CHAPTER 7

Characterizing Cycles

We've already built forecasting models with trend and seasonal components. In this chapter, as well as the next two, we consider a crucial third component, **cycles.** When you think of a "cycle," you probably think of the sort of rigid up-and-down pattern depicted in Figure 7.1. Such cycles can sometimes arise, but cyclical fluctuations in business, finance, economics, and government are typically much less rigid. In fact, when we speak of cycles, we have in mind a much more general, all-encompassing notion of cyclicality: any sort of dynamics not captured by trends or seasonals.

Cycles, according to our broad interpretation, may display the sort of back-and-forth movement characterized in Figure 7.1, but they don't have to. All we require is that there be some dynamics, some persistence, some way in which the present is linked to the past and the future to the present. Cycles are present in most of the series that concern us, and it's crucial that we know how to model and forecast them, because their history conveys information regarding their future.

Trend and seasonal dynamics are simple, so we can capture them with simple models. Cyclical dynamics, however, are more complicated. Because of the wide variety of cyclical patterns, the sorts of models we need are substantially more involved. Thus, we split the discussion into three parts. Here in Chapter 7 we develop methods for *characterizing* cycles, in Chapter 8 we discuss *models* of cycles, and following that, in Chapter 9, we show how to use those models

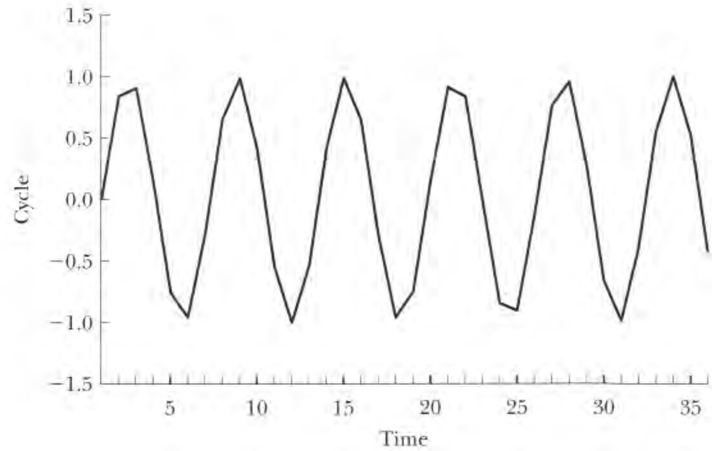


FIGURE 7.1
A Rigid Cyclical Pattern

to *forecast* cycles. All of the material is crucial to a real understanding of forecasting and forecasting models, and it's also a bit difficult the first time around because it's unavoidably rather mathematical, so careful, systematic study is required. The payoff will be large when we arrive at Chapter 10, in which we assemble and apply extensively the ideas for modeling and forecasting trends, seasonals, and cycles developed in Chapters 5–9.

1. Covariance Stationary Time Series

A **realization** of a time series is an ordered set, $\{\ldots, y_{-2}, y_{-1}, y_0, y_1, y_2, \ldots\}$. Typically the observations are ordered in time—hence the name **time series**—but they don't have to be. We could, for example, examine a spatial series, such as office space rental rates as we move along a line from a point in Midtown Manhattan to a point in the New York suburbs 30 miles away. But the most important case for forecasting, by far, involves observations ordered in time, so that's what we'll stress.

In theory, a time series realization begins in the infinite past and continues into the infinite future. This perspective may seem abstract and of limited practical applicability, but it will be useful in deriving certain very important properties of the forecasting models we'll be using soon. In practice, of course, the data we observe are just a finite subset of a realization, $\{y_1, \ldots, y_T\}$, called a **sample path.**

Shortly we'll be building forecasting models for cyclical time series. If the underlying probabilistic structure of the series were changing over time, we'd be doomed—there would be no way to predict the future accurately on the basis of the past, because the laws governing the future would differ from those governing the past. If we want to forecast a series, at a minimum we'd like its mean and its covariance structure (i.e., the covariances between current and past values) to be stable over time, in which case we say that the series is **covariance stationary.**

Let's discuss covariance stationarity in greater depth. The first requirement for a series to be covariance stationary is that the mean of the series be stable over time. The mean of the series at time t is

$$E(y_t) = \mu_t \ .$$

If the mean is stable over time, as required by covariance stationarity, then we can write

$$E(y_t) = \mu \ ,$$

for all t. Because the mean is constant over time, there's no need to put a time subscript on it.

The second requirement for a series to be covariance stationary is that its covariance structure be stable over time. Quantifying stability of the covariance structure is a bit tricky, but tremendously important, and we do it using the **autocovariance function.** The autocovariance at displacement τ is just the covariance between y_t and $y_{t-\tau}$. It will of course depend on τ, and it may also depend on t, so in general we write

$$\gamma(t, \tau) = \mathrm{cov}(y_t, y_{t-\tau}) = E(y_t - \mu)(y_{t-\tau} - \mu) \ .$$

If the covariance structure is stable over time, as required by covariance stationarity, then the autocovariances depend only on displacement, τ, not on time, t, and we write

$$\gamma(t, \tau) = \gamma(\tau) \ ,$$

for all t.

The autocovariance function is important because it provides a basic summary of cyclical dynamics in a covariance stationary series. By examining the autocovariance structure of a series, we learn about its dynamic behavior. We graph and examine the autocovariances as a function of τ. Note that the autocovariance function is symmetric; that is,

$$\gamma(\tau) = \gamma(-\tau) \ ,$$

for all τ. Typically, we'll consider only nonnegative values of τ. Symmetry reflects the fact that the autocovariance of a covariance stationary series depends only on displacement; it doesn't matter whether we go forward or backward. Note also that

$$\gamma(0) = \mathrm{cov}(y_t, y_t) = \mathrm{var}(y_t) \ .$$

There is one more technical requirement of covariance stationarity: We require that the variance of the series—the autocovariance at displacement 0, $\gamma(0)$—be finite. It can be shown that no autocovariance can be larger in absolute value than $\gamma(0)$, so if $\gamma(0) < \infty$, then so, too, are all the other autocovariances.

It may seem that the requirements for covariance stationarity are quite stringent, which would bode poorly for our forecasting models, almost all of which invoke covariance stationarity in one way or another. It is certainly true that many economic, business, financial, and government series are not covariance stationary. An upward trend, for example, corresponds to a steadily increasing mean, and seasonality corresponds to means that vary with the season, both of which are violations of covariance stationarity.

But appearances can be deceptive. Although many series are not covariance stationary, it is frequently possible to work with models that give special treatment to nonstationary components such as trend and seasonality, so that the cyclical component that's left over is likely to be covariance stationary. We'll often adopt that strategy. Alternatively, simple transformations often appear to transform nonstationary series to covariance stationarity. For example, many series that are clearly nonstationary in levels appear covariance stationary in growth rates.

In addition, note that although covariance stationarity requires means and covariances to be stable and finite, it places no restrictions on other aspects of the distribution of the series, such as skewness and kurtosis.[1] The upshot is simple: Whether we work directly in levels and include special components for the nonstationary elements of our models, or we work on transformed data such as growth rates, the covariance stationarity assumption is not as unrealistic as it may seem.

Recall that the correlation between two random variables x and y is defined by

$$\mathrm{corr}(x, y) = \frac{\mathrm{cov}(x, y)}{\sigma_x \sigma_y}.$$

That is, the correlation is simply the covariance, "normalized" or "standardized," by the product of the standard deviations of x and y. Both the correlation and the covariance are measures of linear association between two random variables. The correlation is often more informative and easily interpreted, however, because the construction of the correlation coefficient guarantees that $\mathrm{corr}(x, y) \in [-1, 1]$, whereas the covariance between the same two random variables may take any value. The correlation, moreover, does not depend on the units in which x and y are measured, whereas the covariance does. Thus, for example, if x and y have a covariance of 10 million, they're not necessarily very strongly associated, whereas if they have a correlation of .95, it is unambiguously clear that they are very strongly associated.

In light of the superior interpretability of correlations as compared with covariances, we often work with the correlation, rather than the covariance, between y_t and $y_{t-\tau}$. That is, we work with the **autocorrelation function,** $\rho(\tau)$, rather than the autocovariance function, $\gamma(\tau)$. The autocorrelation function is obtained by dividing the autocovariance function by the variance,

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}, \quad \tau = 0, 1, 2, \ldots.$$

[1] For that reason, covariance stationarity is sometimes called **second-order stationarity** or **weak stationarity.**

The formula for the autocorrelation is just the usual correlation formula, specialized to the correlation between y_t and $y_{t-\tau}$. To see why, note that the variance of y_t is $\gamma(0)$, and by covariance stationarity, the variance of y at any other time $y_{t-\tau}$ is also $\gamma(0)$. Thus,

$$\rho(\tau) = \frac{\operatorname{cov}(y_t, y_{t-\tau})}{\sqrt{\operatorname{var}(y_t)}\sqrt{\operatorname{var}(y_{t-\tau})}} = \frac{\gamma(\tau)}{\sqrt{\gamma(0)}\sqrt{\gamma(0)}} = \frac{\gamma(\tau)}{\gamma(0)},$$

as claimed. Note that we always have $\rho(0) = \frac{\gamma(0)}{\gamma(0)} = 1$, because any series is perfectly correlated with itself. Thus, the autocorrelation at displacement 0 isn't of interest; rather, only the autocorrelations *beyond* displacement 0 inform us about a series' dynamic structure.

Finally, the **partial autocorrelation function,** $p(\tau)$, is sometimes useful. $p(\tau)$ is just the coefficient of $y_{t-\tau}$ in a population linear regression of y_t on $y_{t-1}, \ldots, y_{t-\tau}$.[2] We call such a regression an **autoregression,** because the variable is regressed on lagged values of itself. It's easy to see that the autocorrelations and partial autocorrelations, although related, differ in an important way. The autocorrelations are just the "simple" or "regular" correlations between y_t and $y_{t-\tau}$. The partial autocorrelations, on the other hand, measure the association between y_t and $y_{t-\tau}$ after *controlling* for the effects of $y_{t-1}, \ldots, y_{t-\tau+1}$; that is, they measure the partial correlation between y_t and $y_{t-\tau}$.

As with the autocorrelations, we often graph the partial autocorrelations as a function of τ and examine their qualitative shape, which we'll do soon. Like the autocorrelation function, the partial autocorrelation function provides a summary of a series' dynamics, but as we'll see, it does so in a different way.[3]

All of the covariance stationary processes that we will study subsequently have autocorrelation and partial autocorrelation functions that approach 0, one way or another, as the displacement gets large. In Figure 7.2 we show an autocorrelation function that displays gradual one-sided damping, and in Figure 7.3 we show a constant autocorrelation function; the latter could not be the autocorrelation function of a stationary process, whose autocorrelation function must eventually decay. The precise decay patterns of autocorrelations and partial autocorrelations of a covariance stationary series, however, depend on the specifics of the series, as we'll see in detail in the next chapter. In Figure 7.4, for example, we show an autocorrelation function that displays damped oscillation—the autocorrelations are positive at first, then become negative for a while, then positive again, and so on, while continuously getting

[2] To get a feel for what we mean by "population regression," imagine that we have an infinite sample of data at our disposal, so that the parameter estimates in the regression are not contaminated by sampling variation; that is, they're the true population values. The thought experiment just described is a **population regression.**

[3] Also in parallel to the autocorrelation function, the partial autocorrelation at displacement 0 is always 1 and is therefore uninformative and uninteresting. Thus, when we graph the autocorrelation and partial autocorrelation functions, we'll begin at displacement 1 rather than displacement 0.

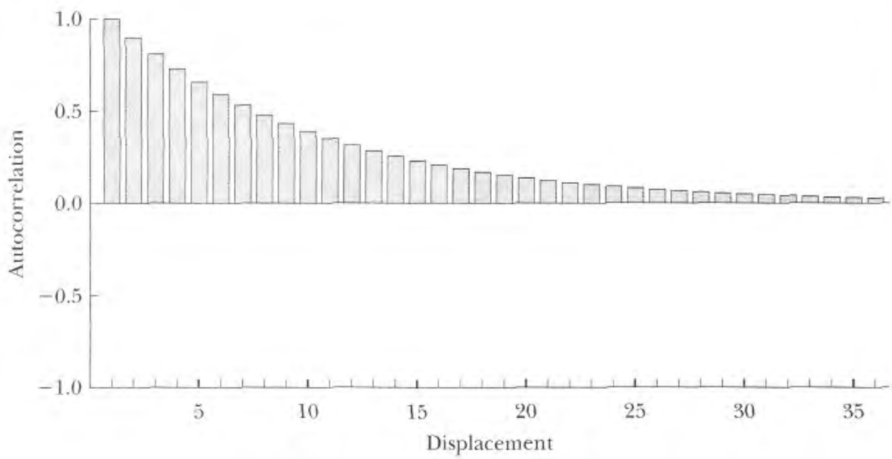


FIGURE 7.2 *Autocorrelation Function, One-Sided Gradual Damping*

smaller in absolute value. Finally, in Figure 7.5 we show an autocorrelation function that differs in the way it approaches 0—the autocorrelations drop abruptly to 0 beyond a certain displacement.

2. White Noise

In this section and throughout the next chapter, we'll study the population properties of certain time series models, or **time series processes,** which are very important for forecasting. Before we estimate time series forecasting models, we need to understand their population properties, assuming that the

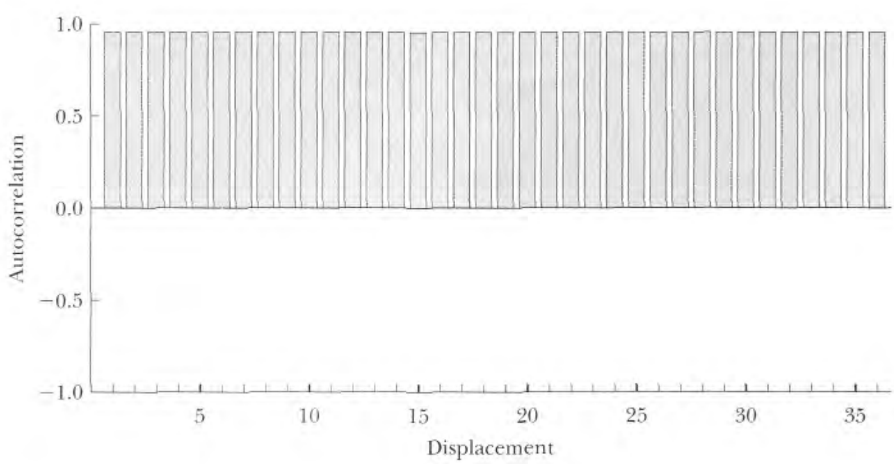
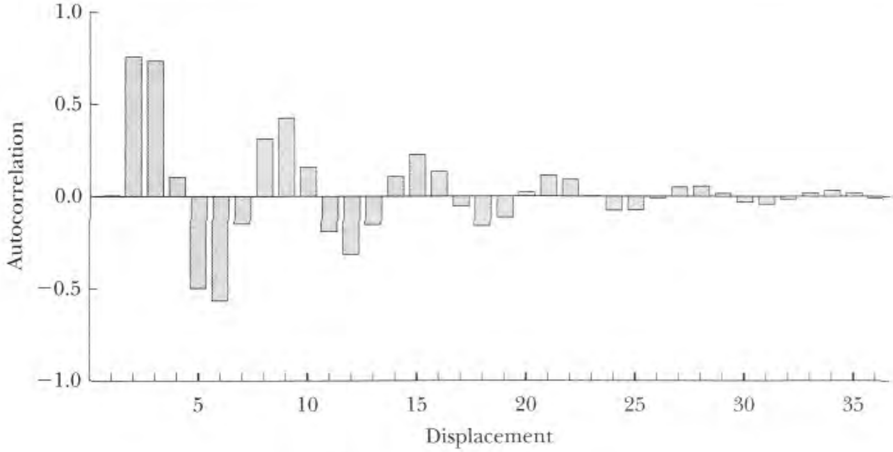


FIGURE 7.3 *Autocorrelation Function, Nondamping*

FIGURE 7.4
Autocorrelation Function, Gradual Damped Oscillation

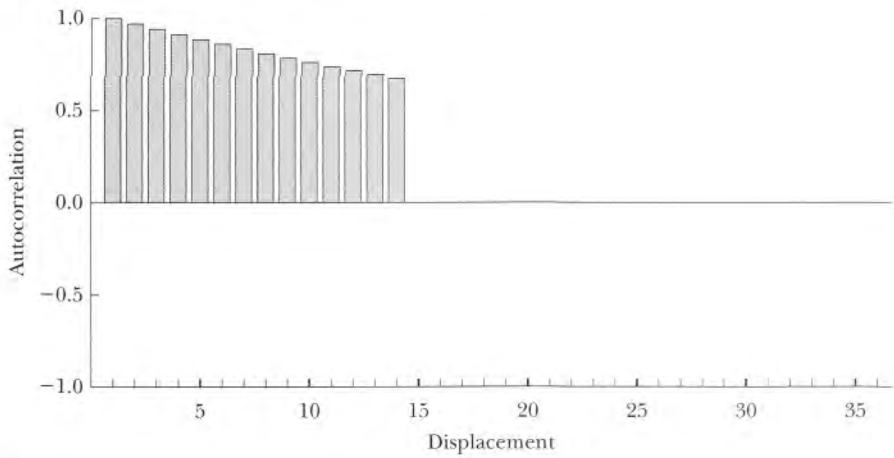


postulated model is true. The simplest of all such time series processes is the fundamental building block from which all others are constructed. In fact, it's so important that we introduce it now. We use y to denote the observed series of interest. Suppose that

$$y_t = \varepsilon_t$$
$$\varepsilon_t \sim (0, \sigma^2)\ ,$$

where the "shock," ε_t, is uncorrelated over time. We say that ε_t, and hence y_t, is **serially uncorrelated.** Throughout, unless explicitly stated otherwise, we assume that $\sigma^2 < \infty$. Such a process, with zero mean, constant variance, and

FIGURE 7.5
Autocorrelation Function, Sharp Cutoff



no serial correlation, is called **zero-mean white noise,** or simply **white noise.**[4] Sometimes for short we write

$$\varepsilon_t \sim WN(0, \sigma^2)$$

and hence

$$y_t \sim WN(0, \sigma^2) .$$

Note that, although ε_t and hence y_t are serially uncorrelated, they are not necessarily serially independent, because they are not necessarily normally distributed.[5] If in addition to being serially uncorrelated, y is serially independent, then we say that y is **independent white noise.**[6] We write

$$y_t \overset{\text{iid}}{\sim} (0, \sigma^2) ,$$

and we say that "y is independently and identically distributed with zero mean and constant variance." If y is serially uncorrelated and normally distributed, then it follows that y is also serially independent, and we say that y is **normal white noise** or **Gaussian white noise.**[7] We write

$$y_t \overset{\text{iid}}{\sim} N(0, \sigma^2) .$$

We read "y is independently and identically distributed as normal, with zero mean and constant variance" or simply "y is Gaussian white noise." In Figure 7.6 we show a sample path of Gaussian white noise, of length $T = 150$, simulated on a computer. There are no patterns of any kind in the series due to the independence over time.

You're already familiar with white noise, although you may not realize it. Recall that the disturbance in a regression model is typically assumed to be white noise of one sort or another. There's a subtle difference here, however. Regression disturbances are not observable, whereas we're working with an observed series. Later, however, we'll see how all of our models for observed series can be used to model unobserved variables such as regression disturbances.

Let's characterize the dynamic stochastic structure of white noise, $y_t \sim WN(0, \sigma^2)$. By construction the unconditional mean of y is

$$E(y_t) = 0 ,$$

and the unconditional variance of y is

$$\text{var}(y_t) = \sigma^2 .$$

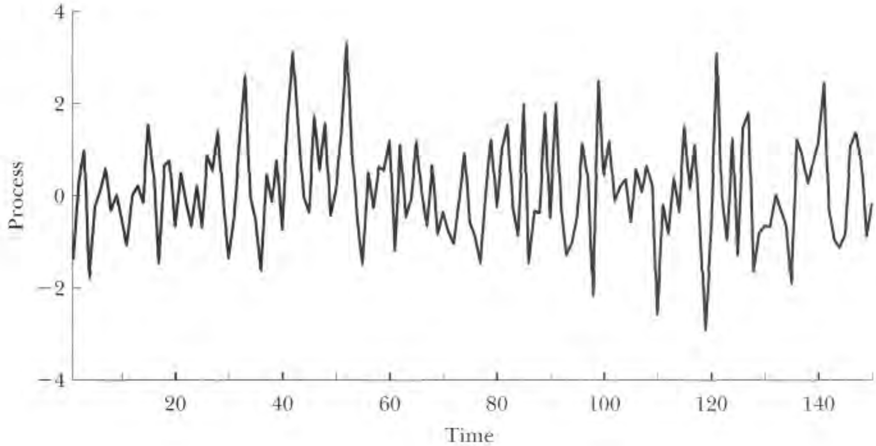
[4] It's called white noise by analogy with white light, which is composed of all colors of the spectrum, in equal amounts. We can think of white noise as being composed of a wide variety of cycles of differing periodicities, in equal amounts.

[5] Recall that zero correlation implies independence only in the normal case.

[6] Another name for independent white noise is **strong white noise,** in contrast to standard serially uncorrelated **weak white noise.**

[7] Karl Friedrich Gauss, one of the greatest mathematicians of all time, discovered the normal distribution some 200 years ago—hence the adjective *Gaussian.*

FIGURE 7.6
Realization of White Noise Process



Note that the unconditional mean and variance are constant. In fact, the unconditional mean and variance must be constant for any covariance stationary process. The reason is that constancy of the unconditional mean was our first explicit requirement of covariance stationarity and that constancy of the unconditional variance follows implicitly from the second requirement of covariance stationarity—that the autocovariances depend only on displacement, not on time.[8]

To understand fully the linear dynamic structure of a covariance stationary time series process, we need to compute and examine its mean and its autocovariance function. For white noise, we've already computed the mean and the variance, which is the autocovariance at displacement 0. We have yet to compute the rest of the autocovariance function; fortunately, however, it's very simple. Because white noise is, by definition, uncorrelated over time, all the autocovariances, and hence all the autocorrelations, are 0 beyond displacement 0.[9] Formally, then, the autocovariance function for a white noise process is

$$\gamma(\tau) = \begin{cases} \sigma^2, & \tau = 0 \\ 0, & \tau \geq 1 \end{cases},$$

and the autocorrelation function for a white noise process is

$$\rho(\tau) = \begin{cases} 1, & \tau = 0 \\ 0, & \tau \geq 1 \end{cases}.$$

In Figure 7.7 we plot the white noise autocorrelation function.

Finally, consider the partial autocorrelation function for a white noise series. For the same reason that the autocorrelation at displacement 0 is always 1, so, too, is the partial autocorrelation at displacement 0. For a white noise

[8] Recall that $\sigma^2 = \gamma(0)$.

[9] If the autocovariances are all 0, so are the autocorrelations, because the autocorrelations are proportional to the autocovariances.

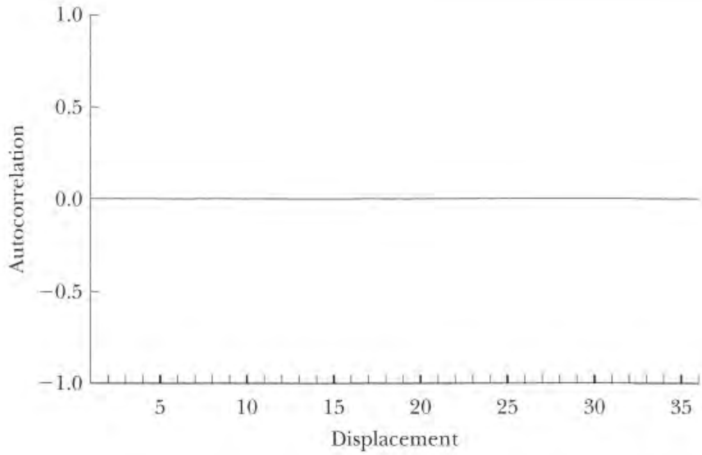


FIGURE 7.7 *Population Autocorrelation Function, White Noise Process*

process, all partial autocorrelations beyond displacement 0 are 0, which again follows from the fact that white noise, by construction, is serially uncorrelated. Population regressions of y_t on y_{t-1}, or on y_{t-1} and y_{t-2} or on any other lags, produce nothing but 0 coefficients, because the process is serially uncorrelated. Formally, the partial autocorrelation function of a white noise process is

$$p(\tau) = \begin{cases} 1, & \tau = 0 \\ 0, & \tau \geq 1 \end{cases} .$$

We show the partial autocorrelation function of a white noise process in Figure 7.8. Again, it's degenerate and exactly the same as the autocorrelation function!

By now you've surely noticed that if you were assigned the task of forecasting independent white noise, you'd likely be doomed to failure. What happens

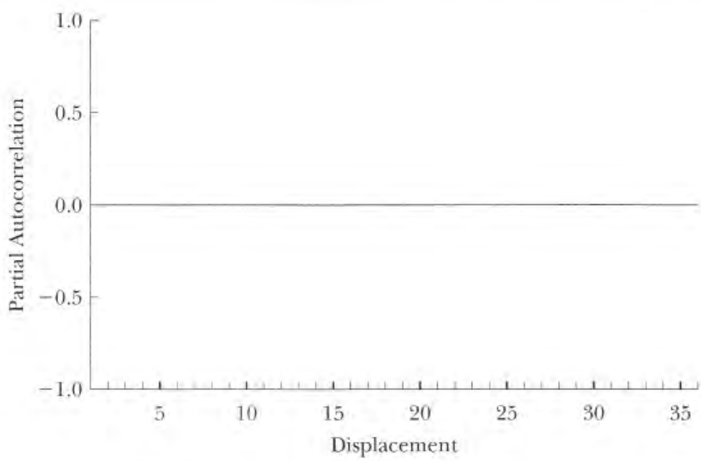


FIGURE 7.8 *Population Partial Autocorrelation Function, White Noise Process*

to a white noise series at any time is uncorrelated with anything in the past; similarly, what happens in the future is uncorrelated with anything in the present or past. But understanding white noise is tremendously important for at least two reasons. First, as already mentioned, processes with much richer dynamics are built up by taking simple transformations of white noise. Second, 1-step-ahead forecast errors from good models should be white noise. After all, if such forecast errors aren't white noise, then they're serially correlated, which means that they're forecastable; and if forecast errors are forecastable, then the forecast can't be very good. Thus, it's important that we understand and be able to recognize white noise.

Thus far we've characterized white noise in terms of its mean, variance, autocorrelation function, and partial autocorrelation function. Another characterization of dynamics, with important implications for forecasting, involves the mean and variance of a process, *conditional* on its past. In particular, we often gain insight into the dynamics in a process by examining its conditional mean, which is a key object for forecasting.[10] In fact, throughout our study of time series, we'll be interested in computing and contrasting the **unconditional mean and variance** and the **conditional mean and variance** of various processes of interest. Means and variances, which convey information about location and scale of random variables, are examples of what statisticians call **moments.** For the most part, our comparisons of the conditional and unconditional moment structure of time series processes will focus on means and variances (they're the most important moments), but sometimes we'll be interested in higher-order moments, which are related to properties such as skewness and kurtosis.

For comparing conditional and unconditional means and variances, it will simplify our story to consider independent white noise, $y_t \overset{\text{iid}}{\sim} (0, \sigma^2)$. By the same arguments as before, the unconditional mean of y is 0, and the unconditional variance is σ^2. Now consider the conditional mean and variance, where the information set Ω_{t-1} on which we condition contains either the past history of the observed series, $\Omega_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$, or the past history of the shocks, $\Omega_{t-1} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$. (They're the same in the white noise case.) In contrast to the unconditional mean and variance, which must be constant by covariance stationarity, the conditional mean and variance need not be constant, and in general we'd expect them *not* to be constant. The unconditionally expected growth of laptop computer sales next quarter may be 10%, but expected sales growth may be much higher, *conditional* on knowledge that sales grew this quarter by 20%. For the independent white noise process, the conditional mean is

$$E(y_t \mid \Omega_{t-1}) = 0 \,,$$

and the conditional variance is

$$\operatorname{var}(y_t \mid \Omega_{t-1}) = E((y_t - E(y_t \mid \Omega_{t-1}))^2 \mid \Omega_{t-1}) = \sigma^2 \,.$$

[10] If you need to refresh your memory on conditional means, consult any good introductory statistics book, such as Wonnacott and Wonnacott (1990).

Conditional and unconditional means and variances are identical for an independent white noise series; there are no dynamics in the process and hence no dynamics in the conditional moments to exploit for forecasting.

3. The Lag Operator

The **lag operator** and related constructs are the natural language in which forecasting models are expressed. If you want to understand and manipulate forecasting models—indeed, even if you simply want to be able to read the software manuals—you have to be comfortable with the lag operator. The lag operator, L, is very simple: It "operates" on a series by lagging it. Hence,

$$Ly_t = y_{t-1} .$$

Similarly,

$$L^2 y_t = L(L(y_t)) = L(y_{t-1}) = y_{t-2} ,$$

and so on. Typically we'll operate on a series not with the lag operator but with a **polynomial in the lag operator.** A lag operator polynomial of degree m is just a linear function of powers of L, up through the mth power,

$$B(L) = b_0 + b_1 L + b_2 L^2 + \cdots b_m L^m .$$

To take a very simple example of a lag operator polynomial operating on a series, consider the mth-order lag operator polynomial L^m, for which

$$L^m y_t = y_{t-m} .$$

A well-known operator, the first-difference operator Δ, is actually a first-order polynomial in the lag operator; you can readily verify that

$$\Delta y_t = (1 - L)y_t = y_t - y_{t-1} .$$

As a final example, consider the second-order lag operator polynomial $(1 + 0.9L + 0.6L^2)$ operating on y_t. We have

$$(1 + 0.9L + 0.6L^2)y_t = y_t + 0.9y_{t-1} + 0.6y_{t-2} ,$$

which is a weighted sum, or **distributed lag,** of current and past values. All forecasting models, one way or another, must contain such distributed lags, because they've got to quantify how the past evolves into the present and future; hence, lag operator notation is a useful shorthand for stating and manipulating forecasting models.

Thus far, we've considered only finite-order polynomials in the lag operator; it turns out that infinite-order polynomials are also of great interest. We write the infinite-order lag operator polynomial as

$$B(L) = b_0 + b_1 L + b_2 L^2 + \cdots = \sum_{i=0}^{\infty} b_i L^i .$$

Thus, for example, to denote an infinite distributed lag of current and past shocks, we might write

$$B(L)\varepsilon_t = b_0\varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \cdots = \sum_{i=0}^{\infty} b_i\varepsilon_{t-i} .$$

At first sight, infinite distributed lags may seem esoteric and of limited practical interest, because models with infinite distributed lags have infinitely many parameters ($b_0, b_1, b_2, \ldots$) and therefore can't be estimated with a finite sample of data. On the contrary, and surprisingly, it turns out that models involving infinite distributed lags are central to time series modeling and forecasting. Wold's theorem, to which we now turn, establishes that centrality.

4. Wold's Theorem, the General Linear Process, and Rational Distributed Lags[11]

Wold's Theorem

Many different dynamic patterns are consistent with covariance stationarity. Thus, if we know only that a series is covariance stationary, it's not at all clear what sort of model we might fit to describe its evolution. The trend and seasonal models that we've studied aren't of use; they're models of specific nonstationary components. Effectively, what we need now is an appropriate model for what's left after fitting the trend and seasonal components—a model for a covariance stationary residual. **Wold's representation theorem** points to the appropriate model.

Theorem

Let $\{y_t\}$ be any zero-mean covariance-stationary process.[12] Then we can write it as

$$y_t = B(L)\varepsilon_t = \sum_{i=0}^{\infty} b_i\varepsilon_{t-i}$$

$$\varepsilon_t \sim WN(0, \sigma^2) ,$$

where $b_0 = 1$ and $\sum_{i=0}^{\infty} b_i^2 < \infty$. In short, the correct "model" for any covariance stationary series is some infinite distributed lag of white noise, called the **Wold representation.** The ε_t's are often called **innovations,** because (as we'll see in Chapter 9) they correspond to the 1-step-ahead forecast errors that we'd make

[11] This section is a bit more abstract than others, but don't be put off. On the contrary, you may want to read it several times. The material in it is crucially important for time series modeling and forecasting and is therefore central to our concerns.

[12] Moreover, we require that the covariance stationary processes don't contain any deterministic components.

if we were to use a particularly good forecast. That is, the ε_t's represent that part of the evolution of y that's linearly unpredictable on the basis of the past of y. Note also that the ε_t's, although uncorrelated, are not necessarily independent. Again, it's only for Gaussian random variables that lack of correlation implies independence, and the innovations are not necessarily Gaussian.

In our statement of Wold's theorem we assumed a zero mean. That may seem restrictive, but it's not. Rather, whenever you see y_t, just read $y_t - \mu$, so that the process is expressed in deviations from its mean. The deviation from the mean has a zero mean, by construction. Working with zero-mean processes therefore involves no loss of generality while facilitating notational economy. We'll use this device frequently.

The General Linear Process

Wold's theorem tells us that when formulating forecasting models for covariance stationary time series, we need only consider models of the form

$$y_t = B(L)\varepsilon_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i}$$

$$\varepsilon_t \sim WN(0, \sigma^2),$$

where the b_i are coefficients with $b_0 = 1$ and $\sum_{i=0}^{\infty} b_i^2 < \infty$. We call this the **general linear process,** "general" because any covariance stationary series can be written that way, and "linear" because the Wold representation expresses the series as a linear function of its innovations.

The general linear process is so important that it's worth examining its unconditional and conditional moment structure in some detail. Taking means and variances, we obtain the unconditional moments

$$E(y_t) = E\left(\sum_{i=0}^{\infty} b_i \varepsilon_{t-i}\right) = \sum_{i=0}^{\infty} b_i E(\varepsilon_{t-i}) = \sum_{i=0}^{\infty} b_i \cdot 0 = 0$$

and

$$\mathrm{var}(y_t) = \mathrm{var}\left(\sum_{i=0}^{\infty} b_i \varepsilon_{t-i}\right) = \sum_{i=0}^{\infty} b_i^2\, \mathrm{var}(\varepsilon_{t-i}) = \sum_{i=0}^{\infty} b_i^2 \sigma^2 = \sigma^2 \sum_{i=0}^{\infty} b_i^2 .$$

At this point, in parallel to our discussion of white noise, we could compute and examine the autocovariance and autocorrelation functions of the general linear process. Those calculations, however, are rather involved, and not particularly revealing, so we'll proceed instead to examine the conditional mean and variance, where the information set Ω_{t-1} on which we condition contains past innovations; that is, $\Omega_{t-1} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$. In this manner, we can see how dynamics are modeled via conditional moments.[13] The

[13] Although Wold's theorem guarantees only serially uncorrelated white noise innovations, we shall sometimes make a stronger assumption of independent white noise innovations to focus the discussion. We do so, for example, in the following characterization of the conditional moment structure of the general linear process.

conditional mean is

$$E(y_t \mid \Omega_{t-1}) = E(\varepsilon_t \mid \Omega_{t-1}) + b_1 E(\varepsilon_{t-1} \mid \Omega_{t-1}) + b_2 E(\varepsilon_{t-2} \mid \Omega_{t-1}) + \cdots$$
$$= 0 + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \cdots = \sum_{i=1}^{\infty} b_i\varepsilon_{t-i} \ ,$$

and the conditional variance is

$$\mathrm{var}(y_t \mid \Omega_{t-1}) = E((y_t - E(y_t \mid \Omega_{t-1}))^2 \mid \Omega_{t-1}) = E(\varepsilon_t^2 \mid \Omega_{t-1}) = E(\varepsilon_t^2) = \sigma^2 .$$

The key insight is that the conditional mean *moves* over time in response to the evolving information set. The model captures the dynamics of the process, and the evolving conditional mean is one crucial way of summarizing them. An important goal of time series modeling, especially for forecasters, is capturing such conditional mean dynamics—the unconditional mean is constant (a requirement of stationarity), but the conditional mean varies in response to the evolving information set.[14]

Rational Distributed Lags

As we've seen, the Wold representation points to the crucial importance of models with infinite distributed lags. Infinite distributed lag models, in turn, are stated in terms of infinite polynomials in the lag operator, which are therefore very important as well. Infinite distributed lag models are not of immediate practical use, however, because they contain infinitely many parameters, which certainly inhibits practical application! Fortunately, infinite polynomials in the lag operator needn't contain infinitely many free parameters. The infinite polynomial $B(L)$ may, for example, be a ratio of finite-order (and perhaps very low-order) polynomials. Such polynomials are called **rational polynomials,** and distributed lags constructed from them are called **rational distributed lags.**

Suppose, for example, that

$$B(L) = \frac{\Theta(L)}{\Phi(L)} \ ,$$

where the numerator polynomial is of degree q,

$$\Theta(L) = \sum_{i=0}^{q} \theta_i L^i \ ,$$

and the denominator polynomial is of degree p,

$$\Phi(L) = \sum_{i=0}^{p} \varphi_i L^i .$$

There are *not* infinitely many free parameters in the $B(L)$ polynomial; instead, there are only $p + q$ parameters (the θ's and the φ's). If p and q are small—say,

[14] Note, however, an embarrassing asymmetry: the conditional variance, like the unconditional variance, is a fixed constant. However, models that allow the conditional variance to change with the information set have been developed recently, as discussed in detail in Chapter 14.

0, 1, or 2—then what seems like a hopeless task—estimation of $B(L)$—may actually be easy.

More realistically, suppose that $B(L)$ is not exactly rational but is approximately rational,

$$B(L) \approx \frac{\Theta(L)}{\Phi(L)} .$$

Then we can find an **approximation of the Wold representation** using a rational distributed lag. Rational distributed lags produce models of cycles that economize on parameters (they're **parsimonious**), while nevertheless providing accurate approximations to the Wold representation. The popular ARMA and ARIMA forecasting models, which we'll study shortly, are simply rational approximations to the Wold representation.

5. Estimation and Inference for the Mean, Autocorrelation, and Partial Autocorrelation Functions

Now suppose we have a sample of data on a time series, and we don't know the true model that generated the data, or the mean, autocorrelation function, or partial autocorrelation function associated with that true model. Instead, we want to use the data to *estimate* the mean, autocorrelation function, and partial autocorrelation function, which we might then use to help us learn about the underlying dynamics and to decide on a suitable model or set of models to fit to the data.

Sample Mean

The mean of a covariance stationary series is $\mu = Ey_t$. A fundamental principle of estimation, called the **analog principle,** suggests that we develop estimators by replacing expectations with sample averages. Thus, our estimator for the population mean, given a sample of size T, is the **sample mean,**

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t .$$

Typically we're not directly interested in the estimate of the mean, but it's needed for estimation of the autocorrelation function.

Sample Autocorrelations

The autocorrelation at displacement τ for the covariance stationary series y is

$$\rho(\tau) = \frac{E((y_t - \mu)(y_{t-\tau} - \mu))}{E((y_t - \mu)^2)} .$$

Application of the analog principle yields a natural estimator,

$$\hat{\rho}(\tau) = \frac{\frac{1}{T}\sum_{t=\tau+1}^{T}((y_t - \bar{y})(y_{t-\tau} - \bar{y}))}{\frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})^2} = \frac{\sum_{t=\tau+1}^{T}((y_t - \bar{y})(y_{t-\tau} - \bar{y}))}{\sum_{t=1}^{T}(y_t - \bar{y})^2} .$$

This estimator, viewed as a function of τ, is called the **sample autocorrelation function** or correlogram. Note that some of the summations begin at $t = \tau + 1$, not at $t = 1$; this is necessary because of the appearance of $y_{t-\tau}$ in the sum. Note that we divide those same sums by T, even though only $(T - \tau)$ terms appear in the sum. When T is large relative to τ (which is the relevant case), division by T or by $T - \tau$ will yield approximately the same result, so it won't make much difference for practical purposes; moreover, there are good mathematical reasons for preferring division by T.[15]

It's often of interest to assess whether a series is reasonably approximated as white noise, which is to say whether all its autocorrelations are 0 in population. A key result, which we simply assert, is that if a series is white noise, then the distribution of the sample autocorrelations in large samples is

$$\hat{\rho}(\tau) \sim N\left(0, \frac{1}{T}\right) .$$

Note how simple the result is. The sample autocorrelations of a white noise series are approximately normally distributed, and the normal is always a convenient distribution to work with. Their mean is 0, which is to say the sample autocorrelations are unbiased estimators of the true autocorrelations, which are in fact 0. Finally, the variance of the sample autocorrelations is approximately $1/T$ (equivalently, the standard deviation is $1/\sqrt{T}$), which is easy to construct and remember. Under normality, taking plus or minus two standard errors yields an approximate 95% confidence interval. Thus, if the series is white noise, then approximately 95% of the sample autocorrelations should fall in the interval $\pm\frac{2}{\sqrt{T}}$. In practice, when we plot the sample autocorrelations for a sample of data, we typically include the "two-standard-error bands," which are useful for making informal graphical assessments of whether and how the series deviates from white noise.

The two-standard-error bands, although very useful, only provide 95% bounds for the sample autocorrelations taken one at a time. Ultimately, we're often interested in whether a series is white noise—that is, whether *all* its autocorrelations are *jointly* 0. A simple extension lets us test that hypothesis. Rewrite the expression

$$\hat{\rho}(\tau) \sim N\left(0, \frac{1}{T}\right)$$

as

$$\sqrt{T}\hat{\rho}(\tau) \sim N(0, 1) .$$

[15] For additional discussion, consult any of the more advanced time series texts mentioned in Chapter 1.

Squaring both sides yields[16]

$$T\hat{\rho}^2(\tau) \sim \chi_1^2 .$$

It can be shown that, in addition to being approximately normally distributed, the sample autocorrelations at various displacements are approximately independent of one another. Recalling that the sum of independent χ^2 variables is also χ^2 with degrees of freedom equal to the sum of the degrees of freedom of the variables summed, we have shown that the **Box-Pierce Q-statistic,**

$$Q_{\text{BP}} = T\sum_{\tau=1}^{m}\hat{\rho}^2(\tau) ,$$

is approximately distributed as a χ_m^2 random variable under the null hypothesis that y is white noise.[17] A slight modification of this, designed to follow more closely the χ^2 distribution in small samples, is

$$Q_{\text{LB}} = T(T+2)\sum_{\tau=1}^{m}\left(\frac{1}{T-\tau}\right)\hat{\rho}^2(\tau) .$$

Under the null hypothesis that y is white noise, Q_{LB} is approximately distributed as a χ_m^2 random variable. Note that the **Ljung-Box Q-statistic** is the same as the Box-Pierce Q-statistic, except that the sum of squared autocorrelations is replaced by a weighted sum of squared autocorrelations, where the weights are $(T+2)/(T-\tau)$. For moderate and large T, the weights are approximately 1, so that the Ljung-Box statistic differs little from the Box-Pierce statistic.

Selection of m is done to balance competing criteria. On the one hand, we don't want m too small, because, after all, we're trying to do a joint test on a large part of the autocorrelation function. On the other hand, as m grows relative to T, the quality of the distributional approximations we've invoked deteriorates. In practice, focusing on m in the neighborhood of $\sqrt{T}$ is often reasonable.

Sample Partial Autocorrelations

Recall that the partial autocorrelations are obtained from population linear regressions, which correspond to a thought experiment involving linear regression using an infinite sample of data. The sample partial autocorrelations correspond to the same thought experiment, except that the linear regression is now done on the (feasible) sample of size T. If the fitted regression is

$$\hat{y}_t = \hat{c} + \hat{\beta}_1 y_{t-1} + \cdots + \hat{\beta}_\tau y_{t-\tau} ,$$

then the **sample partial autocorrelation** at displacement τ is

$$\hat{p}(\tau) \equiv \hat{\beta}_\tau .$$

[16] Recall that the square of a standard normal random variable is a χ^2 random variable with 1 degree of freedom. We square the sample autocorrelations $\hat{\rho}(\tau)$ so that positive and negative values don't cancel when we sum across various values of τ, as we will soon do.

[17] m is a maximum displacement selected by the user. Shortly we'll discuss how to choose it.

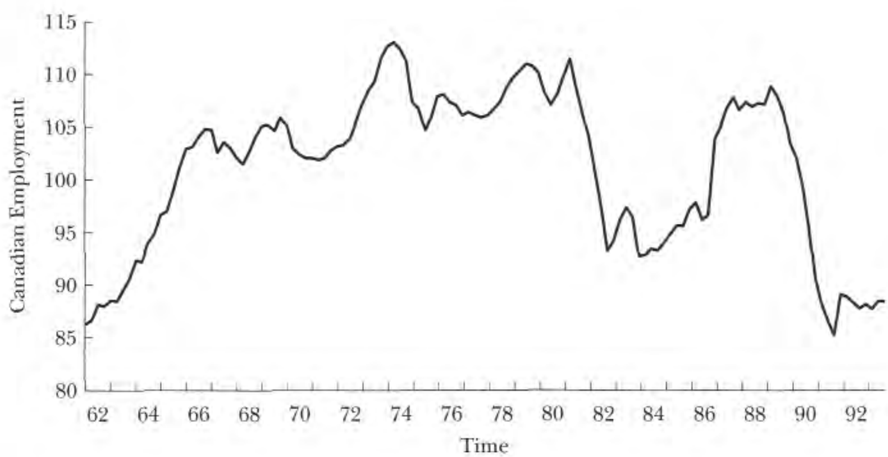
Distributional results identical to those we discussed for the sample autocorrelations hold as well for the sample *partial* autocorrelations. That is, if the series is white noise, approximately 95% of the sample partial autocorrelations should fall in the interval $\pm\frac{2}{\sqrt{T}}$. As with the sample autocorrelations, we typically plot the sample partial autocorrelations along with their two-standard-error bands.

6. Application: Characterizing Canadian Employment Dynamics

To illustrate the ideas we've introduced, we examine a quarterly, seasonally adjusted index of Canadian employment, 1962.1–1993.4, which we plot in Figure 7.9. The series displays no trend, and of course it displays no seasonality because it's seasonally adjusted. It does, however, appear highly serially correlated. It evolves in a slow, persistent fashion—high in business cycle booms and low in recessions.

To get a feel for the dynamics operating in the employment series, we perform a correlogram analysis.[18] The results appear in Table 7.1. Consider first the Q-statistic.[19] We compute the Q-statistic and its p-value under the null hypothesis of white noise for values of m (the number of terms in the sum that

FIGURE 7.9
Canadian Employment Index



[18] A **correlogram analysis** simply means examination of the sample autocorrelation and partial autocorrelation functions (with two-standard-error bands), together with related diagnostics, such as Q-statistics.

[19] We show the Ljung-Box version of the Q-statistic.

TABLE 7.1
Canadian Employment Index, Correlogram

Sample: 1962:1 1993:4
Included observations: 128

	Acorr.	P. Acorr.	Std. Error	Ljung-Box	p-value
1	0.949	0.949	.088	118.07	0.000
2	0.877	−0.244	.088	219.66	0.000
3	0.795	−0.101	.088	303.72	0.000
4	0.707	−0.070	.088	370.82	0.000
5	0.617	−0.063	.088	422.27	0.000
6	0.526	−0.048	.088	460.00	0.000
7	0.438	−0.033	.088	486.32	0.000
8	0.351	−0.049	.088	503.41	0.000
9	0.258	−0.149	.088	512.70	0.000
10	0.163	−0.070	.088	516.43	0.000
11	0.073	−0.011	.088	517.20	0.000
12	−0.005	0.016	.088	517.21	0.000

underlies the Q-statistic) ranging from 1 through 12. The p-value is consistently 0 to four decimal places, so the null hypothesis of white noise is decisively rejected.

Now we examine the sample autocorrelations and partial autocorrelations. The sample autocorrelations are very large relative to their standard errors and display slow one-sided decay.[20] The sample partial autocorrelations, in contrast, are large relative to their standard errors at first (particularly for the one-quarter displacement) but are statistically negligible beyond displacement 2.[21] In Figure 7.10 we plot the sample autocorrelations and partial autocorrelations along with their two-standard-error bands.

It's clear that employment has a strong cyclical component; all diagnostics reject the white noise hypothesis immediately. Moreover, the sample autocorrelation and partial autocorrelation functions have particular shapes—the autocorrelation function displays slow one-sided damping, while the partial autocorrelation function cuts off at displacement 2. You might guess that such patterns, which summarize the dynamics in the series, might be useful for suggesting candidate forecasting models. Such is indeed the case, as we'll see in the next chapter.

[20] We don't show the sample autocorrelation or partial autocorrelation at displacement 0, because as we mentioned earlier, they equal 1.0, by construction, and therefore convey no useful information. We'll adopt this convention throughout.

[21] Note that the sample autocorrelation and partial autocorrelation are identical at displacement 1. That's because at displacement 1, there are no earlier lags to control for when computing the sample partial autocorrelation, so it equals the sample autocorrelation. At higher displacements, of course, the two diverge.

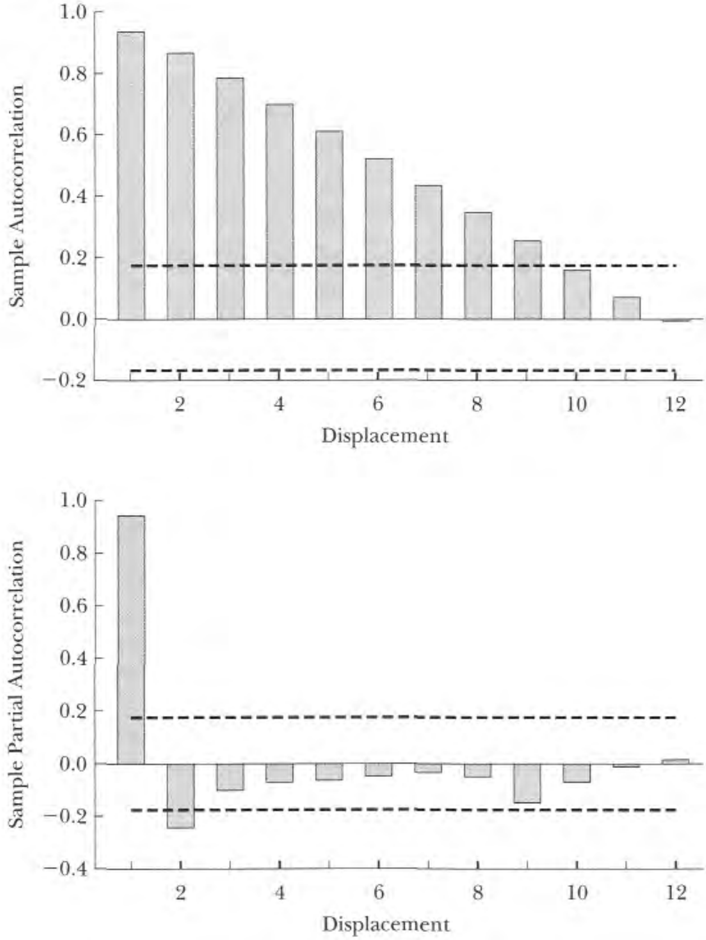


FIGURE 7.10 *Canadian Employment Index: Sample Autocorrelation and Partial Autocorrelation Functions, with Plus or Minus Two-Standard-Error Bands*

Exercises, Problems, and Complements

1. (Lag operator expressions 1) Rewrite the following expressions without using the lag operator.
   a. $(L^7)y_t = \varepsilon_t$

   b. $y_t = \left(\dfrac{2 + 5L + 0.8L^2}{L - 0.6L^3}\right)\varepsilon_t$

   c. $y_t = 2\left(1 + \dfrac{L^3}{L}\right)\varepsilon_t.$

2. (Lag operator expressions 2) Rewrite the following expressions in lag operator form.
   a. $y_t + y_{t-1} + \cdots + y_{t-N} = \alpha + \varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_{t-N}$, where α is a constant
   b. $y_t = \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t$.

3. (Autocorrelation functions of covariance stationary series) While interviewing at a top investment bank, your interviewer is impressed by the fact that you have taken a course on time series forecasting. She decides to test your knowledge of the autocovariance structure of covariance stationary series and lists four autocovariance functions:
   a. $\gamma(t, \tau) = \alpha$,
   b. $\gamma(t, \tau) = e^{-\alpha\tau}$,
   c. $\gamma(t, \tau) = \alpha\tau$, and
   d. $\gamma(t, \tau) = \dfrac{\alpha}{\tau}$,

   where α is a positive constant. Which autocovariance function(s) are consistent with covariance stationarity, and which are not? Why?

4. (Autocorrelation vs. partial autocorrelation) Describe the difference between autocorrelations and partial autocorrelations. How can autocorrelations at certain displacements be positive while the partial autocorrelations at those same displacements are negative?

5. (Conditional and unconditional means) As head of sales of the leading technology and innovation magazine publisher TECCIT, your bonus is dependent on the firm's revenue. Revenue changes from season to season, as subscriptions and advertising deals are entered or renewed. From your experience in the publishing business, you know that the revenue in a season is a function of the number of magazines sold in the previous season and can be described as $y_t = 1000 + 0.9x_{t-1} + \varepsilon_t$, with uncorrelated residuals $\varepsilon_t \sim N(0, 1000)$, where y is revenue and x is the number of magazines sold.
   a. What is the expected revenue for next season conditional on total sales of 6340 this season?
   b. What is unconditionally expected revenue if unconditionally expected sales are 8500?
   c. A rival publisher offers you a contract identical to your current contract (same base pay and bonus). Based on a confidential interview, you know that the same revenue model with identical coefficients is appropriate for your rival. The rival has sold an average of 9000 magazines in previous seasons but only 5650 this season. Will you accept the offer? Why or why not?

6. (White noise residuals) You work for a top five consulting firm and are in the middle of a 1-week vacation, when one of the directors calls you and urges you immediately to join a turnaround project at Stardust Cinemas. You are briefed that despite its bad financial condition, the recently fired CEO had planned to increase Stardust's market share by renovating every theater to include a bar, an arcade, and a restaurant. Your task on the team is to assess whether this renovation should be scrapped or included in a future value creation project. To do so, you spend a long night fitting a trend + seasonal model to a sample of $T = 100$ observations of Stardust's recent box office income data. You find that the residuals (e) from your model approximately follow $e_t = 0.5e_{t-1} + v_t$, where $v_t \overset{iid}{\sim} N(0, 1)$. At 4 A.M. you send your results to your project manager.

a. The next morning you receive an e-mail from your project manager. He thinks that your residuals do not look like white noise. Why? Why care?
b. Assuming that the residuals do indeed follow $e_t = 0.5e_{t-1} + v_t$, what is their autocorrelation function? Discuss.
c. What type of model might be useful for describing the historical path of box office income and its likely future path in the absence of renovations? How would you use it to assess the efficacy of the renovation project, if implemented?

7. (Selecting an employment forecasting model with the AIC and SIC) Use the AIC and SIC to assess the necessity and desirability of including trend and seasonal components in a forecasting model for Canadian employment.
   a. Display the AIC and SIC for a variety of specifications of trend and seasonality. Which would you select using the AIC? SIC? Do the AIC and SIC select the same model? If not, which do you prefer?
   b. Discuss the estimation results and residual plot from your preferred model, and perform a correlogram analysis of the residuals. Discuss, in particular, the patterns of the sample autocorrelations and partial autocorrelations, and their statistical significance.
   c. How, if at all, are your results different from those reported in the text? Are the differences important? Why or why not?

8. (**Simulation of a time series process**) Many cutting-edge estimation and forecasting techniques involve simulation. Moreover, simulation is often a good way to get a feel for a model and its behavior. White noise can be simulated on a computer using **random number generators,** which are available in most statistics, econometrics, and forecasting packages.
   a. Simulate a Gaussian white noise realization of length 200. Call the white noise ε_t. Compute the correlogram. Discuss.
   b. Form the distributed lag $y_t = \varepsilon_t + 0.9\varepsilon_{t-1}$, $t = 2, 3, \ldots, 200$. Compute the sample autocorrelations and partial autocorrelations. Discuss.
   c. Let $y_1 = 1$ and $y_t = 0.9y_{t-1} + \varepsilon_t$, $t = 2, 3, \ldots, 200$. Compute the sample autocorrelations and partial autocorrelations. Discuss.

9. (Sample autocorrelation functions for trending series) A telltale sign of the slowly evolving nonstationarity associated with trend is a sample autocorrelation function that damps extremely slowly.
   a. Find three trending series, compute their sample autocorrelation functions, and report your results. Discuss.
   b. Fit appropriate trend models, obtain the model residuals, compute their sample autocorrelation functions, and report your results. Discuss.

10. (Sample autocorrelation functions for seasonal series) A telltale sign of seasonality is a sample autocorrelation function with sharp peaks at the seasonal displacements (4, 8, 12, etc., for quarterly data; 12, 24, 36, etc., for monthly data; etc.).
    a. Find a series with both trend and seasonal variation. Compute its sample autocorrelation function. Discuss.
    b. Detrend the series. Discuss.
    c. Compute the sample autocorrelation function of the detrended series. Discuss.
    d. Seasonally adjust the detrended series. Discuss.
    e. Compute the sample autocorrelation function of the detrended, seasonally-adjusted series. Discuss.

11. (Volatility dynamics: correlograms of squares) In Chapter 4's Exercises, Problems, and Complements, we suggested that a time series plot of a squared residual, e_t^2, can reveal serial correlation in squared residuals, which corresponds to nonconstant volatility, or heteroskedasticity, in the levels of the residuals. Financial asset returns often display little systematic variation, so instead of examining residuals from a model of returns, we often examine returns directly. In what follows, we will continue to use the notation e_t, but you should interpret e_t as an observed asset return.
    a. Find a high-frequency (e.g., daily) financial asset return series, e_t, plot it, and discuss your results.
    b. Perform a correlogram analysis of e_t, and discuss your results.
    c. Plot e_t^2, and discuss your results.
    d. In addition to plotting e_t^2, examining the correlogram of e_t^2 often proves informative for assessing volatility persistence. Why might that be so? Perform a correlogram analysis of e_t^2, and discuss your results.

Bibliographical and Computational Notes

Wold's theorem was originally proved in a 1938 monograph, later revised as Wold (1954). Rational distributed lags have long been used in engineering, and their use in econometric modeling dates at least to Jorgenson (1966).

Bartlett (1946) derived the standard errors of the sample autocorrelations and partial autocorrelations of white noise. In fact, the plus-or-minus two-standard-error bands are often called the "**Bartlett bands.**"

The two variants of the Q-statistic that we introduced were developed in the 1970s by Box and Pierce (1970) and by Ljung and Box (1978). Some packages compute both variants, and some compute only one (typically Ljung-Box, because it's designed to be more accurate in small samples). In practice, the Box-Pierce and Ljung-Box statistics usually lead to the same conclusions.

For concise and insightful discussion of random number generation, as well as a variety of numerical and computational techniques, see Press *et al.* (1992).

Concepts for Review

Cycle
Realization
Sample path
Covariance stationarity
Autocovariance function
Second-order stationarity
Weak stationarity
Autocorrelation function
Partial autocorrelation function
Population regression
Autoregression
Time series process
Serially uncorrelated
Zero-mean white noise
White noise
Strong white noise
Weak white noise
Independent white noise
Normal white noise
Gaussian white noise
Unconditional mean and variance
Conditional mean and variance

Moments
Lag operator
Polynomial in the lag operator
Distributed lag
Wold's representation theorem
Wold representation
Innovation
General linear process
Rational polynomial
Rational distributed lag
Approximation of the Wold representation
Parsimonious
Analog principle
Sample mean
Sample autocorrelation function
Box-Pierce Q-statistic
Ljung-Box Q-statistic
Sample partial autocorrelation
Correlogram analysis
Simulation of a time series process
Random number generator
Bartlett bands

References and Additional Readings

Bartlett, M. (1946). "On the Theoretical Specification of Sampling Properties of Autocorrelated Time Series." *Journal of the Royal Statistical Society* B, 8, 27–41.

Box, G. E. P., and Pierce, D. A. (1970). "Distribution of Residual Autocorrelations in ARIMA Time-Series Models." *Journal of the American Statistical Association*, 65, 1509–1526.

Jorgenson, D. (1966). "Rational Distributed Lag Functions." *Econometrica*, 34, 135–149.

Ljung, G. M., and G. E. P. Box (1978). "On a Measure of Lack of Fit in Time-Series Models." *Biometrika*, 65, 297–303.

Press, W. H., et al. (1992). *Numerical Recipes: The Art of Scientific Computing.* Cambridge: Cambridge University Press.

Wold, H. O. (1954). *A Study in the Analysis of Stationary Time Series.* 2nd ed. Uppsala, Sweden: Almquist and Wicksell.

Wonnacott, T. H., and Wonnacott, R. J. (1990). *Introductory Statistics.* 5th ed. New York: Wiley.