta B)} I_t + \frac{1}{(1 - \varphi_1 B - \varphi_2 B^2)(1 - B^{12})} a_t, \qquad (83)$$

where

$$I_t = \begin{cases} 0, & t < 81 (\text{Sept., } 2001), \\ 1, & t \geq 81 \text{ (Sept., } 2001). \end{cases}$$

The estimation results are:

Parameter	Estimate	St. Error
ω	$-18,973.5$	1299.3
δ	0.76	0.06
φ_1	0.62	0.1
φ_2	0.21	0.1

The impact of the September 11th tragedy on the airline industry is clearly devastating.

Transfer Function and Time Series Regression Models

In earlier sections, we were concerned with univariate time series models. In this section, we will consider models where an output series is related to one or more input series.

Transfer Function Models

Assume that X_t and Y_t are properly transformed series so that they are both stationary. The *transfer function model* is the following model that relates input and output variables:

$$\begin{aligned} Y_t &= \upsilon_0 X_t + \upsilon_1 X_{t-1} + \upsilon_2 X_{t-2} + \cdots + N_t \\ &= \upsilon(B) X_t + N_t, \end{aligned} \qquad (84)$$

where $\upsilon(B) = \sum_{j=0}^{\infty} \upsilon_j B^j$ is the transfer function for the system and the υ_j are known as impulse response weights. Figure 22.10 illustrated this dynamic system.

In practice, we often represent $\upsilon(B)$ with the following rational function,

$$\upsilon(B) = \frac{\omega_s(B)}{\delta_r(B)} B^b, \qquad (85)$$

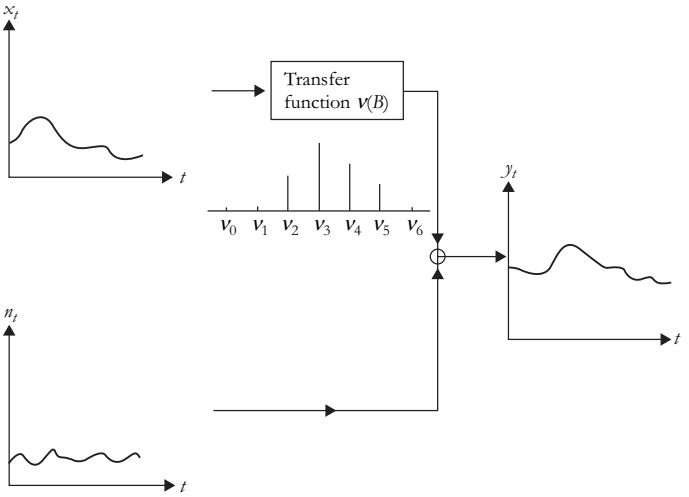

Figure 22.10 Dynamic transfer function system.

where $\omega_s(B) = \omega_0 - \omega_1 B - \cdots - \omega_s B^s$, $\delta_r(B) = 1 - \delta_1 B - \cdots - \delta_r B^r$, b is the delay parameter representing the time lag that elapses before the input variable produces an effect on the output variable, and N_t is the noise series of the system that is independent of the input series. For a stable system, we assume that the roots of $\delta_r(B) = 0$ are outside of the unit circle. When X_t and N_t are assumed to follow some ARMA processes, the system is also known as the *ARMAX model*.

A useful measure for studying the relationship between time series variables is the *cross-correlation function* (*CCF*),

$$\rho_{XY}(k) = \frac{\gamma_{XY}(k)}{\sigma_X \sigma_Y}, \tag{86}$$

where $\gamma_{XY}(k) = E[(X_t - \mu_X)(Y_{t+k} - \mu_Y)]$ is the *cross-covariance function* between X_t and Y_t. The sample CCF is given by,

$$\hat{\rho}_{XY}(k) = \frac{\hat{\gamma}_{XY}(k)}{S_X S_Y}, \tag{87}$$

where

$$\hat{\gamma}_{XY}(k) = \begin{cases} \frac{1}{n}\sum_{t=1}^{n-k}(X_t - \overline{X})(Y_{t+k} - \overline{Y}), & k \geq 0, \\ \frac{1}{n}\sum_{t=1-k}^{n}(X_t - \overline{X})(Y_{t+k} - \overline{Y}), & k < 0, \end{cases} \tag{88}$$

$$S_X = \sqrt{\hat{\gamma}_{XX}(0)},\ S_Y = \sqrt{\hat{\gamma}_{YY}(0)},$$

and $\overline{X}$ and $\overline{Y}$ are the sample means of the X_t and Y_t series, respectively. Under the hypothesis that X_t and Y_t are uncorrelated and X_t is white noise, we have

$$Var[\hat{\rho}_{XY}(k)] \approx (n-k)^{-1}.$$

When the input series is white noise, it can be shown that

$$\upsilon_k = \frac{\sigma_Y}{\sigma_X}\rho_{XY}(k). \tag{89}$$

This result leads to the following procedure of transfer function model identification:

1. Prewhiten the input series:

$$\varphi_X(B)X_t = \theta_X(B)\alpha_t.$$

So

$$\alpha_t = \frac{\varphi_X(B)}{\theta_X(B)}X_t, \tag{90}$$

and α_t is a white noise series with mean 0 and variance σ_α^2.

2. Calculate the filtered output series:

$$\beta_t = \frac{\varphi_X(B)}{\theta_X(B)}Y_t. \tag{91}$$

3. Calculate the sample CCF, $\hat{\rho}_{\alpha\beta}(k)$, between α_t and β_t to estimate υ_k:

$$\hat{\upsilon}_k = \frac{\hat{\sigma}_\beta}{\hat{\sigma}_\alpha}\hat{\rho}_{\alpha\beta}(k). \tag{92}$$

The significance of the CCF and its equivalent υ_k can be tested by comparing it with its standard error $(n-k)^{-1/2}$.

4. Identify the delay parameter b, the order r in $\delta_r(B) = (1 - \delta_1 B - \cdots - \delta_r B^r)$, and the order s in $\omega_s(B) = (\omega_0 - \omega_1 B - \cdots - \omega_s B^s)$ using the pattern of $\hat{\upsilon}_k$. Table 22.9 illustrates some typical impulse weights and their corresponding transfer

Table 22.9. Some Typical Impulse Weights and Their Corresponding Transfer Functions

(b,r,s)	Transfer function	Typical impulse weights
(2, 0, 0)	$\upsilon(B)x_t = \omega_0 x_{t-2}$	
(2, 0, 1)	$\upsilon(B)x_t = (\omega_0 - \omega_1 B)x_{t-2}$	
(2, 0, 2)	$\upsilon(B)x_t = (\omega_0 - \omega_1 B - \omega_2 B^2)x_{t-2}$	
(2, 1, 0)	$\upsilon(B)x_t = \frac{\omega_0}{(1-\delta_1 B)}x_{t-2}$	
(2, 1, 1)	$\upsilon(B)x_t = \frac{(\omega_0 - \omega_1 B)}{(1-\delta_1 B)}x_{t-2}$	
(2, 1, 2)	$\upsilon(B)x_t = \frac{(\omega_0 - \omega_1 B - \omega_2 B^2)}{(1-\delta_1 B)}x_{t-2}$	

functions with $b = 2$. Thus, we have our preliminary transfer function for the system:

$$\upsilon(B)X_t = \frac{\omega_s(B)}{\delta_r(B)}B^b X_t. \tag{93}$$

5. Once we obtain the preliminary transfer function, we can calculate the estimated noise series,

$$\hat{N}_t = Y_t - \frac{\hat{\omega}_s(B)}{\hat{\delta}_r(B)}B^b X_t.$$

We then use identification statistics such as sample ACF and PACF to identify the noise model,

$$\varphi(B)N_t = \theta(B)a_t. \tag{94}$$

Combining Equations 93 and 94, we have our entertained transfer function model:

$$Y_t = \frac{\omega_s(B)}{\delta_r(B)}X_{t-b} + \frac{\theta(B)}{\varphi(B)}a_t. \tag{95}$$

Because Equations 95 can be rewritten in terms of a_t as a function of $Y_t's, X_t's$, and past values of a_t, the estimation methods discussed in the Parameter Estimation section can be used to estimate the parameters. Once the parameters are estimated, we will check the model adequacy by examining the CCF, $\hat{\rho}_{\alpha\hat{a}}(k)$, between α_t and $\hat{a}_t$, and ACF and PACF of $\hat{a}_t$ to make sure they are all insignificant and do not show any patterns as specified in the assumptions of our model. We can then use the adequate model for forecasting, inference, and control. We refer readers to Box, Jenkins, and Reinsel (2008, Chapters 11 and 12) for more details.

Regression Time Series Models

A regression model is used to study the relationship of a dependent variable with one or more independent variables. The standard regression model is represented by the following equation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon,$$

where Y is the dependent variable, $X_1, \cdots, X_k$ are the independent variables, $\beta_0, \beta_1, \cdots, \beta_k$ are the regression coefficients, and ε is the error term. When time series data are used in the model, it becomes *time series regression*, and the model is often written as

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon_t,$$

or equivalently,

$$Y_t = \mathbf{X}'_t\boldsymbol{\beta} + \varepsilon_t, \tag{96}$$

where $\mathbf{X}'_t = [1, X_{1,t}, \cdots, X_{k,t}]$ and $\boldsymbol{\beta} = [\beta_0, \beta_1, \cdots, \beta_k]'$.

The standard regression assumptions for the error variable are that the ε_t are i.i.d. $N(0, \sigma_\varepsilon^2)$. Under these standard assumptions, it is well known that the *ordinary least squares* (*OLS*) *estimator* $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is a minimum variance unbiased estimator and distributed as multivariate normal, $N(\boldsymbol{\beta}, \sigma_\varepsilon^2\mathbf{I})$ When $\mathbf{X}'_t$ is stochastic in Model 96 and conditional on $\mathbf{X}'_t$, the results about the OLS estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ also hold as long as ε_s and $\mathbf{X}'_t$ are independent for all s and t. However, the standard assumptions associated with these models are often violated when time series data are used.

REGRESSION WITH AUTOCORRELATED ERRORS

When $\mathbf{X}'_t$ is a vector of a constant 1 and k lagged values of Y_t—that is, $\mathbf{X}'_t = (1, Y_{t-1}, \cdots, Y_{t-k})$ and ε_t is white noise, the model in Equation 96 states that the variable Y_t is regressed on its own past k lagged values and hence is known as autoregressive model of order k—that is, $AR(k)$ model

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \ldots + \beta_k Y_{t-k} + \varepsilon_t. \quad (97)$$

The OLS estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is still a minimum variance unbiased estimator. However, this result no longer holds when the ε_t are autocorrelated. In fact, when this is the case, the estimator is not consistent and the usual tests of significance are invalid. This is an important caveat.

When time series are used in a model, it is the norm rather than the exception that the error terms are autocorrelated. Even in univariate time series analysis when the underlying process is known to be an AR model as in Equation 97, the error terms ε_t could still be autocorrelated unless the correct order of k is chosen. Thus, a residual analysis is an important step in regression analysis when time series variables are involved in the study.

There are many methods that can be used to test for autocorrelation of the error term. For example, one can use the test based on the Durbin-Watson statistic. More generally, to study the autocorrelation structure of the error term, we can perform the residual analysis with time series model identification statistics like the sample ACF and sample PACF. Through these identification statistics, one can detect not only whether the residuals are autocorrelated but also identify its possible underlying model. A final analysis can then be performed on a model with autocorrelated errors as follows:

$$Y_t = \mathbf{X}'_t\boldsymbol{\beta} + \varepsilon_t \quad (98)$$

for $t = 1, 2, \ldots, n$, where

$$\varepsilon_t = \varphi_1\varepsilon_{t-1} + \ldots + \varphi_p\varepsilon_{t-p} + a_t \quad (99)$$

and the a_t are i.i.d. $N(0, \sigma_a^2)$.

Let

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} \mathbf{X}'_1 \\ \vdots \\ \mathbf{X}'_n \end{bmatrix}, \text{ and } \boldsymbol{\xi} = \begin{bmatrix} \varepsilon'_1 \\ \vdots \\ \varepsilon'_n \end{bmatrix}.$$

The matrix form of the model in Equation 98 is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi} \quad (100)$$

where $\boldsymbol{\xi}$ follows a multivariate normal distribution $N(\mathbf{0}, \boldsymbol{\Sigma})$. When $\varphi_1, \ldots, \varphi_p$, and σ^2 are known in Equation 99, $\boldsymbol{\Sigma}$ can be easily calculated. The diagonal element of $\boldsymbol{\Sigma}$ is the variance of ε_t, the jth off-diagonal element corresponds to the jth autocovariance of ε_t, and they can be easily computed from Equation 99. Given $\boldsymbol{\Sigma}$, the *generalized least squares* (*GLS*) *estimator*,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} \quad (101)$$

is known to be a minimum variance unbiased estimator.

Normally, we will not know the variance–covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{\xi}$ because even if ε_t follows an $AR(p)$ model given in Equation 99, the σ^2 and AR parameters φ_j are usually unknown. As a remedy, the following iterative GLS is often used:

1a. Calculate OLS residuals $\hat{\varepsilon}_t$ from OLS fitting of Model 98.

1b. Estimate φ_j and σ^2 for the $AR(p)$ model in Equation 99 based on the OLS residuals, $\hat{\varepsilon}_t$, using any time series estimation method. For example, a simple conditional OLS estimation can be used.

1c. Compute $\boldsymbol{\Sigma}$ from model in Equation 99 using the values of φ_j and σ^2 obtained in Step 1b.

1d. Compute GLS estimator, $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}$, using the $\boldsymbol{\Sigma}$ obtained in Step 1c.

Compute the residuals $\hat{\varepsilon}_t$ from the GLS model fitting in Step 1d, and repeat the above Steps 1b through 1d until some convergence criterion (such as the maximum absolute value change in the estimates between iterations become less than some specified quantity) is reached.

More generally, the error structure can be modified to include an ARMA model. The above GLS iterative estimation can still be used with the exception that a nonlinear least squares estimation rather than OLS is needed to estimate the parameters in the error model. Alternatively, by substituting the error process in the regression model Equation 98, we can also use the nonlinear estimation or maximum likelihood estimation to jointly estimate the regression and error model parameters $\boldsymbol{\beta}$ and $\varphi'_j s$, which is available in many standard software.

It should be pointed out that although the error term, ε_t, can be autocorrelated in the regression model, it should be stationary. A nonstationary error structure could produce a spurious regression, where a significant regression can be achieved for totally unrelated series.

REGRESSION WITH HETEROSCEDASTICITY

One of the main assumptions of the standard regression model in Equation 96 or the regression model with autocorrelated errors in Equation 98 is that the variance, σ_ε^2, is constant. In many applications, this assumption may not be realistic. For example, in financial investments, it is generally agreed that the stock market's volatility is rarely constant.

A model with a non-constant error variance is called a *heteroscedasticity model*. There are many approaches that can be used to deal with heteroscedasticity. For example, the weighted regression is often used if the error variances at different times are known or if the variance of the error term varies proportionally to the value of an independent variable. In time series regression, we often have a situation where the variance of the error term is related to the magnitude of past errors. This phenomenon leads to the conditional heteroscedasticity model, introduced by Engle (1982), where in terms of Equation 96 we assume that

$$\varepsilon_t = \sigma_t e_t \tag{102}$$

where e_t is the series of i.i.d. random variables with mean 0 and variance 1, and

$$\sigma_t^2 = \theta_0 + \theta_1 \varepsilon_{t-1}^2 + \theta_2 \varepsilon_{t-2}^2 + \ldots + \theta_s \varepsilon_{t-s}^2. \tag{103}$$

Given all of the information up to time $(t-1)$, the conditional variance of the ε_t becomes

$$\begin{aligned} Var_{t-1}(\varepsilon_t) &= E_{t-1}(\varepsilon_t^2) \\ &= E(\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) = \sigma_t^2 \\ &= \theta_0 + \theta_1 \varepsilon_{t-1}^2 + \theta_2 \varepsilon_{t-2}^2 \\ &\quad + \ldots + \theta_s \varepsilon_{t-s}^2, \end{aligned} \tag{104}$$

which is related to the squares of past errors, and it changes over time. A large error through ε_{t-j}^2 gives rise to the variance, which tends to be followed by another large error. This is a common phenomenon of volatility clustering in many financial time series. From the forecasting results, we see that Equation 103 is simply the optimal forecast of ε_t^2 from the following $AR(s)$ model:

$$\varepsilon_t^2 = \theta_0 + \theta_1 \varepsilon_{t-1}^2 + \theta_2 \varepsilon_{t-2}^2 + \ldots + \theta_s \varepsilon_{t-s}^2 + a_t, \tag{105}$$

where the a_t is a $N(0, \sigma_a^2)$ white noise process. Thus, Engle (1982) called the model of the error term ε_t with the variance specification given in Equations 102 and 103 or equivalently in Equation 105 the *autoregressive conditional heteroscedasticity model of order s* (*ARCH*(*s*)).

Bollerslev (1986) extended the ARCH(*s*) model to the *GARCH*(r, s) *model* (*generalized autoregressive conditional heteroscedasticity model of order* (r, s)) so that the conditional variance of the error process is related not only to the squares of past errors but also to the past conditional variances. Thus, we have the following more general case,

$$\varepsilon_t = \sigma_t e_t, \tag{106}$$

where e_t is the series of i.i.d. random variables with mean 0 and variance 1,

$$\begin{aligned} \sigma_t^2 &= \theta_0 + \phi_1 \sigma_{t-1}^2 + \ldots + \phi_r \sigma_{t-r}^2 \\ &\quad + \theta_1 \varepsilon_{t-1}^2 + \ldots + \theta_s \varepsilon_{t-s}^2, \end{aligned} \tag{107}$$

and the roots of $(1 - \phi_1 B - \ldots - \phi_r B^r) = 0$ are outside the unit circle. To guarantee $\sigma_t^2 > 0$, we assume that $\theta_0 > 0$, and ϕ_i and θ_j are non-negative.

More generally, the regression model with autocorrelated error can be combined with the conditional heteroscedasticity model—that is,

$$Y_t = \mathbf{X}'_t \beta + \varepsilon_t, \tag{108}$$

where

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \cdots + \phi_p \varepsilon_{t-p} + a_t, \tag{109}$$

$$\varepsilon_t = \sigma_t e_t, \tag{110}$$

$$\begin{aligned} \sigma_t^2 &= \theta_0 + \phi_1 \sigma_{t-1}^2 + \cdots + \phi_r \sigma_{t-r}^2 \\ &\quad + \theta_1 a_{t-1}^2 + \cdots + \theta_s a_{t-s}^2, \end{aligned} \tag{111}$$

and the e_t are i.i.d. $N(0, 1)$. To test for the heteroscedasticity in this model, we perform the following steps:

1a. Calculate OLS residuals $\hat{\varepsilon}_t$ from the OLS fitting of Equation 108.

1b. Fit an $AR(p)$ Model 109 to the $\hat{\varepsilon}_t$.

1c. Obtain the residuals $\hat{a}_t$ from the AR fitting in Model 109.

1d. Form the series $\hat{a}_t^2$, compute its sample ACF and PACF, and check whether these ACF and PACF follow any pattern. A pattern of these ACF and PACF not only indicates ARCH or GARCH errors, it also forms a good basis for their order specification. Alternatively, we can also use the following portmanteau Q statistic to test for $\rho_i(a_t^2) = 0, i = 1, 2, \ldots, k$,

$$Q(k) = n(n+2) \sum_{i=1}^{k} \frac{\hat{\rho}_i^2(\hat{a}_t^2)}{(n-i)}, \tag{112}$$

which approximately follows a $\chi^2(k)$ distribution. The significance of the $Q(k)$ statistic occurring

only for a small value of k indicates an ARCH model, and a persistent significance for a large value of k implies a GARCH model.

Vector Time Series Models

In transfer function and time series regression models, we study the relationship between an output or a dependent variable and a set of input or independent variables. In many applications, the relationship represented in these models may not be appropriate. In this section, we introduce the extension of the univariate time series models from the section on Univariate Time Series Models to vector time series models and use them to describe the relationships among several time series variables.

Just like univariate time series models are characterized by their moments such as means, variances, ACFs, and PACFs, vector time series models are also characterized by their moments such as mean vectors, variance–covariance matrices, correlation matrix functions, and partial correlation matrix functions.

Correlation and Partial Correlation Matrix Functions

Let $\mathbf{Z}_t = [Z_{1,t}, Z_{2,t}, \cdots, Z_{m,t}]', t = 0, \pm 1, \pm 2, \ldots$, be a m-dimensional jointly stationary real-valued vector process so that $E(Z_{i,t}) = \mu_i$ is constant for each $i = 1, 2, \ldots, m$ and the cross-covariance between $Z_{i,t}$ and $Z_{j,s}$, for all $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, m$, are functions only of the time difference $(s - t)$. Hence, we have the *mean vector*

$$E(\mathbf{Z}_t) = \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{bmatrix}, \tag{113}$$

and the *lag-k covariance matrix*

$$\begin{aligned} \boldsymbol{\Gamma}(k) &= Cov\{\mathbf{Z}_t, \mathbf{Z}_{t+k}\} = E[(\mathbf{Z}_t - \boldsymbol{\mu})(\mathbf{Z}_{t+k} - \boldsymbol{\mu})'] \\ &= E\begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ \vdots \\ Z_{m,t} \end{bmatrix} \begin{bmatrix} Z_{1,t+k} - \mu_1, & Z_{2,t+k} - \mu_2, \\ \cdots, & Z_{m,t+k} - \mu_m \end{bmatrix} \\ &= \begin{bmatrix} \gamma_{11}(k) & \gamma_{12}(k) & \cdots & \gamma_{1m}(k) \\ \gamma_{21}(k) & \gamma_{22}(k) & \cdots & \gamma_{2m}(k) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{m1}(k) & \gamma_{m2}(k) & \cdots & \gamma_{mm}(k) \end{bmatrix}, \end{aligned} \tag{114}$$

where

$$\gamma_{ij}(k) = E(Z_{i,t} - \mu_i)(Z_{j,t+k} - \mu_j)$$

for $k = 0, \pm 1, \pm 2, \ldots, i = 1, 2, \ldots, m$, and $j = 1, 2, \ldots, m$. As a function of k, $\boldsymbol{\Gamma}(k)$ is called the *covariance matrix function* for the vector process $\mathbf{Z}_t$. For $i = j, \gamma_{ii}(k)$ is the autocovariance function for the ith component process $Z_{i,t}$; and for $i \neq j, \gamma_{ij}(k)$ is the cross-covariance function between component series $Z_{i,t}$ and $Z_{j,t}$. The matrix $\boldsymbol{\Gamma}(0)$ is easily seen to be the contemporaneous variance–covariance matrix of the process.

The *correlation matrix function* for the vector process is defined by

$$\boldsymbol{\rho}(k) = \mathbf{D}^{-1/2}\boldsymbol{\Gamma}(k)\mathbf{D}^{-1/2} = \left[\rho_{ij}(k)\right] \tag{115}$$

for $i = 1, 2, \ldots, m$, and $j = 1, 2, \ldots, m$, where $\mathbf{D}$ is the diagonal matrix in which the ith diagonal element is the variance of the ith process; that is, $\mathbf{D} = diag[\gamma_{11}(0), \gamma_{22}(0), \ldots, \gamma_{mm}(0)]$. Thus, the ith diagonal element of $\boldsymbol{\rho}(k)$ is the ACF for the ith component series $Z_{i,t}$ whereas the (i, j)th off-diagonal element of $\boldsymbol{\rho}(k)$ is the cross-correlation function between component series $Z_{i,t}$ and $Z_{j,t}$.

Unlike the correlation matrix function that follows the standard definition given in Equation 115, the concept of a partial correlation matrix function has been introduced much later (the correlation matrix function was introduced before 1900 and the concept of a partial correlation matrix function was introduced only after 1980) and there are different versions.

Heyse and Wei (1985) extended the definition of univariate partial autocorrelation to vector time series and derived the correlation matrix between $\mathbf{Z}_t$ and $\mathbf{Z}_{t+s}$ after removing the linear dependence of each on the intervening vectors $\mathbf{Z}_{t+1}, \ldots, \mathbf{Z}_{t+s-1}$. This correlation matrix is defined as the correlation between the residual vectors

$$\begin{aligned} \mathbf{U}_{s-1,t+s} &= \mathbf{Z}_{t+s} - \boldsymbol{\alpha}_{s-1,1}\mathbf{Z}_{t+s-1} - \cdots - \boldsymbol{\alpha}_{s-1,s-1}\mathbf{Z}_{t+1} \\ &= \begin{cases} \mathbf{Z}_{t+s} - \sum_{j=1}^{s-1} \boldsymbol{\alpha}_{s-1,j}\mathbf{Z}_{t+s-j}, & s \geq 2, \\ \mathbf{Z}_{t+1}, & s = 1, \end{cases} \end{aligned} \tag{116}$$

and

$$\begin{aligned} \mathbf{V}_{s-1,t} &= \mathbf{Z}_t - \boldsymbol{\beta}_{s-1,1}\mathbf{Z}_{t+1} - \cdots - \boldsymbol{\beta}_{s-1,s-1}\mathbf{Z}_{t+s-1} \\ &= \begin{cases} \mathbf{Z}_t - \sum_{j=1}^{s-1} \boldsymbol{\beta}_{s-1,j}\mathbf{Z}_{t+j}, & s \geq 2, \\ \mathbf{Z}_{t+1}, & s = 1. \end{cases} \end{aligned} \tag{117}$$

Let $\mathbf{C}_{\mathbf{VU}}(s)$ be the covariance between $\mathbf{V}_{s-1,t}$ and $\mathbf{U}_{s-1,t+s}$—that is, $\mathbf{C}_{\mathbf{VU}}(s) = Cov(\mathbf{V}_{s-1,t}, \mathbf{U}_{s-1,t+s})$, Heyse and Wei (1985) showed that

$$\begin{aligned}
&\mathbf{C}_{\mathbf{VU}}(s) \\
&= \boldsymbol{\Gamma}(s) - \begin{bmatrix} \boldsymbol{\Gamma}(s-1) & \boldsymbol{\Gamma}(s-2) & \cdots & \boldsymbol{\Gamma}(1) \end{bmatrix} \\
&\begin{bmatrix} \boldsymbol{\Gamma}(0) & \boldsymbol{\Gamma}'(1) & \cdots & \boldsymbol{\Gamma}'(s-2) \\ \boldsymbol{\Gamma}(1) & \boldsymbol{\Gamma}(0) & \cdots & \boldsymbol{\Gamma}'(s-3) \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Gamma}(s-2) & \boldsymbol{\Gamma}(s-3) & \cdots & \boldsymbol{\Gamma}(0) \end{bmatrix}^{-1} \\
&\begin{bmatrix} \boldsymbol{\Gamma}(1) \\ \boldsymbol{\Gamma}(2) \\ \vdots \\ \boldsymbol{\Gamma}(s-1) \end{bmatrix}, \qquad (118)
\end{aligned}$$

where $\boldsymbol{\Gamma}(k) = Cov\{\mathbf{Z}_t, \mathbf{Z}_{t+k}\}$. Note that $Var(\mathbf{U}_{s-1,t+s}) = \mathbf{C}_{\mathbf{UU}}(s)$ and $Var(\mathbf{V}_{s-1,t}) = \mathbf{C}_{\mathbf{VV}}(s)$. Thus, the *partial lag autocorrelation matrix at lag s* is

$$\mathbf{P}(s) = [\mathbf{D}_{\mathbf{V}}(s)]^{-1}\mathbf{C}_{\mathbf{VU}}(s)[\mathbf{D}_{\mathbf{U}}(s)]^{-1}, \qquad (119)$$

where $\mathbf{D}_{\mathbf{V}}(s)$ is the diagonal matrix in which the ith diagonal element is the square root of the ith diagonal element of $\mathbf{C}_{\mathbf{VV}}(s)$ and $\mathbf{D}_{\mathbf{U}}(s)$ is similarly defined for $\mathbf{C}_{\mathbf{UU}}(s)$.

Tiao and Box (1981) defined the *partial autoregression matrix at lag s*, denoted by $\boldsymbol{\Phi}_{s,s}$, to be the last matrix coefficient when the data is fitted to a vector AR process of order s. It can be shown that

$$\boldsymbol{\Phi}_{s,s} = \mathbf{C}'_{\mathbf{VU}}(s)[\mathbf{D}_{\mathbf{V}}(s)]^{-1}. \qquad (120)$$

Ansley and Newbold (1979) defined the *multivariate partial autocorrelation matrix at lag s* to be

$$\mathbf{Q}(s) = [\mathbf{W}_{\mathbf{U}}(s)]^{-1}\mathbf{C}'_{\mathbf{VU}}(s)[\mathbf{W}_{\mathbf{V}}(s)]^{-1}, \qquad (121)$$

where $\mathbf{W}_{\mathbf{U}}(s)$ and $\mathbf{W}_{\mathbf{V}}(s)$ are the symmetric square roots of $\mathbf{C}_{\mathbf{UU}}(s)$ and $\mathbf{C}_{\mathbf{VV}}(s)$, defined such that $|\mathbf{W}_{\mathbf{U}}(s)|^2 = \mathbf{C}_{\mathbf{UU}}(s)$ and $|\mathbf{W}_{\mathbf{V}}(s)|^2 = \mathbf{C}_{\mathbf{VV}}(s)$. However, it should be noted that although $\mathbf{P}(s)$, $\boldsymbol{\Phi}_{s,s}$ and $\mathbf{Q}(s)$ all share the same cut-off property for vector $AR(s)$ models, the elements of $\mathbf{P}(s)$ are proper correlation coefficients but those of $\boldsymbol{\Phi}_{s,s}$ and $\mathbf{Q}(s)$ are not correlation coefficients, with the exception of when $m = 1$—that is, except in the univariate case in which $\mathbf{P}(s) = \boldsymbol{\Phi}_{s,s} = \mathbf{Q}(s)$.

Vector Autoregressive, Vector Moving Average, and Vector Autoregressive Moving Average Models

STATIONARY VECTOR TIME SERIES MODELS

A m-dimensional stationary vector time series process $\mathbf{Z}_t$ can always be written as a linear combination of a sequence of vector white noises—that is,

$$\begin{aligned}
\mathbf{Z}_t &= \boldsymbol{\mu} + \mathbf{a}_t + \boldsymbol{\psi}_1\mathbf{a}_{t-1} + \boldsymbol{\psi}_2\mathbf{a}_{t-2} + \cdots \\
&= \boldsymbol{\mu} + \sum_{k=0}^{\infty} \boldsymbol{\psi}_k\mathbf{a}_{t-k}, \qquad (122)
\end{aligned}$$

where the $\mathbf{A}_t$ is a sequence of m-dimensional white noise processes with mean $\mathbf{0}$ vector and covariance matrix function

$$E(\mathbf{a_t}\mathbf{a}'_{t+k}) = \begin{cases} \boldsymbol{\Sigma}, \text{ if } k = 0, \\ \mathbf{0}, \text{ if } k \neq 0, \end{cases} \qquad (123)$$

and $\boldsymbol{\Sigma}$ is a $m \times m$ symmetric positive definite matrix. $\boldsymbol{\psi}_0 = \mathbf{I}$ is the $m \times m$ identity matrix and the $\boldsymbol{\psi}_k$ is a sequence of absolutely summable $m \times m$ coefficient matrices in the sense that if we let $\boldsymbol{\psi}_k = [\psi_{ij,k}]$, then each of the $m \times m$ sequences $\psi_{ij,k}$ is absolutely summable—that is, $\sum_{k=0}^{\infty} |\psi_{ij,k}| < \infty$ for $i = 1, \ldots, m$ and $j = 1, \ldots, m$. The Equation 122 is known as the *vector moving average* (*VMA*) *representation*. A vector time series process $\mathbf{Z}_t$ is said to be invertible if it can be written as a *vector autoregressive* (*VAR*) *representation*

$$\begin{aligned}
\dot{\mathbf{Z}}_t &= \boldsymbol{\Pi}_1\dot{\mathbf{Z}}_{t-1} + \boldsymbol{\Pi}_2\dot{\mathbf{Z}}_{t-2} + \cdots + \mathbf{a}_t \\
&= \sum_{k=1}^{\infty} \boldsymbol{\Pi}_k\dot{\mathbf{Z}}_{t-k} + \mathbf{a}_t, \qquad (124)
\end{aligned}$$

so that the sequence of $m \times m$ AR coefficient matrices $\boldsymbol{\Pi}_k$ is absolutely summable where $\dot{\mathbf{Z}}_t = \mathbf{Z}_t - \boldsymbol{\mu}$.

A useful class of parsimonious vector time series models is the *vector autoregressive moving average* (*VARMA*) *process*

$$\boldsymbol{\Phi}_p(B)\dot{\mathbf{Z}}_t = \boldsymbol{\Theta}_q(B)\mathbf{a}_t, \qquad (125)$$

where $\boldsymbol{\Phi}_p(B) = \boldsymbol{\Phi}_0 - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p$ and $\boldsymbol{\Theta}_q(B) = \boldsymbol{\Theta}_0 - \boldsymbol{\Theta}_1 B - \cdots - \boldsymbol{\Theta}_q B^q$ are AR and MA matrix polynomials of order p and q, respectively; $\boldsymbol{\Phi}_0$ and $\boldsymbol{\Theta}_0$ are nonsingular $m \times m$ matrices; and the $\mathbf{A}_t$ is a sequence of m-dimensional white noise processes with mean zero vector and positive definite variance–covariance matrix $\boldsymbol{\Sigma}$. Because one can always invert $\boldsymbol{\Phi}_0$ and $\boldsymbol{\Theta}_0$ and combine them into $\boldsymbol{\Sigma}$, with no loss of generality, we will assume in the following discussion that $\boldsymbol{\Phi}_0 = \boldsymbol{\Theta}_0 = \mathbf{I}$, the $m \times m$ identity matrix.

Following the extension of the stationary univariate time series models of the Univariate Time Series Models section, we have the following stationary vector time series models.

1. *VAR*(p) *models*:

$$(\mathbf{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p)\dot{\mathbf{Z}}_t = \mathbf{a}_t, \qquad (126)$$

where the zeros of $\left|\mathbf{I} - \mathbf{\Phi}_1 B - \cdots - \mathbf{\Phi}_p B^p\right|$ lie outside of the unit circle or, equivalently, the roots of $\left|\lambda^p \mathbf{I} - \lambda^{p-1}\mathbf{\Phi}_1 - \cdots - \mathbf{\Phi}_p\right| = 0$ are all inside of the unit circle.

2. *VMA*(q) *models*:

$$\dot{\mathbf{Z}}_t = (\mathbf{I} - \mathbf{\Theta}_1 B - \cdots - \mathbf{\Theta}_q B^q)\mathbf{a}_t. \tag{127}$$

3. *VARMA*(p, q) *models*:

$$(\mathbf{I} - \mathbf{\Phi}_1 B - \cdots - \mathbf{\Phi}_p B^p)\dot{\mathbf{Z}}_t = (\mathbf{I} - \mathbf{\Theta}_1 B - \cdots - \mathbf{\Theta}_p B^p)\mathbf{a}_t. \tag{128}$$

Consider the following *VAR*(1) model,

$$(\mathbf{I} - \mathbf{\Phi} B)\dot{\mathbf{Z}}_t = \mathbf{a}_t, \tag{129}$$

or

$$\dot{\mathbf{Z}}_t = \mathbf{\Phi}\dot{Z}_{t-1} = \mathbf{a}_t, \tag{130}$$

where the $\mathbf{A}_t$ is a m-dimensional white noise process with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$. For $m = 2$, we have

$$\begin{bmatrix} \dot{Z}_{1,t} \\ \dot{Z}_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} \dot{Z}_{1,t-1} \\ \dot{Z}_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}, \tag{131}$$

or

$$\begin{aligned} \dot{Z}_{1,t} &= \phi_{11}\dot{Z}_{1,t-1} + \phi_{12}\dot{Z}_{2,t-1} + a_{1,t} \\ \dot{Z}_{2,t} &= \phi_{21}\dot{Z}_{1,t-1} + \phi_{22}\dot{Z}_{2,t-1} + a_{2,t}. \end{aligned} \tag{132}$$

Thus, apart from current shocks, each $\dot{Z}_{i,t}$ depends not only on its past values of $\dot{Z}_{i,t}$ but also the past values of other variables $\dot{Z}_{j,t}$. For the *VAR*(1) to be stationary, the zero of the determinant equation $|\mathbf{I} - \mathbf{\Phi} B|$ must be outside the unit circle or the eigenvalues of $|\lambda\mathbf{I} - \mathbf{\Phi}| = 0$ are inside the unit circle.

It is important that one should not conclude from Equation 132 that there is no contemporaneous relationship between $\dot{Z}_{1,t}$ and $\dot{Z}_{2,t}$. In the form of VARMA models, because of our choice of $\mathbf{\Phi}_0 = \mathbf{\Theta}_0 = \mathbf{I}$, the contemporaneous relationship between components of vector series is modeled through the off-diagonal elements of $\mathbf{\Sigma}$.

It is also interesting to note that when $\phi_{12} = 0$ in Equation 131, then we have

$$\begin{bmatrix} 1 - \phi_{11}B & 0 \\ -\phi_{21}B & 1 - \phi_{22}B \end{bmatrix} \begin{bmatrix} \dot{Z}_{1,t} \\ \dot{Z}_{2,t} \end{bmatrix} = \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}, \tag{133}$$

or

$$\begin{cases} \dot{Z}_{1,t} = \dfrac{1}{1 - \phi_{11}B} a_{1,t}, \\ \dot{Z}_{2,t} = \dfrac{\phi_{21}B}{1 - \phi_{22}B}\dot{Z}_{1,t} + \dfrac{1}{1 - \phi_{22}B} a_{2,t}. \end{cases} \tag{134}$$

Thus, the model can be reduced to a transfer function type of model. However, from Equation 134, we should not mistakenly think that $\dot{Z}_{2,t}$ is affected only by the past values of $\dot{Z}_{1,t}$. As pointed out earlier, the contemporaneous relationship between $\dot{Z}_{1,t}$ and $\dot{Z}_{2,t}$ is contained in the off-diagonal elements of $\mathbf{\Sigma}$. Unless $\mathbf{\Sigma}$ is a diagonal matrix, $\dot{Z}_{1,t}$ and $a_{2,t}$ are correlated, which clearly violates the fundamental assumption of the transfer function model. To make it a proper transfer function model, one needs to use some transformations so that the error term in the equation is independent of input variables. We refer readers to Wei (2006, chapter 16) for more details.

NONSTATIONARY VECTOR TIME SERIES MODELS AND COINTEGRATED PROCESSES

In univariate time series analysis, a nonstationary time series is reduced to a stationary time series by proper power transformations and differencing. They can still be used in vector time series analysis. However, it should be noted that these transformations should be applied to component series individually because not all component series can be reduced to stationary by exactly the same power transformation and the same number of differencing. To be more flexible, after applying proper power transformations to the component series, we will use the following presentation for a nonstationary vector time series model:

$$\mathbf{\Phi}_p(B)\mathbf{D}(B)\dot{\mathbf{Z}}_t = \mathbf{\Theta}_q(B)\mathbf{A}_t, \tag{135}$$

where

$$D(B) = \begin{bmatrix} (1-B)^{d_1} & 0 & \cdots & 0 & 0 \\ 0 & (1-B)^{d_2} & \ddots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \ddots & & 0 \\ 0 & \cdots & & 0 & (1-B)^{d_m} \end{bmatrix}. \tag{136}$$

In many applications, the $d_i's$ in Equation 136 may be equal. In this case, one needs to be very careful in constructing a vector time series model because its component series could be cointegrated. A vector time series is said to be cointegrated if each component series is nonstationary but some linear combinations of them become stationary. For example, consider the following two-dimensional *VAR*(1) process

$$\begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\phi & 0 \end{bmatrix} \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}. \tag{137}$$

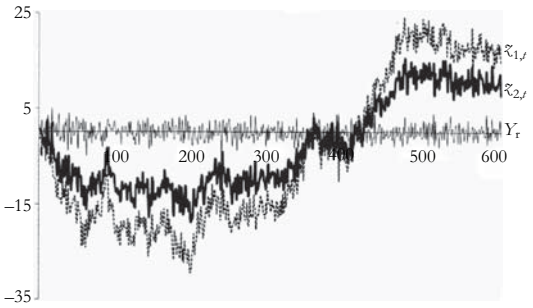

Figure 22.11 A cointegrated process where $Z_{1,t}$ and $Z_{2,t}$ are each nonstationary, but their linear combination $Y_t = 0.6Z_{1,t} + Z_{2,t}$ is stationary.

Clearly, the component $Z_{1,t} = Z_{1,t-1} + a_{1,t}$ is a random walk, which is nonstationary. For the component $Z_{2,t}$, we have $Z_{2,t} = -\phi Z_{1,t-1} + a_{2,t}$, which as the sum of a constant multiple of $Z_{1,t}$ and a white noise process is also nonstationary. However, the linear combination, $Y_t = \phi Z_{1,t} + Z_{2,t} = \phi Z_{1,t} - \phi Z_{1,t-1} + a_{2,t} = \phi a_{1,t} + a_{2,t}$, is stationary. Hence, $Z_{1,t}$ and $Z_{2,t}$ are cointegrated. Figure 22.11 illustrates the phenomenon.

For a cointegrated nonstationary vector process $\mathbf{Z}_t$, one cannot consider its differences, $\Delta\mathbf{Z}_t$ where $\Delta = (1 - B)$, and build a model only in terms of the differences. In other words, its AR representation in terms of only its differences, $(\mathbf{I} - \mathbf{\Phi}_1 B - \cdots - \mathbf{\Phi}_p B^p)\Delta\mathbf{Z}_t = \mathbf{a}_t$, does not exist for any p. The vector AR representation of a *cointegrated process* must be in terms of $\mathbf{Z}_t$ directly. If a presentation using its differences, $\Delta\mathbf{Z}_t$, and lagged values, $\Delta\mathbf{Z}_j$ for $j < t$, is preferable, one has to use the error-correction representation by including an error-correction term in the model. For more details, we refer interested readers to Granger (1986), Engle and Granger (1987), and Wei (2006, Chapters 16 and 17).

Vector Time Series Model Building

IDENTIFICATION OF VECTOR TIME SERIES MODELS

In constructing a vector time series model, just like in univariate time series model building, the first step is to plot the vector time series, as shown in Figure 22.12, for the vector series of sales and advertizing expenditures of a company. By plotting all of the component series in one graph, we obtain a good idea of the movements of different components and the general pattern of their relationships. In principle, vector time series model-building procedure is similar to the univariate time series model building procedure discussed in the Time Series Model Building section. We identify an underlying model from its correlation and partial correlation matrix functions. Table 22.10 gives a useful summary.

Given an observed vector time series $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$, we compute its sample correlation and partial correlation matrices after proper transformations are applied to reduce a nonstationary series to a stationary series.

Sample Correlation Matrix Function

The *sample correlation matrix function* is computed as

$$\hat{\boldsymbol{\rho}}(k) = [\hat{\rho}_{ij}(k)], \tag{138}$$

where the $\hat{\rho}_{ij}(k)$ is the sample cross-correlation function for the ith and jth component series,

$$\hat{\rho}_{ij}(k) = \frac{\sum_{t=1}^{n-k}(Z_{i,t} - \overline{Z}_i)(Z_{j,t+k} - \overline{Z}_j)}{[\sum_{t=1}^{n}(Z_{i,t} - \overline{Z}_i)^2 \sum_{t=1}^{n}(Z_{j,t} - \overline{Z}_j)^2]^{1/2}}, \tag{139}$$

and $\overline{Z}_i$ and $\overline{Z}_j$ are the sample means of the corresponding component series. For a stationary vector process, Hannan (1970, p. 228) showed that $\hat{\boldsymbol{\rho}}(k)$ is a consistent estimator that is asymptotically normally distributed. When the vector process is white noise, we have

$$Cov[\hat{\rho}_{ij}(k), \hat{\rho}_{ij}(k+s)] \approx \frac{1}{(n-k)}, \tag{140}$$

and

$$Var[\hat{\rho}_{ij}(k)] \approx \frac{1}{(n-k)}. \tag{141}$$

For large samples, $(n - k)$ is often replaced by n in the above expressions.

Sample Partial Lag Correlation Matrix Function

The *sample partial lag correlation matrices*, denoted by $\hat{\mathbf{P}}(s)$, are obtained by using $\hat{\mathbf{\Gamma}}(j)$ in place of $\mathbf{\Gamma}(j)$ for $j = 0, 1, \ldots, (s-1)$ in $\mathbf{P}(s)$, as shown in Equations 118 and 119. Because $\hat{\mathbf{P}}(s)$ is a proper correlation matrix, the results of sample correlation matrices can be used for its inference. Specifically, the elements of $\hat{\mathbf{P}}(s)$, denoted by $\hat{p}_{ij}(s)$, are independent and asymptotically normally distributed with mean 0 and variance *1/n*. Thus,

$$X(s) = n\sum_{i=1}^{m}\sum_{j=1}^{m}[\hat{p}_{ij}(s)]^2 \tag{142}$$

is asymptotically distributed as a χ^2 with m^2 degrees of freedom.

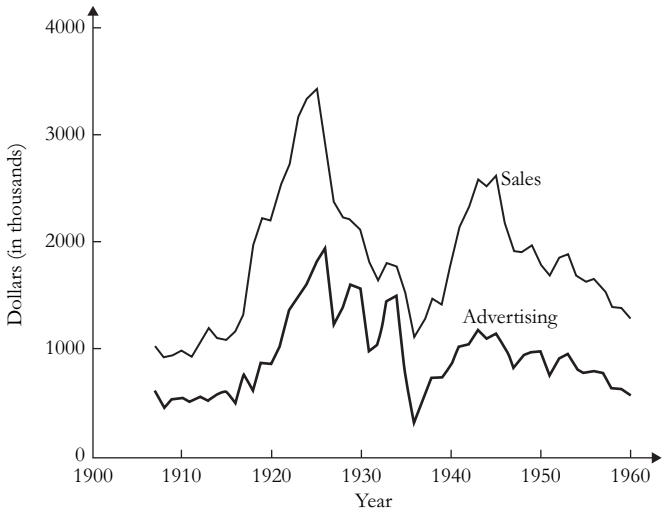

Figure 22.12 Example of a vector time series.

Table 22.10. Characteristics of Stationary Vector Time Series Models

Process	Correlation matrix function	Partial correlation matrix function
VAR(p)	Non-zero matrix with diminishing elements	Zero matrix after lag p
VMA(q)	Zero matrix after lag q	Non-zero matrix with diminishing elements
VARMA(p, q)	Non-zero matrix with diminishing elements	Non-zero matrix with diminishing elements

PARAMETER ESTIMATION, DIAGNOSTIC CHECKING, AND FORECASTING

Once a tentative model is identified, efficient estimates of the parameter matrices $\mathbf{\Phi}_i$, $\mathbf{\Theta}_j$, and $\mathbf{\Sigma}$ are obtained using a maximum likelihood method that is available in many statistical packages such as SAS, SCA, and SPSS. The adequacy of the fitted model can be checked through a careful analysis of the residuals

$$\hat{\mathbf{a}}_t = \dot{\mathbf{Z}}_t - \hat{\mathbf{\Phi}}_1\dot{\mathbf{Z}}_{t-1} - \cdots - \hat{\mathbf{\Phi}}_p\dot{\mathbf{Z}}_{t-p} + \hat{\mathbf{\Theta}}_1\hat{\mathbf{A}}_{t-1} + \cdots + \hat{\mathbf{\Theta}}_q\hat{\mathbf{a}}_{t-q}, \qquad (143)$$

where $\dot{\mathbf{Z}}_t$ is now used to denote $\mathbf{Z}_t$ if $\boldsymbol{\mu} = 0$ and $(\mathbf{Z}_t - \hat{\boldsymbol{\mu}})$ otherwise. For an adequate model, the sequence of residual vectors should behave as a vector white noise process.

After residual analysis, if the model is adequate, then it can be used for forecasting future values. For the general model in Equation 128, the ℓ − step ahead forecast at time n is given by,

$$\hat{\dot{\mathbf{Z}}}_n(\ell) = \hat{\mathbf{\Phi}}_1\hat{\dot{\mathbf{Z}}}_n(\ell-1) + \cdots + \hat{\mathbf{\Phi}}_p\hat{\dot{\mathbf{Z}}}_n(\ell-p) + \hat{\mathbf{a}}_n(\ell) - \hat{\mathbf{\Theta}}_1\hat{\mathbf{a}}_n(\ell-1) - \cdots - \hat{\mathbf{\Theta}}_q\hat{\mathbf{a}}_n(\ell-q), \qquad (144)$$

where $\hat{\dot{\mathbf{Z}}}_n(j) = \dot{\mathbf{Z}}_{n+j}$ for $j \leq 0, \hat{\mathbf{a}}_{n+j} = \mathbf{0}$ for $j > 0$, and $\hat{\mathbf{a}}_{n+j} = \mathbf{a}_{n+j}$ when $j \leq 0$. It can also be used for inference and control using the estimates of parameters and the relationship presented in the vector model.

Concluding Remarks and Future Directions

In this chapter, we have discussed many useful time domain methods and their applications in time series analysis. These include AR, MA, ARMA, and ARIMA models, intervention models, outlier detection, transfer function models, time series regression, GARCH model, vector time series models, cointegrated processes, and their iterative model-building processes and applications. Although most time series data used for our illustrations are from business and social sciences, these models and methods are general statistical methodology and can be used in any field where time series analysis is needed.

We do not cover state space models, fractional differencing, and nonlinear time series models, but with the background provided in this chapter, readers should be able to pick up these topics on their own without difficulty.

After making a conjecture or a proposition about the underlying phenomenon, a researcher often wants to test his or her proposition against observable data. In a time series study, this time series data set may be available in many forms. For example, one can choose a weekly data set, a monthly data set, or a quarterly data set. What time unit should we use in the analysis? Does the time unit chosen make a difference? It is important to point out that the same time unit should be used in both the underlying proposition and the data analysis. This normally would not be a problem in a setting where controlled experiments are possible. However, in many social science studies, a controlled experiment may not be possible, and data are often available only through aggregation or systematic sampling. In such a case, one must be very careful and aware of the consequences of aggregation and systematic sampling on model structure, parameter estimation, and forecasting discussed in Wei (2006, Chapter 20) and other references therein.

In some studies, there is a natural time unit to be used in the analysis. For example, regarding patient care at a hospital, a doctor using a certain medication will monitor a patient at certain time intervals (such as hourly) simply based on the instructions of the drug company. In some studies, there may be no apparent natural time unit, and data are available in different intervals. This is often true in many empirical studies. Natural issues to address in such cases are whether there is a best time unit to be used in the analysis and, if there is, how to determine what it is. These questions are challenging and their answers remain to be discovered.

Because of high-speed internet and the power and speed of the new generation of computers, a researcher is facing some very challenging phenomena. First, he/she has to deal with an ever increasing amount of data. To find useful information and hidden patterns underlying the data, a researcher may use various data-mining methods and techniques. Adding a time dimension to these large databases certainly introduces new aspects and challenges. In the process, one may also encounter cases where the underlying distribution is non normal, which was often assumed to be the underlying distribution for most traditional time series models.

Appendix

Summary Table of Some Commonly Used Terms, Notations, and Equations in Time Series Analysis

Terms	Notations	Equations
Mean function	μ_t	$\mu_t = E(Z_t) = \mu$ for a stationary process
Variance function	σ_t^2	$\sigma_t^2 = Var(Z_t) = E(Z_t - \mu)^2 = \sigma^2$ for a stationary process
Auto*correlation function*	*ACF*	$\rho_k = \frac{Cov(Z_t, Z_{t+k})}{\sqrt{Var(Z_t)}\sqrt{Var(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0}$
Partial autocorrelation function	PACF	$\varphi_{kk} = Corr(Z_t, Z_{t+k} \mid Z_{t+1}, \ldots, Z_{t+k-1})$
White noise process	a_t	$Z_t = a_t$
Autoregressive model of order p	$AR(p)$	$Z_t = \theta_0 + \varphi_1 Z_{t-1} + \cdots + \varphi_p Z_{t-p} + a_t$ $\varphi_p(B) Z_t = \theta_0 + a_t$ $\varphi_p(B) = 1 - \varphi_1 B - \cdots - \varphi_p B^p$
Moving average model of q	$MA(q)$	$Z_t = \mu + a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}$ $Z_t = \mu + \theta_q(B) a_t$ $\theta_q(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$
Autoregressive moving average model of order (p, q)	$ARMA(p, q)$	$\varphi_p(B) Z_t = \theta_0 + \theta_q(B) a_t$

Appendix (Continued)

Summary Table of Some Commonly Used Terms, Notations, and Equations in Time Series Analysis

Terms	Notations	Equations
Autoregressive integrated moving average model order (p, d, q)	$ARIMA(p, d, q)$	$\varphi_p(B)(1-B)^d Z_t = \theta_0 + \theta_q(B)a_t$
Cross-correlation function	CCF	$\rho_{XY}(k) = \frac{\gamma_{XY}(k)}{\sigma_X \sigma_Y}$ $\gamma_{XY}(k) = E[(X_t - \mu_X)(Y_{t+k} - \mu_Y)]$
Transfer function model	ARMAX	$Y_t = \frac{\omega_s(B)}{\delta_r(B)} X_{t-b} + \frac{\theta_q(B)}{\varphi_p(B)} a_t$ $\omega_s(B) = \omega_0 - \omega_1 B - \cdots - \omega_s B^s$ $\delta_r(B) = 1 - \delta_1 B - \cdots - \delta_r B^r$
Autoregressive conditional heteroscedasticity model of order s	$ARCH(s)$	$\varepsilon_t = \sigma_t e_t$ where $e_t \sim i.i.d.\, N(0,1)$ and $\sigma_t^2 = \theta_0 + \theta_1 \varepsilon_{t-1}^2 + \cdots + \theta_s \varepsilon_{t-s}^2$
Generalized autoregressive conditional heteroscedasticity model of order (r, s)	$GARCH(r, s)$	$\varepsilon_t = \sigma_t e_t$ where $e_t \sim i.i.d.\, N(0,1)$ and $\sigma_t^2 = \theta_0 + \phi_1 \sigma_{t-1}^2 + \cdots + \phi_r \sigma_{t-r}^2$ $+ \theta_1 \varepsilon_{t-1}^2 + \cdots + \theta_s \varepsilon_{t-s}^2$
Vector autoregressive model of order p	$VAR(p)$	$(\mathbf{I} - \mathbf{\Phi}_1 B - \cdots - \mathbf{\Phi}_p B^p)\dot{\mathbf{Z}}_t = \mathbf{A}_t$ $\dot{\mathbf{Z}}_t = \mathbf{Z}_t - \boldsymbol{\mu}$
Vector moving average model of order q	$VMA(q)$	$\dot{\mathbf{Z}}_t = (\mathbf{I} - \mathbf{\Theta}_1 B - \cdots - \mathbf{\Theta}_q B^q)\mathbf{A}_t$
Vector autoregressive moving average model of order (p, q)	$VARMA(p, q)$	$(\mathbf{I} - \mathbf{\Phi}_1 B - \cdots - \mathbf{\Phi}_p B^p)\dot{\mathbf{Z}}_t = (\mathbf{I} - \mathbf{\Theta}_1 B - \cdots - \mathbf{\Theta}_q B^q)\mathbf{A}_t$

In addition to the large amount of data, one is also encountering more and more high-dimensional data sets. Traditional time series methods are not designed to deal with these kinds of high-dimensional variables. Even with today's computer power and speed, there are many difficult problems that remain to be solved. As most statistical methods are developed for a random sample, the use of highly correlated time series data certainly introduces a new set of complications and challenges.

Acknowledgments

The authors thank the referees and the editor for their helpful comments and suggestions, which have improved the presentation of this paper.

Author Note

William W.S. Wei, Department of Statistics, Temple University, 1810 North 13th Street, Philadelphia, PA 19122, USA, Email: wwei@temple.edu.

References

Akaike, H. (1974). A new look at the statistical identification, *IEEE Transactions on Automatic Control*, *AC-19*, 716–723.

Ansley, C. F. & Newbold, P. (1979). Multivariate partial autocorrelations, *American Statistical Association Proceedings of Business and Economic Statistics Section,* 349–353.

Bartlett, M. S. (1946). On the theoretical specification of sampling properties of autocorrelated time series, *Journal of Royal Statistical Society, B8*, 27–41.

Box, G. E. P. & Cox, D. R. (1964). An analysis of transformations, *Journal of Royal Statistical Society, B26*, 211–252.

Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control, 4th Ed.,* New York: Wiley.

Box, G. E. P. & Tiao, G.C. (1975). Intervention analysis with applications to economic and environmental problems, *Journal of American Statistical Association, 70,* 70–79.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, *31*, 307–327.

Chan, N. H. & Wei, C. Z. (1988). Limiting distribution of least squares estimates of unstable autoregressive processes, *Annals of Statistics*, *16*, 367–401.

Chang, I., Tiao, G.C., & Chen C. (1988). Estimation of time series parameters in the presence of outliers, *Technometrics, 30* (2), 193–204.

Chen, C. & Liu, L. M. (1991). Forecasting time series with outliers, *Journal of Forecasting*, *12*, 13–35.

Dickey, D. A. & Fuller, W. A. (1979). Distribution of the estimates for autoregressive time series with a unit root, *Journal of American Statistical Association*, *74*, 427–431.

Dickey, D. A., Hasza, D. P., & Fuller, W. A. (1984). Testing for unit roots in seasonal time series, *Journal of American Statistical Association*, *79*, 355–367.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, *50*, 987–1007.

Engle, R. F. & Granger, C. W. J. (1987). Co-intrgration and error correction: representation, estimation, and testing, *Econometrica*, *55*, 251–276.

Fox, A.J. (1972). Outliers in time series, *Journal of Royal Statistical Society*, *B43*, 350–363.

Granger, C. W. J. (1986). Developments in the study of co-integrated economic variables, *Oxford Bulletin of Economics and Statistics*, *48*, 213–228.

Hannan, E.J. (1970). *Multiple Time Series*, New York: Wiley.

Heyse, J. F. & Wei, W. W. S. (1985). Inverse and partial lag autocorrelation for vector time series, *American Statistical Association Proceedings of Business and Economic Statistics Section,* 233–237.

Lee, J.H. & Wei, W.W.S. (1995). A model-independent outlier detection procedure, *Journal of Applied Statistical Science*, *2*, 345–359.

Ljung, G. M. & Box, G. E. P. (1978). On a measure of lack of fit in time series models, *Biometrika*, *65*, 297–303.

Parzen, E. (1977). Multiple time series modeling: Determining the order of approximating autoregressive schemes. In P. Krishnaiah (Ed.). *Multivariate Analysis IV* (pp. 283–295). Amsterdam, The Netherlands: North-Holland.

R Foundation (2012). The R project for statistical computing, version 2.15.0, Vienna, Austria. URL http://www.r-project.org.

SAS Institute, Inc. (2009). SAS for Windows, 9.2, Cary, North Carolina.

SCA Corp. (2008). SCA WorkBench User's Guide, Release 5.4, Villa Park, Illinois.

Schwartz, G. (1978). Estimating the dimension of a model, *Annal of Statistics*, *6*, 461–464.

SPSS, Inc. (2009). SPSS 15.0 for Windows, Chicago, Illinois.

Tiao, G. C. & Box, G. E. P. (1981). Modeling multiple time series with applications, *Journal of American Statistical Association*, *76*, 802–816.

Tsay, R. S. & Tiao, G. C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and non-stationary ARIMA models, *Journal of American Statistical Association*, *79*, 84–96.

Tsay, R. S. & Tiao, G. C. (1985). Use of canonical analysis in time series model identification, *Biometrika*, *72*(2), 299–315.

Tsay, R. S. (1988). Outliers, level shifts, and variance changes in time series. *Journal of Forecasting*, *7*, 1–22.

Wei, William W.S. (2006). *Time Series Analysis—Univariate and Multivariate Methods*, *2nd Ed.*, Boston, MA: Pearson Addison-Wesley.

Wei, William W.S. (2008). Spectral analysis. In S. Menard (Ed.). *Hand Book of Longitudinal Research* (pp. 601–620). Burlington, MA: Academic Press..