

Applied Econometric Time Series

WALTER ENDERS
Iowa State University



JOHN WILEY & SONS, INC.

Acquisitions Editor Whitney Blake
Marketing Manager Karen Allman
Senior Production Editor Jeanine Furino
Manufacturing Manager Susan Stetzer
Illustration Coordinator Gene Aiello
Text Designer Alan Barnett
Cover Designer Carol C. Grobe

This book was set in Times Roman by General Graphic Services and
printed and bound by Courier Stoughton. The cover was printed by NEBC.

Recognizing the importance of preserving what has been written, it is a
policy of John Wiley & Sons, Inc. to have books of enduring value published
in the United States printed on acid-free paper, and we exert our best
efforts to that end.

Copyright © 1995, by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of this work beyond that permitted by
Sections 107 and 108 of the 1976 United States Copyright Act without the
permission of the copyright owner is unlawful. Requests for permission or fur-
ther information should be addressed to the Permissions Department, John
Wiley & Sons, Inc.

Library of Congress Cataloging-in-Publication Data

Enders, Walter, 1948-

Applied econometric time series / Walter Enders. — 1st ed.

p. cm. — (Wiley series in probability and mathematical statistics)

Includes bibliographical references.

ISBN 0-471-03941-1

I. Econometrics. 2. Time-series analysis. I. Title.

II. Series.

HB139.E55 1995

330'.01'5195—dc20

94-27849

CIP

To my mother, father, and Linda

Printed in the United States of America

10 9 8 7 6 5 4

Chapter 2

STATIONARY TIME-SERIES MODELS

The theory of linear difference equations can be extended to allow the forcing process $\{x_t\}$ to be stochastic. This class of linear stochastic difference equations underlies much of the theory of time-series econometrics. Especially important is the Box–Jenkins (1976) methodology for estimating time-series models of the form:

$$y_t = a_0 + a_1 y_{t-1} + \cdots + a_p y_{t-p} + \epsilon_t + \beta_1 \epsilon_{t-1} + \cdots + \beta_q \epsilon_{t-q}$$

Such models are called autoregressive integrated moving average (ARIMA) time-series models. The aims of this chapter are to:

1. Present the theory of stochastic linear difference equations and consider the time-series properties of stationary ARIMA models; a stationary ARIMA model is called an ARMA model. It is shown that the stability conditions of the previous chapter are necessary conditions for stationarity.
2. Develop the tools used in estimating ARMA models. Especially useful are the autocorrelation and partial autocorrelation functions. It is shown how the Box–Jenkins methodology relies on these tools to estimate an ARMA model from sample data.
3. Consider various test statistics to check for model adequacy. Several examples of estimated ARMA models are analyzed in detail. It is shown how a properly estimated model can be used for forecasting.

1. STOCHASTIC DIFFERENCE EQUATION MODELS

In this chapter, we continue to work with **discrete**, rather than continuous, time-series models. Recall from the discussion in Chapter 1 that we can evaluate the function $y = f(t)$ at t_0 and $t_0 + h$ to form

$$\Delta y = f(t_0 + h) - f(t_0)$$

As a practical matter, most economic time-series data are collected for discrete time periods. Thus, we consider only the equidistant intervals $t_0, t_0 + h, t_0 + 2h, t_0 + 3h, \dots$ and conveniently set $h = 1$. Be careful to recognize, however, that a discrete time series implies t , but not necessarily y_t , is discrete. For example, although Scotland's annual rainfall is a continuous variable, the sequence of such annual rainfall totals for years 1 through t is a discrete time series. In many economic applications, t refers to "time" so that h represents the change in time. However, t need not refer to the type of time interval as measured by a clock or calendar. Instead of allowing our measurement units to be minutes, days, quarters, or years, we can use t to refer to an ordered event number. We could let y_t denote the outcome of spin t on a roulette wheel; y_t can then take on any of the 38 values 00, 0, 1, ..., 36.

A discrete variable y is said to be a **random variable** (i.e., stochastic) if for any real number r , there exists a probability $p(y \leq r)$ that y takes on a value less than or equal to r . This definition is fairly general; in common usage, it is typically implied that there is at least one value of r for which $0 < p(y = r) < 1$. If there is some r for which $p(y = r) = 1$, y is deterministic rather than random.

It is useful to consider the elements of an observed time series $\{y_0, y_1, y_2, \dots, y_T\}$ as being realizations (i.e., outcomes) of a stochastic process. As in Chapter 1, we continue to let the notation y_t refer to an element of the entire sequence $\{y_t\}$. In our roulette example, y_t denotes the outcome of spin t on a roulette wheel. If we observe spins 1 through T , we can form the sequence y_1, y_2, \dots, y_T , or more compactly, $\{y_t\}$. In the same way, the term y_t could be used to denote GNP in time period t . Since we cannot forecast GNP perfectly, y_t is a random variable. Once we learn the value of GNP in period t , y_t becomes one of the realized values from a stochastic process. (Of course, measurement error may prevent us from ever knowing the "true" value of GNP.)

For discrete variables, the probability distribution of y_t is given by a formula (or table) that specifies each possible realized value of y_t and the probability associated with that realization. If the realizations are linked across time, there exists the joint probability distribution $p(y_1 = r_1, y_2 = r_2, \dots, y_T = r_T)$, where r_i is the realized value of y in period i . Having observed the first t realizations, we can form the expected value of y_{t+1}, y_{t+2}, \dots , conditioned on the observed values of y_1 through y_t .¹ This conditional mean, or expected value, of y_{t+i} is denoted by $E_t(y_{t+i} | y_1, y_{t-1}, \dots, y_t)$ or $E_t y_{t+i}$.

Of course, if y_t refers to the outcome of spinning a fair roulette wheel, the probability distribution is easily characterized. In contrast, we may never be able to completely describe the probability distribution for GNP. Nevertheless, the task of economic theorists is to develop models that capture the essence of the true data-generating process. Stochastic difference equations are one convenient way of modeling dynamic economic process. To take a simple example, suppose that the Federal Reserve's money supply target grows 3% each year. Hence,

$$m_t^* = 1.03m_{t-1}^* \quad (2.1)$$

or given the initial condition m_0^* , the particular solution is

$$m_t^* = (1.03)^t m_0^*$$

where m_t^* = the logarithm of the money supply target in year t
 m_0^* = the initial condition for the target money supply in period zero

Of course, the actual money supply m_t and target need not be equal. Suppose that at the end of period $t - 1$, there exist m_{t-1} outstanding dollars that are carried forward into period t . Hence, at the beginning of t there are m_{t-1} dollars so that the gap between actual and desired money holdings is $m_t^* - m_{t-1}$. Suppose that the Fed cannot perfectly control the money supply but attempts to change the money supply by ρ percent ($\rho < 100\%$) of any gap between the desired and actual money supply. We can model this behavior as

$$\Delta m_t = \rho(m_t^* - m_{t-1}) + \epsilon_t$$

or using (2.1), we obtain

$$m_t = \rho(1.03)^t m_0^* + (1 - \rho)m_{t-1} + \epsilon_t \quad (2.2)$$

where ϵ_t = the uncontrollable portion of the money supply

We assume the mean of ϵ_t is zero in all time periods.

Although the economic theory is overly simple, the model does illustrate the key points discussed above. Note the following:

1. Although the money supply is a continuous variable, (2.2) is a discrete difference equation. Since the forcing process $\{\epsilon_t\}$ is stochastic, the money supply is stochastic; we can call (2.2) a linear stochastic difference equation.
2. If we knew the distribution of $\{\epsilon_t\}$, we could calculate the distribution for each element in the $\{m_t\}$ sequence. Since (2.2) shows how the realizations of the $\{m_t\}$ sequence are linked across time, we would be able to calculate the various joint probabilities. Notice that the distribution of the money supply sequence is completely determined by the parameters of the difference equation (2.2) and distribution of the $\{\epsilon_t\}$ sequence.
3. Having observed the first t observations in the $\{m_t\}$ sequence, we can make forecasts of m_{t+1}, m_{t+2}, \dots . For example, if we update (2.2) by one period and take the conditional expectation, the forecast of m_{t+1} is $\rho(1.03)^{t+1}m_0^* + (1 - \rho)m_t$. Hence, $E_t m_{t+1} = \rho(1.03)^{t+1}m_0^* + (1 - \rho)m_t$.

Before we proceed too far along these lines, let us go back to the basic building block of discrete stochastic time-series models: the **white-noise process**. A sequence $\{\epsilon_t\}$ is a white-noise process if each value in the sequence has a mean of

zero, a constant variance, and is serially uncorrelated. Formally, if the notation $E(x)$ denotes the theoretical mean value of x , the sequence $\{\epsilon_t\}$ is a white-noise process if for each time period t ,

$$\begin{aligned} E(\epsilon_t) &= E(\epsilon_{t-1}) = \dots = 0 \\ E(\epsilon_t^2) &= E(\epsilon_{t-1}^2) = \dots = \sigma^2 \quad [\text{or } \text{var}(\epsilon_t) = \text{var}(\epsilon_{t-1}) = \dots = \sigma^2] \end{aligned}$$

and for all j

$$E(\epsilon_t \epsilon_{t-s}) = E(\epsilon_{t-j} \epsilon_{t-j-s}) = 0 \text{ for all } s \quad [\text{or } \text{cov}(\epsilon_t, \epsilon_{t-s}) = \text{cov}(\epsilon_{t-j}, \epsilon_{t-j-s}) = 0]$$

In the remainder of this text, $\{\epsilon_t\}$ will always refer to a white-noise process and σ^2 to the variance of that process. When it is necessary to refer to two or more white-noise processes, symbols such as $\{\epsilon_{1,t}\}$ and $\{\epsilon_{2,t}\}$ will be used. Now, use a white-noise process to construct the more interesting time series:

$$x_t = \sum_{i=0}^q \beta_i \epsilon_{t-i} \quad (2.3)$$

For each period t , x_t is constructed by taking the values $\epsilon_t, \epsilon_{t-1}, \dots, \epsilon_{t-q}$ and multiplying each by the associated value of β_i . A sequence formed in this manner is called a **moving average** of order q and denoted by $\text{MA}(q)$. To illustrate a typical moving average process, suppose you win \$1 if a fair coin shows a head and lose \$1 if it shows a tail. Denote the outcome on toss t by ϵ_t (i.e., for toss t , ϵ_t is either $+\$1$ or $-\$1$). If you wish to keep track of your “hot streaks,” you might want to calculate your average winnings on the last four tosses. For each coin toss t , your average winnings on the last four tosses are $1/4\epsilon_t + 1/4\epsilon_{t-1} + 1/4\epsilon_{t-2} + 1/4\epsilon_{t-3}$. In terms of (2.3), this sequence is a moving average process such that $\beta_i = 0.25$ for $i \leq 3$ and zero otherwise.

Although the $\{\epsilon_t\}$ sequence is a white-noise process, the constructed $\{x_t\}$ sequence will *not* be a white-noise process if two or more of the β_i differ from zero. To illustrate using an $\text{MA}(1)$ process, set $\beta_0 = 1$, $\beta_1 = 0.5$, and all other $\beta_i = 0$. In this circumstance, $E(x_t) = E(\epsilon_t + 0.5\epsilon_{t-1}) = 0$ and $\text{var}(x_t) = \text{var}(\epsilon_t + 0.5\epsilon_{t-1}) = 1.25\sigma^2$. You can easily convince yourself that $E(x_t) = E(x_{t-s})$ and $\text{var}(x_t) = \text{var}(x_{t-s})$ for all s . Hence, the first two conditions for $\{x_t\}$ to be a white-noise process are satisfied. However, $E(x_t x_{t-1}) = E[(\epsilon_t + 0.5\epsilon_{t-1})(\epsilon_{t-1} + 0.5\epsilon_{t-2})] = E[\epsilon_t \epsilon_{t-1} + 0.5(\epsilon_{t-1})^2 + 0.5\epsilon_t \epsilon_{t-2} + 0.25\epsilon_{t-1} \epsilon_{t-2}] = 0.5\sigma^2$. Given there exists a nonzero value of s such that $E(x_t x_{t-s}) \neq 0$, the $\{x_t\}$ sequence is not a white-noise process.

Exercise 1 at the end of this chapter asks you to find the mean, variance, and covariance of your “hot streaks” in coin tossing. For practice, you should complete that exercise before continuing.

2. ARMA MODELS

It is possible to combine a moving average process with a linear difference equation to obtain an autoregressive moving average model. Consider the p th-order difference equation:

$$y_t = a_0 + \sum_{i=1}^p a_i y_{t-i} + x_t \quad (2.4)$$

Now let $\{x_t\}$ be the $\text{MA}(q)$ process given by (2.3) so that we can write

$$y_t = a_0 + \sum_{i=1}^p a_i y_{t-i} + \sum_{i=0}^q \beta_i \epsilon_{t-i} \quad (2.5)$$

We follow the convention of normalizing units so that β_0 is always equal to unity. If the characteristic roots of (2.5) are all in the unit circle, $\{y_t\}$ is called an **autoregressive moving average** (ARMA) model for y_t . The autoregressive part of the model is the “difference equation” given by the homogeneous portion of (2.4) and the moving average part is the $\{x_t\}$ sequence. If the homogeneous part of the difference equation contains p lags and the model for x_t q lags, the model is called an $\text{ARMA}(p, q)$ model. If $q = 0$, the process is called a pure autoregressive process denoted by $\text{AR}(p)$, and if $p = 0$, the process is a pure moving average process denoted by $\text{MA}(q)$. In an ARMA model, it is perfectly permissible to allow p and/or q to be infinite. In this chapter, we consider only models in which all the characteristic roots of (2.4) are within the unit circle. However, if one or more characteristic roots is greater than or equal to unity, the $\{y_t\}$ sequence is said to be an **integrated** process and (2.5) is called an autoregressive integrated moving average (ARIMA) model.

Treating (2.5) as a difference equation suggests that we can “solve” for y_t in terms of the $\{\epsilon_t\}$ sequence. The solution of an $\text{ARMA}(p, q)$ model expressing y_t in terms of the $\{\epsilon_t\}$ sequence is the **moving average representation** of y_t . The procedure is no different from that discussed in Chapter 1. For the $\text{AR}(1)$ model $y_t = a_0 + a_1 y_{t-1} + \epsilon_t$, the moving average representation was shown to be

$$y_t = a_0 / (1 - a_1) + \sum_{i=0}^{\infty} a_1^i \epsilon_{t-i}$$

For the general $\text{ARMA}(p, q)$ model, rewrite (2.5) using lag operators so that

$$\left(1 - \sum_{i=1}^p a_i L^i \right) y_t = a_0 + \sum_{i=0}^q \beta_i \epsilon_{t-i}$$

so that the *particular* solution for y_t is

$$y_t = \frac{a_0 + \sum_{i=0}^q \beta_i \epsilon_{t-i}}{1 - \sum_{i=1}^p a_i L^i} \quad (2.6)$$

Fortunately, it will not be necessary for us to expand (2.6) to obtain the specific coefficient for each element in $\{\epsilon_t\}$. The important point to recognize is that the expansion will yield an $MA(\infty)$ process. The issue is whether such an expansion is convergent so that the stochastic difference equation given by (2.6) is stable. As you will see in the next section, the stability condition is that the characteristic roots of the polynomial $(1 - \sum a_i L^i)$ must lie outside of the unit circle. It is also shown that if y_t is a linear stochastic difference equation, the stability condition is a necessary condition for the time series $\{y_t\}$ to be stationary.

3. STATIONARITY

Suppose that the quality control division of a manufacturing firm samples four machines each hour. Every hour, quality control finds the mean of the machines' output levels. The plot of each machine's hourly output is shown in Figure 2.1. If y_{it} represents machine y_i 's output at hour t , the means (\bar{y}_t) are readily calculated as

$$\bar{y}_t = \sum_{i=1}^4 y_{it} / 4$$

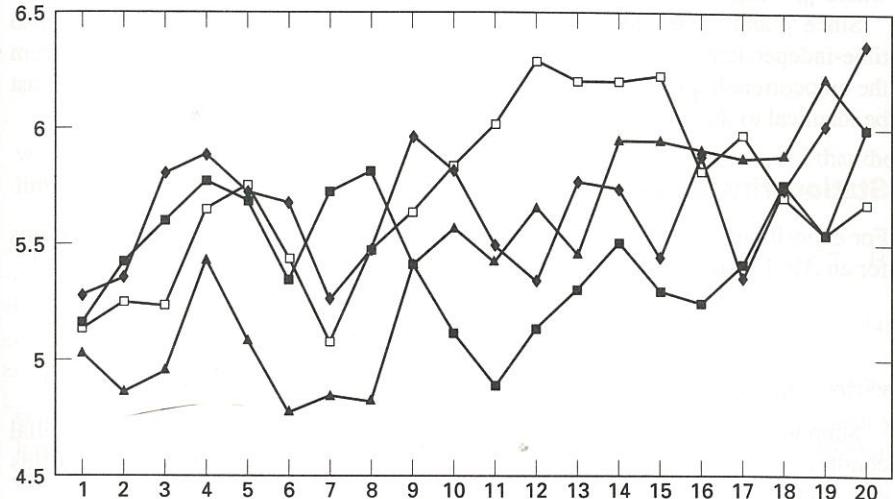
For hours 5, 10, and 15, these mean values are 5.57, 5.59, and 5.73, respectively.

The sample variance for each hour can similarly be constructed. Unfortunately, applied econometricians do not usually have the luxury of being able to obtain an **ensemble** (i.e., multiple time-series data of the same process over the same time period). Typically, we observe only one set of realizations for any particular series. Fortunately, if $\{y_t\}$ is a **stationary** series, the mean, variance, and autocorrelations can usually be well approximated by sufficiently long **time averages** based on the single set of realizations. Suppose you observed only the output of machine 1 for 20 periods. If you knew that the output was stationary, you could approximate the mean level of output by

$$\bar{y}_t = \sum_{t=1}^{20} y_{1t} / 20$$

In using this approximation, you would be assuming that the mean was the same for each period. In this example, the means of the four series are 5.45, 5.66, 5.45,

Figure 2.1 Hourly output of four machines.



and 5.71. Formally, a stochastic process having a finite mean and variance is **covariance stationary** if for all t and $t-s$,

$$E(y_t) = E(y_{t-s}) = \mu \quad (2.7)$$

$$E[(y_t - \mu)^2] = E[(y_{t-s} - \mu)^2] = \sigma_y^2 \quad [\text{var}(y_t) = \text{var}(y_{t-s}) = \sigma_y^2] \quad (2.8)$$

$$E[(y_t - \mu)(y_{t-s} - \mu)] = E[(y_{t-j} - \mu)(y_{t-j-s} - \mu)] = \gamma_s \quad [\text{cov}(y_t, y_{t-s}) = \text{cov}(y_{t-j}, y_{t-j-s})] \quad (2.9)$$

where μ , σ_y^2 and all γ_s are constants

In (2.9), allowing $s = 0$ means that γ_0 is equivalent to the variance of y_t . Simply put, a time series is covariance stationary if its mean and all autocovariances are unaffected by a change of time origin. In the literature, a covariance stationary process is also referred to as a weakly stationary, second-order stationary, or wide-sense stationary process. A **strongly** stationary process need not have a finite mean and/or variance (i.e., μ and/or γ_0 need not be finite); this terminology implies that weak stationarity can be a more stringent condition than strong stationarity. The text considers only covariance stationary series so that there is no ambiguity in using the terms stationary and covariance stationary interchangeably. One further word about terminology. In multivariate models, the term autocovariance is reserved for the covariance between y_t and its own lags. Cross-covariance refers to the covariance between one series and another. In univariate time-series models, there is no ambiguity and the terms autocovariance and covariance are used interchangeably.

For a covariance stationary series, we can define the **autocorrelation** between y_t and y_{t-s} as

$$\rho_s \equiv \gamma_s / \gamma_0$$

where γ_0 and γ_s are defined by (2.9).

Since γ_s and γ_0 are time-independent, the autocorrelation coefficients ρ_s are also time-independent. Although the autocorrelation between y_t and y_{t-1} can differ from the autocorrelation between y_t and y_{t-2} , the autocorrelation between y_t and y_{t-1} must be identical to that between y_{t-s} and y_{t-s-1} . Obviously, $\rho_0 = 1$.

Stationarity Restrictions for an AR(1) Process

For expositional convenience, first consider the necessary and sufficient conditions for an AR(1) process to be stationary. Let

$$y_t = a_0 + a_1 y_{t-1} + \epsilon_t$$

where ϵ_t = white noise

Suppose that the process started in period zero, so that y_0 is a deterministic initial condition. In Section 3 of the last chapter, it was shown that the solution to this equation is (also see Question 2 at the end of this chapter)

$$y_t = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i \epsilon_{t-i} \quad (2.10)$$

Taking the expected value of (2.10), we obtain

$$E y_t = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 \quad (2.11)$$

Updating by s periods yields

$$E y_{t+s} = a_0 \sum_{i=0}^{t+s-1} a_1^i + a_1^{t+s} y_0 \quad (2.12)$$

If we compare (2.11) and (2.12), it is clear that both means are time-dependent. Since $E y_t$ is not equal to $E y_{t+s}$, the sequence cannot be stationary. However, if t is large, we can consider the limiting value of y_t in (2.10). If $|a_1| < 1$, the expression $(a_1)^t y_0$ converges to zero as t becomes infinitely large and the sum $a_0[1 + a_1 + (a_1)^2 + (a_1)^3 + \dots]$ converges to $a_0/(1 - a_1)$. Thus, as $t \rightarrow \infty$ and if $|a_1| < 1$

$$\lim y_t = a_0/(1 - a_1) + \sum_{i=0}^{\infty} a_1^i \epsilon_{t-i} \quad (2.13)$$

Now take expectations of (2.13) so that for sufficiently large values of t , $E y_t = a_0/(1 - a_1)$. Thus, the mean value of y_t is finite and time-independent so that $E y_t =$

$E y_{t-s} = \mu$ for all t . Turning to the limiting value of the variance, we find

$$\begin{aligned} E(y_t - \mu)^2 &= E[(\epsilon_t + a_1 \epsilon_{t-1} + (a_1)^2 \epsilon_{t-2} + \dots)^2] \\ &= \sigma^2 [1 + (a_1)^2 + (a_1)^4 + \dots] = \sigma^2 / [1 - (a_1)^2] \end{aligned}$$

which is also finite and time-independent. Finally, it is easily demonstrated that the limiting values of all autocovariances are finite and time-independent:

$$\begin{aligned} E[(y_t - \mu)(y_{t-s} - \mu)] &= E\{[\epsilon_t + a_1 \epsilon_{t-1} + (a_1)^2 \epsilon_{t-2} + \dots][\epsilon_{t-s} + a_1 \epsilon_{t-s-1} + (a_1)^2 \epsilon_{t-s-2} + \dots]\} \\ &= \sigma^2 (a_1)^s [1 + (a_1)^2 + (a_1)^4 + \dots] \\ &= \sigma^2 (a_1)^s / [1 - (a_1)^2] \end{aligned} \quad (2.14)$$

In summary, if we can use the limiting value of (2.10), the $\{y_t\}$ sequence will be stationary. For any given y_0 and $|a_1| < 1$, it follows that t must be sufficiently large. Thus, if a sample is generated by a process that has recently begun, the realizations may not be stationary. It is for this very reason that many econometricians assume that the data-generating process has been occurring for an infinitely long time. In practice, the researcher must be wary of any data generated from a "new" process. For example, $\{y_t\}$ could represent the daily change in the dollar/mark exchange rate beginning immediately after the demise of the Bretton Woods fixed exchange rate system. Such a series may not be stationary due to the fact there were deterministic initial conditions (exchange rate changes were essentially zero in the Bretton Woods era). The careful researcher wishing to use stationary series might consider excluding some of these earlier observations from the period of analysis.

Little would change had we not been given the initial condition. Without the initial value y_0 , the sum of the homogeneous and particular solutions for y_t is

$$y_t = a_0 / (1 - a_1) + \sum_{i=0}^{\infty} a_1^i \epsilon_{t-i} + A(a_1)^t \quad (2.15)$$

where A = an arbitrary constant

If we take the expectation of (2.15), it is clear that the $\{y_t\}$ sequence cannot be stationary unless the expression $A(a_1)^t$ is equal to zero. Either the sequence must have started infinitely long ago (so that $A = 0$) or the arbitrary constant A must be zero. Recall that the arbitrary constant has the interpretation of a deviation from long-run equilibrium. A succinct way to state the stability conditions is the following:

1. The homogeneous solution must be zero. Either the sequence must have started infinitely far in the past or the process must always be in equilibrium (so that the arbitrary constant is zero).
2. The characteristic root a_1 must be less than unity in absolute value.

These two conditions readily generalize to all ARMA(p, q) processes. We know that the homogeneous solution to (2.5) has the form

$$\sum_{i=1}^p A_i \alpha_i^t$$

or if the roots are repeated,

$$\alpha \sum_{i=1}^m A_i t^i + \sum_{i=m+1}^p A_i \alpha_i^t$$

where the A_i represent p arbitrary values, α are the repeated roots, and the α_i are the $(p - m)$ distinct roots.

If any portion of the homogeneous equation is present, the mean, variance, and all covariances will be time-dependent. Hence, for any ARMA(p, q) model, stationarity necessitates that the homogeneous solution be zero. The next section addresses the stationarity restrictions for the particular solution.

4. STATIONARITY RESTRICTIONS FOR AN ARMA(p, q) MODEL

As a prelude to the stationarity conditions for the general ARMA(p, q) model, first consider the restrictions necessary to ensure that an ARMA(2, 1) model is stationary. Since the magnitude of the intercept term does not affect the stability (or stationarity) conditions, set $a_0 = 0$ and write

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \epsilon_t + \beta_1 \epsilon_{t-1} \quad (2.16)$$

From the previous section, we know that the homogeneous solution must be zero. As such, it is only necessary to find the particular solution. Using the method of undetermined coefficients, we can write the challenge solution as

$$y_t = \sum_{i=0}^{\infty} \alpha_i \epsilon_{t-i} \quad (2.17)$$

For (2.17) to be a solution of (2.16), the various α_i must satisfy

$$\begin{aligned} \alpha_0 \epsilon_t + \alpha_1 \epsilon_{t-1} + \alpha_2 \epsilon_{t-2} + \alpha_3 \epsilon_{t-3} + \dots &= a_1 (\alpha_0 \epsilon_{t-1} + \alpha_1 \epsilon_{t-2} + \alpha_2 \epsilon_{t-3} + \alpha_3 \epsilon_{t-4} + \dots) \\ &\quad + a_2 (\alpha_0 \epsilon_{t-2} + \alpha_1 \epsilon_{t-3} + \alpha_2 \epsilon_{t-4} + \alpha_3 \epsilon_{t-5} + \dots) + \epsilon_t + \beta_1 \epsilon_{t-1} \end{aligned}$$

To match coefficients on the terms containing $\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \dots$, it is necessary to set

1. $\alpha_0 = 1$
2. $\alpha_1 = a_1 \alpha_0 + \beta_1 \Rightarrow \alpha_1 = a_1 + \beta_1$
3. $\alpha_i = a_1 \alpha_{i-1} + a_2 \alpha_{i-2} \quad \text{for all } i \geq 2$

The key point is that for $i \geq 2$, the coefficients satisfy the difference equation $\alpha_i = a_1 \alpha_{i-1} + a_2 \alpha_{i-2}$. If the characteristic roots of (2.16) are within the unit circle, the $\{\alpha_i\}$ must constitute a convergent sequence. For example, reconsider the case in which $a_1 = 1.6$, $a_2 = -0.9$, and let $\beta_1 = 0.5$. Worksheet 2.1 shows that the coefficients satisfying (2.17) are 1, 2.1, 2.46, 2.046, 1.06, -0.146, . . . (also see Worksheet 1.2 of the previous chapter).

WORKSHEET 2.1 Coefficients of the ARMA(2,1) Process:

$$y_t = 1.6 y_{t-1} - 0.9 y_{t-2} + \epsilon_t + 0.5 \epsilon_{t-1}$$

If we use the method of undetermined coefficients, the α_i must satisfy

$$\begin{aligned} \alpha_0 &= 1 \\ \alpha_1 &= 1.6 + 0.5 \quad \text{hence, } \alpha_1 = 2.1 \\ \alpha_i &= 1.6 \cdot \alpha_{i-1} - 0.9 \cdot \alpha_{i-2} \quad \text{for all } i = 2, 3, 4 \dots \end{aligned}$$

Notice that the coefficients follow a second-order difference equation with imaginary roots. With de Moivre's theorem, the coefficients will satisfy

$$\alpha_i = 0.949^i \beta_1 \cos(0.567i + \beta_2)$$

Imposing the initial conditions for α_0 and α_1 yields

$$1 = \beta_1 \cos(\beta_2) \quad \text{and} \quad 2.1 = 0.949 \beta_1 \cos(0.567 + \beta_2)$$

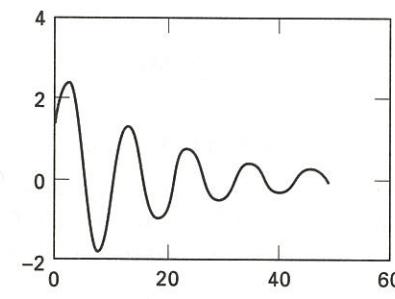
Since $\beta_1 = 1/\cos(\beta_2)$, we seek the solution to

$$\cos(\beta_2) - (0.949/2.1) \cdot \cos(0.567 + \beta_2) = 0$$

From a trig table, the solution for β_2 is -1.197 . Hence, the α_i satisfy

$$-1/1.197 \cdot 0.949^i \cdot \cos(0.567 \cdot i - 1.197)$$

Alternatively, we can use the initial values of α_0 and α_1 to find the other α_i by iteration. The sequence of the α_i is shown in the graph below.



The first 10 values of the sequence are

i	0	1	2	3	4	5	6	7	8	9	10
α_i	1.00	2.10	2.46	2.046	1.06	-0.146	1.187	-1.786	-1.761	-1.226	-0.378

To verify that the $\{y_t\}$ sequence generated by (2.17) is stationary, take the expectation of (2.17) to form $Ey_t = Ey_{t-i} = 0$ for all t and i . Hence, the mean is finite and time-invariant. Since the $\{\epsilon_t\}$ sequence is assumed to be a white-noise process, the variance of y_t is constant and time-independent, that is,

$$\begin{aligned}\text{Var}(y_t) &= E[(\alpha_0\epsilon_t + \alpha_1\epsilon_{t-1} + \alpha_2\epsilon_{t-2} + \alpha_3\epsilon_{t-3} + \dots)^2] \\ &= \sigma^2 \sum_{i=0}^{\infty} \alpha_i^2\end{aligned}$$

Hence, $\text{var}(y_t) = \text{var}(y_{t-s})$ for all t and s . Finally, the covariance between y_t and y_{t-s} is

$$\begin{aligned}\text{Cov}(y_t, y_{t-1}) &= E[(\epsilon_t + \alpha_1\epsilon_{t-1} + \alpha_2\epsilon_{t-2} + \dots)(\epsilon_{t-1} + \alpha_1\epsilon_{t-2} + \alpha_2\epsilon_{t-3} + \alpha_3\epsilon_{t-4} + \dots)] \\ &= \sigma^2(\alpha_1 + \alpha_2\alpha_1 + \alpha_3\alpha_2 + \dots) \\ \text{Cov}(y_t, y_{t-2}) &= E[(\epsilon_t + \alpha_1\epsilon_{t-1} + \alpha_2\epsilon_{t-2} + \dots)(\epsilon_{t-2} + \alpha_1\epsilon_{t-3} + \alpha_2\epsilon_{t-4} + \alpha_3\epsilon_{t-5} + \dots)] \\ &= \sigma^2(\alpha_2 + \alpha_3\alpha_1 + \alpha_4\alpha_2 + \dots)\end{aligned}$$

so that

$$\text{Cov}(y_t, y_{t-s}) = \sigma^2(\alpha_s + \alpha_{s+1}\alpha_1 + \alpha_{s+2}\alpha_2 + \dots) \quad (2.18)$$

Hence, $\text{cov}(y_t, y_{t-s})$ is constant and independent of t . Instead, if the characteristic roots of (2.16) do not lie within the unit circle, the $\{\alpha_i\}$ sequence will *not* be convergent. As such, the $\{y_t\}$ sequence *cannot* be convergent.

It is not too difficult to generalize these results to the entire class of ARMA(p, q) models. Begin by considering the conditions ensuring the stationarity of a pure MA(∞) process. By appropriately restricting the β_i , all the finite-order MA(q) processes can be obtained as special cases. Consider

$$x_t = \sum_{i=0}^{\infty} \beta_i \epsilon_{t-i}$$

where $\{\epsilon_t\}$ = a white-noise process with variance σ^2

We have already determined that $\{x_t\}$ is not a white-noise process; now the issue is whether $\{x_t\}$ is covariance stationary? (If you need to refresh your memory concerning mathematical expectations, you should consult the appendix to this chapter

before proceeding.) Considering conditions (2.7), (2.8), and (2.9), we ask the following:

1. Is the mean finite and time-independent? Take the expected value of x_t and remember that the expectation of a sum is the sum of the individual expectations. Hence,

$$\begin{aligned}E(x_t) &= E(\epsilon_t + \beta_1\epsilon_{t-1} + \beta_2\epsilon_{t-2} + \dots) \\ &= E\epsilon_t + \beta_1 E\epsilon_{t-1} + \beta_2 E\epsilon_{t-2} + \dots = 0\end{aligned}$$

Repeat the procedure with x_{t-s} :

$$E(x_{t-s}) = E(\epsilon_{t-s} + \beta_1\epsilon_{t-s-1} + \beta_2\epsilon_{t-s-2} + \dots) = 0$$

Hence, all elements in the $\{x_t\}$ sequence have the same finite mean ($\mu = 0$).

2. Is the variance finite and time-independent? Form $\text{var}(x_t)$ as

$$\text{Var}(x_t) = E[(\epsilon_t + \beta_1\epsilon_{t-1} + \beta_2\epsilon_{t-2} + \dots)^2]$$

Square the term in parentheses and take expectations. Since $\{\epsilon_t\}$ is a white-noise process, all terms $E\epsilon_t\epsilon_{t-s} = 0$ for $s \neq 0$. Hence,

$$\begin{aligned}\text{Var}(x_t) &= E(\epsilon_t)^2 + (\beta_1)^2 E(\epsilon_{t-1})^2 + (\beta_2)^2 E(\epsilon_{t-2})^2 + \dots \\ &= \sigma^2[1 + (\beta_1)^2 + (\beta_2)^2 + \dots]\end{aligned}$$

As long as $\sum(\beta_i)^2$ is finite, it follows that $\text{var}(x_t)$ is finite. Thus, $\sum(\beta_i)^2$ being finite is a necessary condition for $\{x_t\}$ to be stationary. To determine whether $\text{var}(x_t) = \text{var}(x_{t-s})$, form

$$\text{Var}(x_{t-s}) = E[(\epsilon_{t-s} + \beta_1\epsilon_{t-s-1} + \beta_2\epsilon_{t-s-2} + \dots)^2] = \sigma^2[1 + (\beta_1)^2 + (\beta_2)^2 + \dots]$$

Thus, $\text{var}(x_t) = \text{var}(x_{t-s})$ for all t and $t-s$.

3. Are all autocovariances finite and time-independent? First form $E(x_t x_{t-s})$ as

$$E(x_t x_{t-s}) = E[(\epsilon_t + \beta_1\epsilon_{t-1} + \beta_2\epsilon_{t-2} + \dots)(\epsilon_{t-s} + \beta_1\epsilon_{t-s-1} + \beta_2\epsilon_{t-s-2} + \dots)]$$

Carrying out the multiplication and noting that $E(\epsilon_t\epsilon_{t-s}) = 0$ for $s \neq 0$, we get

$$E(x_t x_{t-s}) = \sigma^2(\beta_s + \beta_1\beta_{s+1} + \beta_2\beta_{s+2} + \dots)$$

Restricting the sum $\beta_s + \beta_1\beta_{s+1} + \beta_2\beta_{s+2} + \dots$ to be finite means that $E(x_t x_{t-s})$ is finite. Given this second restriction, it is clear that the covariance between x_t and x_{t-s} depends on only the number of periods separating the variables (i.e., the value of s), but *not* the time subscript t .

In summary, the necessary and sufficient conditions for any MA process to be stationary are for the sums of (1), $\Sigma(\beta_i)^2$, and of (2), $(\beta_s + \beta_1\beta_{s+1} + \beta_2\beta_{s+2} + \dots)$, to be finite. Since (2) must hold for all values of s and $\beta_0 = 1$, condition (1) is redundant. The direct implication is that a finite-order MA process will always be stationary. For an infinite-order process, (2) must hold for all $s \geq 0$.

Stationarity Restrictions for the Autoregressive Coefficients

Now consider the pure autoregressive model:

$$y_t = a_0 + \sum_{i=1}^p a_i y_{t-i} + \epsilon_t \quad (2.19)$$

If the characteristic roots of the homogeneous equation of (2.19) all lie inside the unit circle, it is possible to write the particular solution as

$$y_t = \frac{a_0}{1 - \sum_{i=1}^p a_i} + \sum_{i=0}^{\infty} \alpha_i \epsilon_{t-i} \quad (2.20)$$

where the α_i = undetermined coefficients

Although it is possible to find the undetermined coefficients $\{\alpha_i\}$, we know that (2.20) is a convergent sequence so long as the characteristic roots of (2.19) are inside the unit circle. To sketch the proof, the method of undetermined coefficients allows us to write the particular solution in the form of (2.20). We also know that the sequence $\{\alpha_i\}$ will eventually solve the difference equation:

$$\alpha_i - a_1\alpha_{i-1} - a_2\alpha_{i-2} - \dots - a_p\alpha_{i-p} = 0 \quad (2.21)$$

If the characteristic roots of (2.21) are all inside the unit circle, the $\{\alpha_i\}$ sequence will be convergent. Although (2.20) is an infinite-order moving average process, the convergence of the MA coefficients implies that $\Sigma\alpha_i^2$ is finite. Hence, we can use (2.20) to check the three conditions for stationarity. Since $\alpha_0 = 1$,

$$1. E[y_t] = E[y_{t-s}] = a_0/(1 - \Sigma a_i)$$

You should recall from Chapter 1 that a necessary condition of all characteristic roots to lie inside the unit circle is $1 - \Sigma a_i \neq 0$. Hence, the mean of the sequence is finite and time-invariant:

$$2. \text{Var}(y_t) = E[(\epsilon_t + \alpha_1\epsilon_{t-1} + \alpha_2\epsilon_{t-2} + \alpha_3\epsilon_{t-3} + \dots)^2] = \sigma^2 \Sigma \alpha_i^2$$

and

$$\text{Var}(y_{t-s}) = E[(\epsilon_{t-s} + \alpha_1\epsilon_{t-s-1} + \alpha_2\epsilon_{t-s-2} + \alpha_3\epsilon_{t-s-3} + \dots)^2] = \sigma^2 \Sigma \alpha_i^2$$

Given that $\Sigma \alpha_i^2$ is finite, the variance is finite and time-independent.

$$\begin{aligned} 3. \text{Cov}(y_t, y_{t-s}) &= E[(\epsilon_t + \alpha_1\epsilon_{t-1} + \alpha_2\epsilon_{t-2} + \dots)(\epsilon_{t-s} + \alpha_1\epsilon_{t-s-1} + \alpha_2\epsilon_{t-s-2} + \dots)] \\ &= \sigma^2 (\alpha_s + \alpha_1\alpha_{s+1} + \alpha_2\alpha_{s+2} + \dots) \end{aligned}$$

Thus, the covariance between y_t and y_{t-s} is constant and time-invariant for all t and $t - s$.

Nothing of substance is changed by combining the AR(p) and MA(q) models into the general ARMA(p, q) model:

$$\begin{aligned} y_t &= a_0 + \sum_{i=1}^p a_i y_{t-i} + x_t \\ x_t &= \sum_{i=0}^q \beta_i \epsilon_{t-i} \end{aligned} \quad (2.22)$$

If the roots of the inverse characteristic equation lie outside of the unit circle [i.e., if the roots of the homogeneous form of (2.22) lie inside the unit circle] and the $\{x_t\}$ sequence is stationary, the $\{y_t\}$ sequence will be stationary. Consider

$$y_t = \frac{a_0}{1 - \sum_{i=1}^p a_i L^i} + \frac{\epsilon_t}{1 - \sum_{i=1}^p a_i L^i} + \frac{\beta_1 \epsilon_{t-1}}{1 - \sum_{i=1}^p a_i L^i} + \frac{\beta_2 \epsilon_{t-2}}{1 - \sum_{i=1}^p a_i L^i} + \dots \quad (2.23)$$

With very little effort, you can convince yourself that the $\{y_t\}$ sequence satisfies the three conditions for stationarity. Each of the expressions on the right-hand side of (2.23) is stationary as long as the roots of $1 - \Sigma a_i L^i$ are outside the unit circle. Given that $\{x_t\}$ is stationary, only the roots of the autoregressive portion of (2.22) determine whether the $\{y_t\}$ sequence is stationary.

What about the possibility of using the forward-looking solution? For example, in Cagan's monetary model you saw that the forward-looking solution yields a convergent sequence. Time-series econometrics rules out this type of perfect foresight/forward-looking solution. It is the *expectation* of future events (not the realized value of future events) that affects the present. After all, if you had perfect foresight, econometric forecasting would be unnecessary.

5. THE AUTOCORRELATION FUNCTION

The autocovariances and autocorrelations of the type found in (2.18) serve as useful tools in the Box-Jenkins (1976) approach to identifying and estimating time-series models. We illustrate by considering four important examples: the AR(1), AR(2), MA(1) and ARMA(1, 1) models. For the AR(1) model, $y_t = a_0 + a_1 y_{t-1} + \epsilon_t$, (2.14) shows

$$\gamma_0 = \sigma^2/[1 - (a_1)^2]$$

$$\gamma_s = \sigma^2(a_1)^s/[1 - (a_1)^2]$$

Forming the autocorrelations by dividing each γ_s by γ_0 , we find that $\rho_0 = 1$, $\rho_1 = a_1$; $\rho_2 = (a_1)^2, \dots, \rho_s = (a_1)^s$. For an AR(1) process, a necessary condition for stationarity is for $|a_1| < 1$. Thus, the plot of ρ_s against s —called the autocorrelation function (ACF) or **correlogram**—should converge to zero geometrically if the series is stationary. If a_1 is positive, convergence will be direct, and if a_1 is negative, the autocorrelations will follow a damped oscillatory path around zero. The first two graphs on the left-hand side of Figure 2.2 show the theoretical autocorrelation functions for $a_1 = 0.7$ and $a_1 = -0.7$, respectively. Here, ρ_0 is not shown since its value is necessarily unity.

The Autocorrelation Function of an AR(2) Process

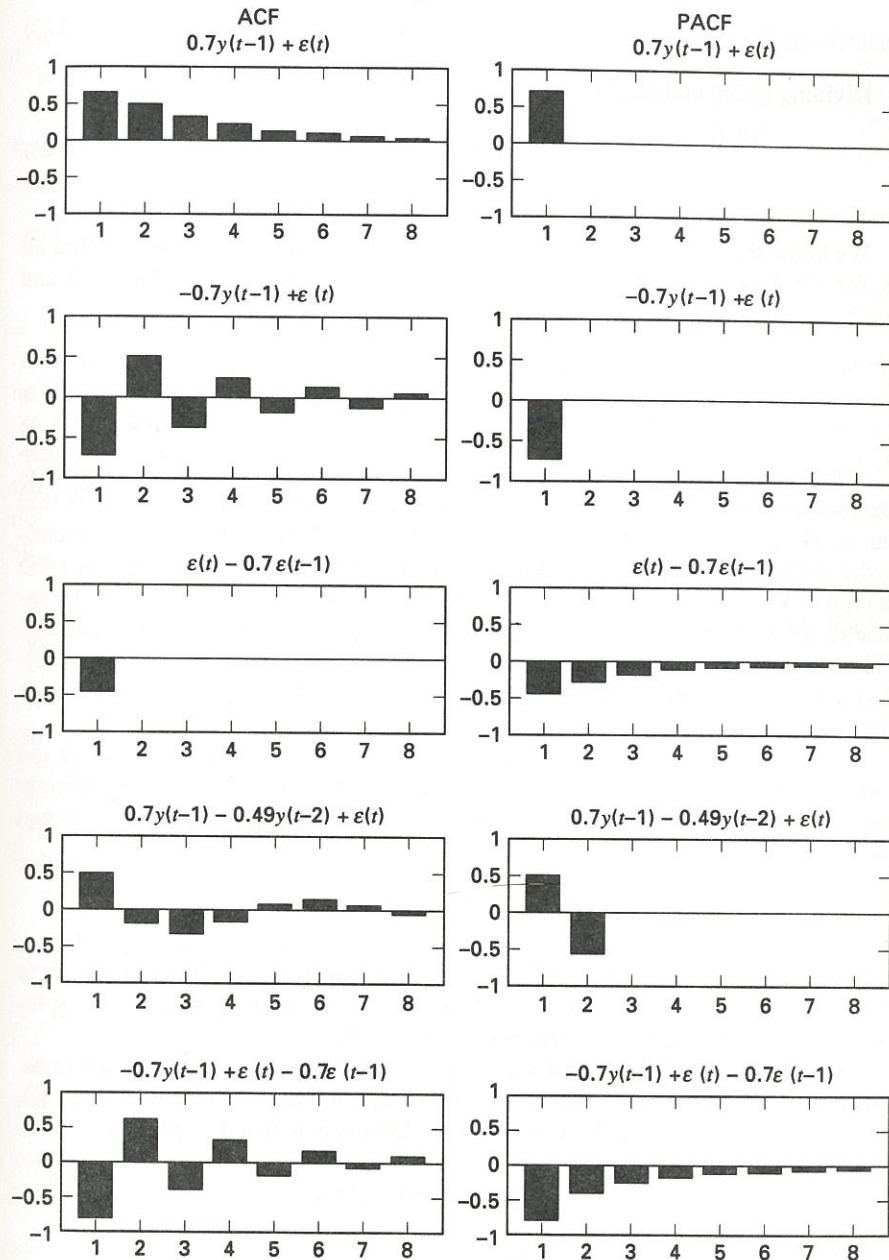
Now consider the more complicated AR(2) process $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \epsilon_t$. We omit an intercept term (a_0) since it has no effect on the ACF. For the second-order process to be stationary, we know that it is necessary to restrict the roots of $(1 - a_1 L - a_2 L^2)$ to be outside the unit circle. In Section 4, we derived the autocovariances of an ARMA(2, 1) process by use of the method of undetermined coefficients. Now we want to illustrate an alternative technique using the Yule-Walker equations. Multiply the second-order difference equation by y_{t-s} for $s = 0, s = 1, s = 2, \dots$ and take expectations to form

$$\begin{aligned} E y_t y_t &= a_1 E y_{t-1} y_t + a_2 E y_{t-2} y_t + E \epsilon_t y_t \\ E y_t y_{t-1} &= a_1 E y_{t-1} y_{t-1} + a_2 E y_{t-2} y_{t-1} + E \epsilon_t y_{t-1} \\ E y_t y_{t-2} &= a_1 E y_{t-1} y_{t-2} + a_2 E y_{t-2} y_{t-2} + E \epsilon_t y_{t-2} \\ &\vdots \\ E y_t y_{t-s} &= a_1 E y_{t-1} y_{t-s} + a_2 E y_{t-2} y_{t-s} + E \epsilon_t y_{t-s} \end{aligned} \quad (2.24)$$

By definition, the autocovariances of a stationary series are such that $E y_t y_{t-s} = E y_{t-s} y_t = E y_{t-k} y_{t-k-s} = \gamma_s$. We also know that the coefficient on ϵ_t is unity so that $E \epsilon_t y_t = \sigma^2$. Since $E \epsilon_t y_{t-s} = 0$, we can use the equations in (2.24) to form

$$\gamma_0 = a_1 \gamma_1 + a_2 \gamma_2 + \sigma^2 \quad (2.25)$$

Figure 2.2 Theoretical ACF and PACF patterns.



$$\gamma_1 = a_1 \gamma_0 + a_2 \gamma_1 \quad (2.26)$$

$$\gamma_s = a_1 \gamma_{s-1} + a_2 \gamma_{s-2} \quad (2.27)$$

Dividing (2.26) and (2.27) by γ_0 yields

$$\rho_1 = a_1 \rho_0 + a_2 \rho_1 \quad (2.28)$$

$$\rho_s = a_1 \rho_{s-1} + a_2 \rho_{s-2} \quad (2.29)$$

We know that $\rho_0 = 1$, so that from (2.28), $\rho_1 = a_1/(1 - a_2)$. Hence, we can find all ρ_s for $s \geq 2$ by solving the difference equation (2.29). For example, for $s = 2$ and $s = 3$,

$$\rho_2 = (a_1)^2/(1 - a_2) + a_2$$

$$\rho_3 = a_1[(a_1)^2/(1 - a_2) + a_2] + a_2 a_1/(1 - a_2)$$

Although the values of the ρ_s are cumbersome to derive, we can easily characterize their properties. Given the solutions for ρ_0 and ρ_1 , the key point to note is that the ρ_s all satisfy the difference equation (2.29). As in the general case of a second-order difference equation, the solution may be oscillatory or direct. Note that the stationarity condition for y_t necessitates that the characteristic roots of (2.29) lie inside of the unit circle. Hence, the $\{\rho_s\}$ sequence must be convergent. The correlogram for an AR(2) process must be such that $\rho_0 = 1$ and ρ_1 is determined by (2.28). These two values can be viewed as "initial values" for the second-order difference equation (2.29).

The fourth graph on the left-hand side of Figure 2.2 shows the ACF for the process $y_t - 0.7y_{t-1} - 0.49y_{t-2} + \epsilon_t$. The properties of the various ρ_s follow directly from the homogeneous equation $y_t - 0.7y_{t-1} - 0.49y_{t-2} = 0$. The roots are obtained from the solution to

$$\alpha = \{0.7 \pm [(-0.7)^2 - 4(0.49)]^{1/2}\}/2$$

Since the discriminant $d = (-0.7)^2 - 4(0.49)$ is negative, the characteristic roots are imaginary so that the solution oscillates. However, since $a_2 = -0.49$, the solution is convergent and the $\{y_t\}$ sequence is stationary.

Finally, we may wish to find the covariances rather than the autocorrelations. Since we know all the autocorrelations, if we can find the variance of y_t (i.e., γ_0), we can find all the other γ_s . To find γ_0 , use (2.25) and note that $\rho_i = \gamma_i/\gamma_0$, so

$$\text{Var}(y_t)(\rho_0 - a_1 \rho_1 - a_2 \rho_2) = \sigma^2$$

Substitution for ρ_0 , ρ_1 , and ρ_2 yields

$$\gamma_0 = \text{var}(y_t) = [(1 - a_2)/(1 + a_2)] \left[\frac{\sigma^2}{(a_1 + a_2 - 1)(a_2 - a_1 - 1)} \right]$$

The Autocorrelation Function of an MA(1) Process

Next consider the MA(1) process $y_t = \epsilon_t + \beta \epsilon_{t-1}$. Again, obtain the Yule–Walker equations by multiplying y_t by each y_{t-s} and take expectations:

$$\gamma_0 = \text{var}(y_t) = E[y_t y_t] = E[(\epsilon_t + \beta \epsilon_{t-1})(\epsilon_t + \beta \epsilon_{t-1})] = (1 + \beta^2)\sigma^2$$

$$\gamma_1 = E[y_t y_{t-1}] = E[(\epsilon_t + \beta \epsilon_{t-1})(\epsilon_{t-1} + \beta \epsilon_{t-2})] = \beta \sigma^2$$

and

$$\gamma_s = E[y_t y_{t-s}] = E[(\epsilon_t + \beta \epsilon_{t-1})(\epsilon_{t-s} + \beta \epsilon_{t-s-1})] = 0 \quad \text{for all } s > 1$$

Hence, by dividing each γ_s by γ_0 , it is immediately seen that the ACF is simply $\rho_0 = 1$, $\rho_1 = \beta/(1 + \beta^2)$, and $\rho_s = 0$ for all $s > 1$. The third graph on the left-hand side of Figure 2.2 shows the ACF for the MA(1) process $y_t = \epsilon_t - 0.7\epsilon_{t-1}$. As an exercise, you should demonstrate that the ACF for the MA(2) process $y_t = \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2}$ has two spikes and then cuts to zero.

The Autocorrelation Function of an ARMA(1, 1) Process

Finally, let $y_t = a_1 y_{t-1} + \epsilon_t + \beta_1 \epsilon_{t-1}$. Using the now familiar procedure, we find the Yule–Walker equations:

$$E[y_t y_t] = a_1 E[y_{t-1} y_t] + E[\epsilon_t y_t] + \beta_1 E[\epsilon_{t-1} y_t] \Rightarrow \gamma_0 = a_1 \gamma_1 + \sigma^2 + \beta_1(a_1 + \beta_1)\sigma^2 \quad (2.30)$$

$$E[y_t y_{t-1}] = a_1 E[y_{t-1} y_{t-1}] + E[\epsilon_t y_{t-1}] + \beta_1 E[\epsilon_{t-1} y_{t-1}] \Rightarrow \gamma_1 = a_1 \gamma_0 + \beta_1 \sigma^2 \quad (2.31)$$

$$E[y_t y_{t-2}] = a_1 E[y_{t-2} y_{t-2}] + E[\epsilon_t y_{t-2}] + \beta_1 E[\epsilon_{t-1} y_{t-2}] \Rightarrow \gamma_2 = a_1 \gamma_1 \quad (2.32)$$

$$\vdots$$

$$E[y_t y_{t-s}] = a_1 E[y_{t-s} y_{t-s}] + E[\epsilon_t y_{t-s}] + \beta_1 E[\epsilon_{t-1} y_{t-s}] \Rightarrow \gamma_s = a_1 \gamma_{s-1} \quad (2.33)$$

Solving (2.30) and (2.31) simultaneously for γ_0 and γ_1 yields

$$\gamma_0 = \frac{1 + \beta_1^2 + 2a_1\beta_1}{(1 - a_1^2)} \sigma^2$$

$$\gamma_1 = \frac{(1 + a_1\beta_1)(a_1 + \beta_1)}{(1 - a_1^2)} \sigma^2$$

Hence,

$$\rho_1 = \frac{(1 + a_1\beta_1)(a_1 + \beta_1)}{(1 + \beta_1^2 + 2a_1\beta_1)} \quad (2.34)$$

and $\rho_s = a_1 \rho_{s-1}$ for all $s \geq 2$.

Thus, the ACF for an ARMA(1, 1) process is such that the magnitude of ρ_1 depends on both a_1 and β_1 . Beginning with this value of ρ_1 , the ACF of an ARMA(1, 1) process looks like that of the AR(1) process. If $0 < a_1 < 1$, convergence will be direct, and if $-1 < a_1 < 0$, the autocorrelations will oscillate. The ACF for the function $y_t = -0.7y_{t-1} + \epsilon_t - 0.7\epsilon_{t-1}$ is shown in the last graph on the left-hand side of Figure 2.2. The top portion of Worksheet 2.2 derives these autocorrelations.

We leave you with the exercise of deriving the correlogram of the ARMA(2, 1) process used in Worksheet 2.1. You should be able to recognize the point that the correlogram can reveal the pattern of the autoregressive coefficients. For an ARMA(p, q) model beginning at lag q , the values of the ρ_i will satisfy

$$\rho_i = a_1\rho_{i-1} + a_2\rho_{i-2} + \dots + a_p\rho_{i-p}$$

The first $p - 1$ values can be treated as initial conditions that satisfy the Yule-Walker equations.

6. THE PARTIAL AUTOCORRELATION FUNCTION

In an AR(1) process, y_t and y_{t-2} are correlated even though y_{t-2} does not directly appear in the model. The correlation between y_t and y_{t-2} (i.e., ρ_2) is equal to the correlation between y_t and y_{t-1} (i.e., ρ_1) multiplied by the correlation between y_{t-1} and y_{t-2} (i.e., ρ_1 again) so that $\rho_2 = \rho_1^2$. It is important to note that all such "indirect" correlations are present in the ACF of any autoregressive process. In contrast, the **partial autocorrelation** between y_t and y_{t-s} eliminates the effects of the intervening values y_{t-1} through y_{t-s+1} . As such, in an AR(1) process, the partial autocorrelation between y_t and y_{t-2} is equal to zero. The most direct way to find the partial autocorrelation function is to first form the series $\{y_t^*\}$ by subtracting the mean of y (μ) from each observation: $y_t^* \equiv y_t - \mu$. Next, form the first-order autoregression equation:

$$y_t^* = \phi_{11}y_{t-1}^* + e_t$$

where: e_t = an error term

Here, the symbol $\{e_t\}$ is used since this error process may not be white-noise.

Since there are no intervening values, ϕ_{11} is both the autocorrelation and partial autocorrelation between y_t and y_{t-1} . Now form the second-order autoregression equation:

$$y_t^* = \phi_{21}y_{t-1}^* + \phi_{22}y_{t-2}^* + e_t$$

Here, ϕ_{22} is the partial autocorrelation coefficient between y_t and y_{t-2} . In other words, ϕ_{22} is the correlation between y_t and y_{t-2} controlling for (i.e., "netting out")

the effect of y_{t-1} . Repeating this process for all additional lags s yields the partial autocorrelation function (PACF). In practice, with sample size T , only $T/4$ lags are used in obtaining the sample PACF.

Since most statistical computer packages perform these transformations, there is little need to elaborate on the computational procedure. However, it should be pointed out that a simple computational method relying on the so-called Yule-Walker equations is available. One can form the partial autocorrelations from the autocorrelations as

$$\phi_{11} = \rho_1 \quad (2.35)$$

$$\phi_{22} = (\rho_2 - \rho_1^2)/(1 - \rho_1^2) \quad (2.36)$$

and for additional lags,

$$\phi_{ss} = \frac{\rho_s - \sum_{j=1}^{s-1} \phi_{s-1,j}\rho_{s-j}}{1 - \sum_{j=1}^{s-1} \phi_{s-1,j}\rho_j}, \quad s = 3, 4, 5, \dots \quad (2.37)$$

where $\phi_{sj} = \phi_{s-1,j} - \phi_{ss}\phi_{s-1,s-j}$, $j = 1, 2, 3, \dots, s - 1$.

For an AR(p) process, there is no direct correlation between y_t and y_{t-s} for $s > p$. Hence, all values of ϕ_{ss} for $s > p$ will be zero and the PACF for a pure AR(p) process should cut to zero for all lags greater than p . This is a useful feature of the PACF that can aid in the identification of an AR(p) model. In contrast, consider the PACF for the MA(1) process $y_t = \epsilon_t + \beta\epsilon_{t-1}$. As long as $\beta \neq -1$, we can write $y_t/(1 + \beta L) = \epsilon_t$, which we know has the infinite-order autoregressive representation:

$$y_t - \beta y_{t-1} + \beta^2 y_{t-2} - \beta^3 y_{t-3} + \dots = \epsilon_t$$

As such, the PACF will *not* jump to zero since y_t will be correlated with all its own lags. Instead, the PACF coefficients exhibit a geometrically decaying pattern. If $\beta < 0$, decay is direct, and if $\beta > 0$, the PACF coefficients oscillate.

Worksheet 2.2 illustrates the procedure used in constructing the PACF for the ARMA(1, 1) model shown in the fifth graph on the right-hand side of Figure 2.2:

$$y_t = -0.7y_{t-1} + \epsilon_t - 0.7\epsilon_{t-1}$$

First calculate the autocorrelations. Clearly, $\rho_0 = 1$; use Equation (2.34) to calculate as $\rho_1 = -0.8445$. Thereafter, the ACF coefficients decay at the rate $\rho_i = (-0.7)\rho_{i-1}$ for $i \geq 2$. Using (2.35) and (2.36), we obtain $\phi_{11} = -0.8445$ and $\phi_{22} = -0.4250$. All subsequent ϕ_{ss} and ϕ_{sj} can be calculated from (2.37) as in Worksheet 2.2.

WORKSHEET 2.2 Calculation of the partial autocorrelations of

$$y_t = -0.7y_{t-1} + \epsilon_t - 0.7\epsilon_{t-1}$$

STEP 1: Calculate the autocorrelations. Use (2.34) to calculate ρ_1 as

$$\rho_1 = \frac{(1+0.49)(-0.7-0.7)}{1+0.49+2(0.49)} = -0.8445$$

The remaining correlations decay at the rate $\rho_i = -0.7\rho_{i-1}$, so that

$$\begin{array}{llll} \rho_2 = 0.591 & \rho_3 = -0.414 & \rho_4 = 0.290 & \rho_5 = -0.203 \\ \rho_6 = 0.142 & \rho_7 = -0.010 & \rho_8 = 0.070 & \rho_9 = -0.049 \end{array}$$

STEP 2: Calculate the first two partial autocorrelations using (2.35) and (2.36).

Hence,

$$\begin{aligned} \phi_{11} &= \rho_1 = -0.844 \\ \phi_{22} &= [0.591 - (-0.8445)^2]/[1 - (-0.8445)^2] = -0.425 \end{aligned}$$

STEP 3: Construct all remaining ϕ_{ss} iteratively using (2.37). To find ϕ_{33} , note that

$$\phi_{21} = \phi_{11} - \phi_{22}\phi_{11} = -1.204 \text{ and form}$$

$$\begin{aligned} \phi_{33} &= \left(\rho_3 - \sum_{j=1}^2 \phi_{2j}\rho_{3-j} \right) \left(1 - \sum_{j=1}^2 \phi_{2j}\rho_j \right)^{-1} \\ &= [-0.414 - (-1.204)(0.591) - (-0.425)(-0.8445)]/ \\ &\quad [1 - (-1.204)(-0.8445) - (-0.425)(0.591)] \\ &= -0.262 \end{aligned}$$

Similarly, to find ϕ_{44} , use

$$\phi_{44} = \left(\rho_4 - \sum_{j=1}^3 \phi_{3j}\rho_{4-j} \right) \left(1 - \sum_{j=1}^3 \phi_{3j}\rho_j \right)^{-1}$$

Since $\phi_{3j} = \phi_{2j} - \phi_{33}\phi_{2,2-j}$, it follows that $\phi_{31} = -1.315$ and $\phi_{32} = -0.74$. Hence,

$$\phi_{44} = -0.173$$

If we continue in this fashion, it is possible to demonstrate that $\phi_{55} = -0.117$, $\phi_{66} = -0.081$, $\phi_{77} = -0.056$, and $\phi_{88} = -0.039$.

Table 2.1: Properties of the ACF and PACF

Process	ACF	PACF
White-noise	All $\rho_s = 0$.	All $\phi_{ss} = 0$.
AR(1): $a_1 > 0$	Direct exponential decay: $\rho_s = a_1^s$.	$\phi_{11} = \rho_1$; $\phi_{ss} = 0$ for $s \geq 2$.
AR(1): $a_1 < 0$	Oscillating decay: $\rho_s = a_1^s$.	$\phi_{11} = \rho_1$; $\phi_{ss} = 0$ for $s \geq 2$.
AR(p)	Decays toward zero. Coefficients may oscillate.	Spikes through lag p . All $\phi_{ss} = 0$ for $s > p$.
MA(1): $\beta > 0$	Positive spike at lag 1. $\rho_s = 0$ for $s \geq 2$.	Oscillating decay: $\phi_{11} > 0$.
MA(1): $\beta < 0$	Negative spike at lag 1. $\rho_s = 0$ for $s \geq 2$.	Decay: $\phi_{11} < 0$.
ARMA(1, 1): $a_1 > 0$	Exponential decay beginning at lag 1. Sign $\rho_1 = \text{sign}(a_1 + \beta)$.	Oscillating decay beginning at lag 1. $\phi_{11} = \rho_1$.
ARMA(1, 1): $a_1 < 0$	Oscillating decay beginning at lag 1. Sign $\rho_1 = \text{sign}(a_1 + \beta)$.	Exponential decay beginning at lag 1. $\phi_{11} = \rho_1$ and $\text{sign}(\phi_{ss}) = \text{sign}(\phi_{11})$.
ARMA (p, q)	Decay (either direct or oscillatory) beginning at lag q .	Decay (either direct or oscillatory) beginning at lag p .

More generally, the PACF of a stationary ARMA(p, q) process must ultimately decay toward zero beginning at lag p . The decay pattern depends on the coefficients of the polynomial $(1 + \beta_1L + \beta_2L^2 + \dots + \beta_qL^q)$. Table 2.1 summarizes some of the properties at the ACF and PACF for various ARMA processes. Also, the right-hand-side graphs of Figure 2.2 show the partial autocorrelation functions of the five indicated processes.

For stationary processes, the key points to note are the following:

1. The ACF of an ARMA(p, q) process will begin to decay at lag q . Beginning at lag q , the coefficients of the ACF (i.e., the ρ_i) will satisfy the difference equation $(\rho_i = a_1\rho_{i-1} + a_2\rho_{i-2} + \dots + a_p\rho_{i-p})$. Since the characteristic roots are inside the unit circle, the autocorrelations will decay beginning at lag q . Moreover, the pattern of the autocorrelation coefficients will mimic that suggested by the characteristic roots.
2. The PACF of an ARMA(p, q) process will begin to decay at lag p . Beginning at lag p , the coefficients of the PACF (i.e., the ϕ_{ss}) will mimic the ACF coefficients from the model $y_t/(1 + \beta_1L + \beta_2L^2 + \dots + \beta_qL^q)$.

We can illustrate the usefulness of the ACF and PACF functions using the model $y_t = a_0 + 0.7y_{t-1} + \epsilon_t$. If we compare the top two graphs of Figure 2.2, the ACF shows the monotonic decay of the autocorrelations, while the PACF exhibits the single spike at lag 1. Suppose that a researcher collected sample data and plotted the ACF and PACF functions. If the actual patterns compared favorably to the theoretical patterns, the researcher might try to estimate data using an AR(1) model.

Correspondingly, if the ACF exhibited a single spike and the PACF monotonic decay (see the third graph of the figure for the model $y_t = \epsilon_t - 0.7\epsilon_{t-1}$), the researcher might try an MA(1) model.

7. SAMPLE AUTOCORRELATIONS OF STATIONARY SERIES

In practice, the theoretical mean, variance, and autocorrelations of a series are unknown to the researcher. Given that a series is stationary, we can use the sample mean, variance, and autocorrelations to estimate the parameters of the actual data-generating process. Let there be T observations labeled y_1 through y_T . We can let \bar{y} , $\hat{\sigma}^2$, and r_s be estimates of μ , σ^2 , and ρ_s , respectively, where:

$$\bar{y} = \frac{\sum_{t=1}^T y_t}{T} \quad (2.38)$$

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^T (y_t - \bar{y})^2}{T} \quad (2.39)$$

and for each value of $s = 1, 2, \dots$,

$$r_s = \frac{\sum_{t=s+1}^T (y_t - \bar{y})(y_{t-s} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2} \quad (2.40)$$

The sample autocorrelation function [i.e., the ACF derived from (2.40)] and PACF can be compared to various theoretical functions to help identify the actual nature of the data-generating process. Box and Jenkins (1976) discuss the distribution of the sample values of r_s under the null that y_t is stationary with normally distributed errors. Allowing $\text{var}(r_s)$ to denote the sampling variance of r_s , they obtain

$$\begin{aligned} \text{Var}(r_s) &= T^{-1} && \text{for } s = 1 \\ &= \left(1 + 2 \sum_{j=1}^{s-1} r_j^2\right) T^{-1} && \text{for } s > 1 \end{aligned} \quad (2.41)$$

if the true value of $r_s = 0$ [i.e., if the true data-generating process is an MA($s - 1$) process]. Moreover, in large samples (i.e., for large values of T), r_s will be normally

distributed with a mean equal to zero. For the PACF coefficients, under the null hypothesis of an AR(p) model (i.e., under the null that all $\phi_{p+i,p+i}$ are zero), the variance of the $\phi_{p+i,p+i}$ is approximately T^{-1} .

In practice, we can use these sample values to form the sample autocorrelation and partial autocorrelation functions and test for significance using (2.41). For example, if we use a 95% confidence interval (i.e., two standard deviations), and the calculated value of r_1 exceeds $2T^{-1/2}$, it is possible to reject the null hypothesis that the first-order autocorrelation is not statistically different from zero. Rejecting this hypothesis means rejecting an MA($s - 1$) = MA(0) process and accepting the alternative $q > 0$. Next, try $s = 2$; $\text{var}(r_2)$ is $(1 + 2r_1^2)/T$. If r_1 is 0.5 and T 100, the variance of r_2 is 0.015 and the standard deviation about 0.123. Thus, if the calculated value of r^2 exceeds $2(0.123)$, it is possible to reject the hypothesis $r_2 = 0$. Here, rejecting the null means accepting the alternative that $q > 1$. Repeating for the various values of s is helpful in identifying the order to the process. In practice, the maximum number of sample autocorrelations and partial autocorrelations to use is $T/4$.

When looking over a large number of autocorrelations, we will see that some exceed two standard deviations as a result of pure chance even though the true values in the data-generating process are zero. The Q -statistic can be used to test whether a group of autocorrelations is significantly different from zero. Box and Pierce (1970) used the sample autocorrelations to form the statistic

$$Q = T \sum_{k=1}^s r_k^2$$

If the data are generated from a stationary ARMA process, Q is asymptotically χ^2 distributed with s degrees of freedom. The intuition behind the use of the statistic is that high sample autocorrelations lead to large values of Q . Certainly, a white-noise process (in which all autocorrelations should be zero) would have a Q value of zero. If the calculated value of Q exceeds the appropriate value in a χ^2 table, we can reject the null of no significant autocorrelations. Note that rejecting the null means accepting an alternative that at least one autocorrelation is not zero.

A problem with the Box-Pierce Q -statistic is that it works poorly even in moderately large samples. Ljung and Box (1978) report superior small sample performance for the modified Q -statistic calculated as

$$Q = T(T+2) \sum_{k=1}^s r_k^2 / (T-k) \quad (2.42)$$

If the sample value of Q calculated from (2.42) exceeds the critical value of χ^2 with s degrees of freedom, then *at least* one value of r_k is statistically different from zero at the specified significance level. The Box-Pierce and Ljung-Box Q -statistics also serve as a check to see if the *residuals* from an estimated ARMA(p, q) model

behave as a white-noise process. However, when we form the s correlations from an estimated ARMA(p, q) model, the degrees of freedom are reduced by the number of estimated coefficients. Hence, if using the residuals of an ARMA(p, q) model, Q has a χ^2 with $s-p-q$ degrees of freedom (if a constant is included, the degrees of freedom are $s-p-q-1$).

Model Selection Criteria

One natural question to ask of any estimated model is: How well does it fit the data? Adding additional lags for p and/or q will necessarily reduce the sum of squares of the estimated residuals. However, adding such lags entails the estimation of additional coefficients and an associated loss of degrees of freedom. Moreover, the inclusion of extraneous coefficients will reduce the forecasting performance of the fitted model. There exist various model selection criteria that trade off a reduction in the sum of squares of the residuals for a more **parsimonious** model. The two most commonly used model selection criteria are the Akaike information criterion (AIC) and Schwartz Bayesian criterion (SBC), calculated as

$$\begin{aligned} \text{AIC} &= T \ln(\text{residual sum of squares}) + 2n \\ \text{SBC} &= T \ln(\text{residual sum of squares}) + n \ln(T) \end{aligned}$$

where n = number of parameters estimated ($p + q +$ possible constant term);
 T = number of usable observations.

Typically in creating lagged variables, some observations are lost. To adequately compare the alternative models, T should be kept fixed. For example, with 100 data points, estimate an AR(1) and AR(2) using only the last 98 observations in each estimation. Compare the two models using $T = 98$.²

Ideally, the AIC and SBC will be as small as possible (note that both can be negative). We can use these criteria to aid in selecting the most appropriate model; model A is said to fit better than model B if the AIC (or SBC) for A is smaller than that for model B . In using the criteria to compare alternative models, we must estimate over the same sample period so that they will be comparable. For each, increasing the number of regressors increases n , but should have the effect of reducing the residual sum of squares. Thus, if a regressor has no explanatory power, adding it to the model will cause both the AIC and SBC to increase. Since $\ln(T)$ will be greater than 2, the SBC will always select a more parsimonious model than the AIC; the marginal cost of adding regressors is greater with the SBC than the AIC.

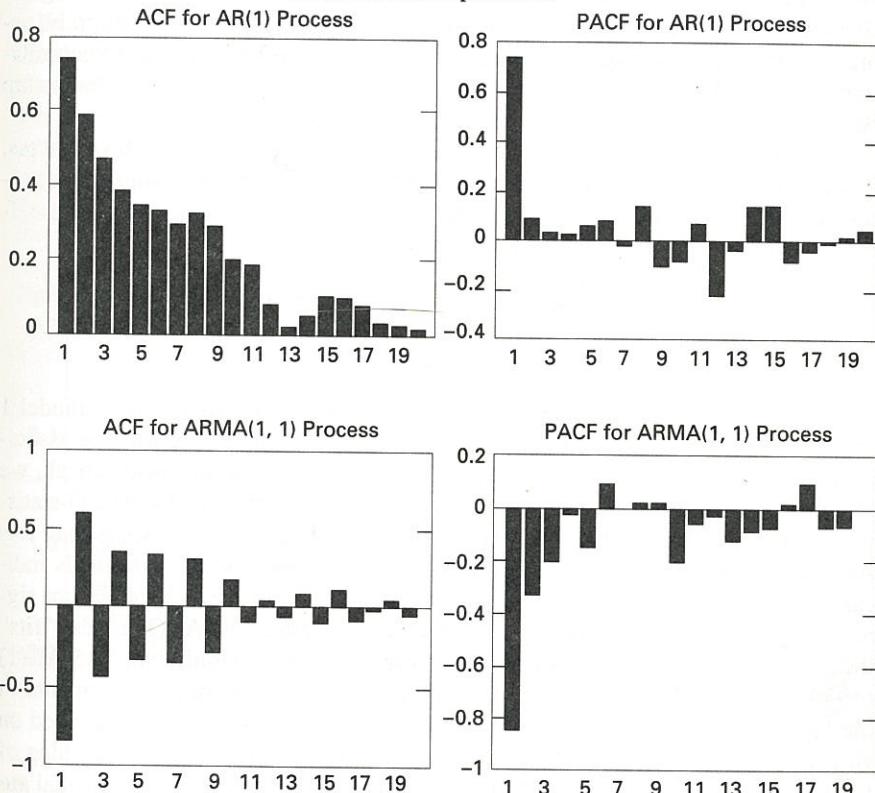
Of the two criteria, the SBC has superior large sample properties. Let the true order of the data-generating process be (p^*, q^*) and suppose that we use the AIC and SBC to estimate all ARMA models of order (p, q) where $p \geq p^*$ and $q \geq q^*$. Both the AIC and SBC will select models of orders greater than or equal to (p^*, q^*) as the sample size approaches infinity. However, the SBC is asymptotically consistent, whereas the AIC is biased toward selecting an overparameterized model.

Estimation of an AR(1) Model

Let us use a specific example to see how the sample autocorrelation function and partial autocorrelation function can be used as an aid in identifying an ARMA model. A computer program was used to draw 100 normally distributed random numbers with a theoretical variance equal to unity. Call these random variates ϵ_t , where t runs from 1 to 100. Beginning with $t = 1$, values of y_t were generated using the formula $y_t = 0.7y_{t-1} + \epsilon_t$, and initial condition $y_0 = 0$. Note that the problem of nonstationarity is avoided since the initial condition is consistent with long-run equilibrium. The upper-left-hand graph of Figure 2.3 shows the sample correlogram and upper-right-hand graph the sample PACF. You should take a minute to compare the ACF and PACF to those of the theoretical processes illustrated in Figure 2.2.

In practice, we never know the true data-generating process. However, suppose we were presented with these 100 sample values and asked to uncover the true process. The first step might be to compare the sample ACF and PACF to those of the various theoretical models. The decaying pattern of the ACF and the single

Figure 2.3 ACF and PACF for two simulated processes.



large spike in the sample PACF suggest an AR(1) model. The first three autocorrelations are $r_1 = 0.74$, $r_2 = 0.58$, and $r_3 = 0.47$, which are somewhat greater than the theoretical values of 0.7, 0.49 ($0.7^2 = 0.49$), and 0.343. In the PACF, there is a sizable spike of 0.74 at lag one and all other partial autocorrelations (except for lag 12) are very small.

Under the null hypothesis of an MA(0) process, the standard deviation of r_1 is $T^{-1/2} = 0.1$. Since the sample value of $r_1 = 0.74$ is more than seven standard deviations from zero, we can reject the null that r_1 equals zero. The standard deviation of r_2 is obtained by applying (2.41) to the sampling data, where $s = 2$:

$$\text{Var}(r_2) = [1 + 2(0.74)^2]/100 = 0.021$$

Since $(0.021)^{1/2} = 0.1449$, the sample value of r_2 is approximately four standard deviations from zero; at conventional significance levels, we can reject the null hypothesis that r_2 equals zero. We can similarly test the significance of the other values of the autocorrelations.

As you can see in the second part of the figure, other than $\phi_{1,1}$, all partial autocorrelations (except for lag 12) are less than $2T^{-1/2} = 0.2$. The decay of the ACF and single spike of the PACF give the strong impression of a first-order autoregressive model. If we did not know the true underlying process and happened to be using monthly data, we might be concerned with the significant partial autocorrelation at lag 12. After all, with monthly data we might expect some direct relationship between y_t and y_{t-12} .

Although we know that the data were actually generated from an AR(1) process, it is illuminating to compare the estimates of two different models. Suppose we estimate an AR(1) model and also try to capture the spike at lag 12 with an MA coefficient. Thus, we can consider the two tentative models:

$$\text{Model 1: } y_t = a_1 y_{t-1} + \epsilon_t$$

$$\text{Model 2: } y_t = a_1 y_{t-1} + \epsilon_t + \beta_{12} \epsilon_{t-12}$$

Table 2.2 reports the results of the two estimations.³ The coefficient of model 1 satisfies the stability condition $|a_1| < 1$ and has a low standard error (the associated t -statistic for a null of zero is more than 12). As a useful diagnostic check, we plot the correlogram of the residuals of the fitted model in Figure 2.4. The Q -statistics for these residuals indicate that each one of the autocorrelations is less than two standard deviations from zero. The Ljung–Box Q -statistics of these residuals indicate that as a group, lags 1 through 8, 1 through 16, and 1 through 24 are not significantly different from zero. This is strong evidence that the AR(1) model “fits” the data well. After all, if residual autocorrelations were significant, the AR(1) model would not be utilizing all available information concerning movements in the $\{y_t\}$ sequence. For example, suppose we wanted to forecast y_{t+1} conditioned on all available information up to and including period t . With model 1, the value of y_{t+1} is: $y_{t+1} = a_1 y_t + \epsilon_{t+1}$. Hence, the forecast from model 1 is $a_1 y_t$. If the residual au-

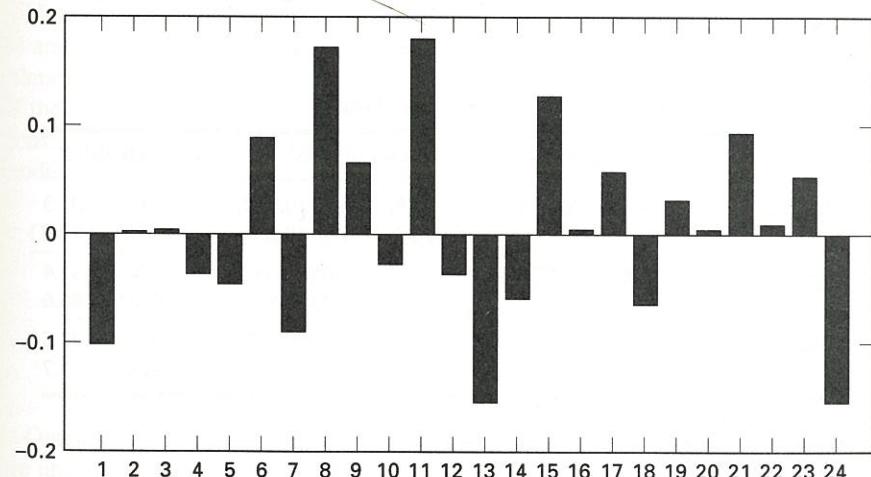
Table 2.2: Estimates of an AR(1) Model

	Model 1 $y_t = a_1 y_{t-1} + \epsilon_t$	Model 2 $y_t = a_1 y_{t-1} + \epsilon_t + \beta_{12} \epsilon_{t-12}$
Degrees of freedom	99	98
Sum of squared residuals	85.21	85.17
Estimated a_1 (standard error)	0.7910 (0.0622)	0.7953 (0.0683)
Estimated β (standard error)		-0.033 (0.1134)
AIC/SBC	AIC = 442.07/SBC = 444.67	AIC = 444.01/SBC = 449.21
Ljung–Box Q -statistics for the residuals (significance level in parentheses)	$Q(8) = 6.43(0.490)$ $Q(16) = 15.86(0.391)$ $Q(24) = 21.74(0.536)$	$Q(8) = 6.48(0.485)$ $Q(16) = 15.75(0.400)$ $Q(24) = 21.56(0.547)$

tocorrelations had been significant, this forecast would not be capturing all the available information set.

Examining the results for model 2, note that both models yield similar estimates for the first-order autoregressive coefficient and associated standard error. However, the estimate for β_{12} is of poor quality; the insignificant t value suggests that it should be dropped from the model. Moreover, comparing the AIC and SBC values of the two models suggests that any benefits of a reduced residual sum of squares are overwhelmed by the detrimental effects of estimating an additional parameter. All these indicators point to the choice of model 1.

Figure 2.4 ACF of residuals from model 1.



Exercise 7 at the end of this chapter entails various estimations using this data set. In this exercise you are asked to show that the AR(1) model performs better than some alternative specifications. It is important that you complete this exercise.

Estimation of an ARMA(1, 1) Model

A second $\{y_t\}$ sequence was constructed to illustrate the estimation of an ARMA(1, 1). Given 100 normally distributed values of the $\{\epsilon_t\}$, 100 values of $\{y_t\}$ were generated using

$$y_t = -0.7y_{t-1} + \epsilon_t - 0.7\epsilon_{t-1}$$

where y_0 and ϵ_0 were both set equal to zero.

Both the sample ACF and PACF from the simulated data (see the second set of graphs in Figure 2.3) are roughly equivalent to those of the theoretical model shown in Figure 2.2. However, if the true data-generating process was unknown, the researcher might be concerned about certain discrepancies. An AR(2) model could yield a sample ACF and PACF similar to those in the figure. Table 2.3 reports the results of estimating the data using the following three models:

$$\text{Model 1: } y_t = a_1 y_{t-1} + \epsilon_t$$

$$\text{Model 2: } y_t = a_1 y_{t-1} + \epsilon_t + \beta_1 \epsilon_{t-1}$$

$$\text{Model 3: } y_t = a_1 y_{t-1} + a_2 y_{t-2} + \epsilon_t$$

In examining Table 2.3, notice that all the estimated values of a_1 are highly significant; each of the estimated values is *at least* eight standard deviations from zero. It is clear that the AR(1) model is inappropriate. The Q -statistics for model 1 indicate that there is significant autocorrelation in the residuals. The estimated ARMA(1, 1) model does not suffer from this problem. Moreover, both the AIC and SBC select model 2 over model 1.

Table 2.3: Estimates of an ARMA(1, 1) Model

	Estimates ^a	Q-Statistics ^b	AIC/SBC
Model 1	$a_1: -0.835 (0.053)$	$Q(8) = 26.19 (0.000)$ $Q(24) = 41.10 (0.001)$	AIC = 507.3 SBC = 509.9
Model 2	$a_1: -0.679 (0.076)$ $\beta_1: -0.676 (0.081)$	$Q(8) = 3.86 (0.695)$ $Q(24) = 14.23 (0.892)$	AIC = 481.4 SBC = 486.6
Model 3	$a_1: -1.16 (0.093)$ $a_2: -0.378 (0.092)$	$Q(8) = 11.44 (0.057)$ $Q(24) = 22.59 (0.424)$	AIC = 492.5 SBC = 497.7

^aStandard errors in parentheses.

^bLjung–Box Q -statistics of the residuals from the fitted model. Significance levels in parentheses.

With the same type of reasoning, model 2 is preferred to model 3. Note that for each model, the estimated coefficients are highly significant and the point estimates imply convergence. Although the Q -statistic at 24 lags indicates that these two models do not suffer from correlated residuals, the Q -statistic at 8 lags indicates serial correlation in the residuals of model 3. Thus, the AR(2) model does not capture short-term dynamics as well as the ARMA(1, 1) model. Also note that the AIC and SBC both select model 2.

Estimation of an AR(2) Model

A third data series was simulated as

$$y_t = 0.7y_{t-1} - 0.49y_{t-2} + \epsilon_t$$

The estimated coefficients of the ACF and PACF of the series are

ACF:

Lag:	1:	0.4655046	-0.1607289	-0.3216291	-0.1077528	-0.0518159	-0.1649841
	7:	-0.0995764	0.1283475	0.1795718	0.0343415	-0.0869808	-0.1133948
	13:	-0.1639613	-0.0579051	0.1151097	0.2540039	0.0460659	-0.1745434
	19:	-0.1503307	0.0100510	0.0318942	-0.0869327	-0.0456013	0.0516806

PACF:

1:	0.4655046	-0.4818344	0.0225089	0.0452089	-0.2528370	-0.1206075
7:	0.1011489	0.0367555	-0.0758751	0.0229422	-0.0203879	-0.1391730
13:	-0.1671389	0.2066915	0.0074996	0.0851050	-0.2156580	0.0131360
19:	-0.0223151	-0.0324078	0.0148130	-0.0609358	0.0374894	-0.1842465

Note the large autocorrelation at lag 16 and large partial autocorrelations at lags 14 and 17. Given the way the process was simulated, the presence of these autocorrelations is due to nothing more than chance. However, an econometrician unaware of the actual data-generating process might be concerned about these autocorrelations. By using the 100 observations of the series, the coefficients of the AR(2) model are estimated as

Coefficient	Estimate	Standard Error	t-Statistic	Significance
a_1	0.692389807	0.089515769	7.73484	0.00000000
a_2	-0.480874620	0.089576524	-5.36831	0.000000055
AIC = 219.87333, SBC = 225.04327				

Overall, the model appears to be adequate. However, the two AR(2) coefficients are unable to capture the correlations at very long lags. For example, the partial au-

tocorrelations of the *residuals* for lags 14 and 17 are both greater than 0.2 in absolute value. The calculated Ljung–Box statistic for 16 lags is 24.6248 (which is significant at the 0.038 level). At this point, it might be tempting to try to model the correlation at lag 16 by including the moving average term $\beta_{16}\epsilon_{t-16}$. Such an estimation results in

Coefficient	Estimate	Standard Error	t-Statistic	Significance
a_1	0.716681247	0.091069451	7.86961	0.000000000
a_2	-0.464999924	0.090958095	-5.11224	0.00000165
β_{16}	0.305813568	0.109936945	2.78172	0.00652182

$$\text{AIC} = 213.40055, \quad \text{SBC} = 221.15545$$

All estimated coefficients are significant and the Ljung–Box Q -statistics are all insignificant at conventional levels. In conjunction with the fact that the AIC and SBC both select this second model, the researcher unaware of the true process might be tempted to conclude that the data-generating process includes a moving average term at lag 16.

A useful check of model adequacy is to split the sample into two parts. If a coefficient is present in the data-generating process, its influence should be seen in both subsamples. If the simulated data are split into two parts, the ACF and PACF using observations 50 through 100 are

ACF:

1:	0.4599901	-0.2066698	-0.2803821	0.0347391	0.0954499	-0.1525615
7:	-0.1329561	0.1017190	0.1807890	0.0274346	-0.0085572	0.0085072
13:	-0.0582938	-0.0876270	0.0426585	0.2051536	0.0643034	-0.1615533
19:	-0.1839472	-0.0504421	-0.0328661	-0.1452760	-0.1071579	-0.0360288

PACF:

1:	0.4599901	-0.5305123	0.1926653	0.0632359	-0.1967718	-0.1295519
7:	0.2337759	-0.0784309	0.0035535	0.0647291	0.1537812	-0.2568410
13:	0.0258134	0.1455056	0.0384080	0.0042725	-0.0542786	-0.0094085
19:	-0.1372884	-0.0763979	-0.0341192	-0.0830831	-0.0287218	-0.2200985

As you can see, the size of partial autocorrelations at lags 14 and 17 is diminished. Now, estimating a pure AR(2) model over this second part of the sample yields

Coefficient	Estimate	Standard Error	t-Statistic	Significance
a_1	0.713855785	0.120541523	5.92207	0.000000031
a_2	-0.537843744	0.120420318	-4.46639	0.00004687

$$\begin{aligned} Q(8) &= 7.8296, & \text{significance level } 0.25085764 \\ Q(16) &= 15.9331, & \text{significance level } 0.31747712 \\ Q(24) &= 26.0648, & \text{significance level } 0.24890909 \end{aligned}$$

All estimated coefficients are significant and the Ljung–Box Q -statistics do not indicate any significant autocorrelations in the residuals. In fact, this model does estimate the actual data-generating process quite well. In this example, the large spurious autocorrelations of the long lags do not appear in the second sample period. Thus, it is hard to maintain that the correlation at lag 16 is meaningful. Most sophisticated practitioners are cautious about trying to fit any model to the very long lags. As you can infer from (2.41), the variance of r_s can be sizable when s is large. Moreover, in small samples, a few “unusual” observations can create the appearance of significant autocorrelations at long lags. Since econometric estimation involves unknown data-generating processes, the more general point is that we always need to be wary of our estimated model. Fortunately, Box and Jenkins (1976) established a set of procedures that can be used to check a model’s adequacy.

8. BOX-JENKINS MODEL SELECTION

The estimates of the AR(1), ARMA(1, 1) and AR(2) models in the previous section illustrate the Box–Jenkins (1976) strategy for appropriate model selection. Box and Jenkins popularized a three-stage method aimed at selecting an appropriate model for the purpose of estimating and forecasting a univariate time series. In the **identification stage**, the researcher visually examines the time plot of the series, autocorrelation function, and partial correlation function. Plotting each observation of the $\{y_t\}$ sequence against t provides useful information concerning outliers, missing values, and structural breaks in the data. Nonstationary variables may have a pronounced trend or appear to meander without a constant long-run mean or variance. Missing values and outliers can be corrected at this point. At one time, the standard practice was to first-difference any series deemed to be nonstationary. Currently, a large literature is evolving that develops formal procedures to check for nonstationarity. We defer this discussion until Chapter 4 and assume that we are working with stationary data. A comparison of the sample ACF and PACF to those of various theoretical ARMA processes may suggest several plausible models. In the **estimation stage**, each of the tentative models is fit and the various a_i and β_i coefficients are examined. In this second stage, the estimated models are compared using the criteria listed below.

Parsimony

A fundamental idea in the Box–Jenkins approach is the principle of **parsimony**. Parsimony (meaning sparseness or stinginess) should come as second nature to economists. Incorporating additional coefficients will necessarily increase fit (e.g., the value of R^2 will increase) at a cost of reducing degrees of freedom. Box and

Jenkins argue that parsimonious models produce better forecasts than overparameterized models. A parsimonious model fits the data well without incorporating any needless coefficients. The aim is to approximate the true data-generating process but not to pin down the exact process. The goal of parsimony suggested eliminating the MA(12) coefficient in the simulated AR(1) model above.

In selecting an appropriate model, the econometrician needs to be aware that several very different models may have very similar properties. As an extreme example, note that the AR(1) model $y_t = 0.5y_{t-1} + \epsilon_t$ has the equivalent infinite-order moving average representation $y_t = \epsilon_t + 0.5\epsilon_{t-1} + 0.25\epsilon_{t-2} + 0.125\epsilon_{t-3} + 0.0625\epsilon_{t-4} + \dots$. In most samples, approximating this MA(∞) process with an MA(2) or MA(3) model will give a very good fit. However, the AR(1) model is the more parsimonious model and is preferred.

Also be aware of the **common factor problem**. Suppose we wanted to fit the ARMA(2, 3) model:

$$(1 - a_1L - a_2L^2)y_t = (1 + \beta_1L + \beta_2L^2 + \beta_3L^3)\epsilon_t \quad (2.43)$$

Also suppose that $(1 - a_1L - a_2L^2)$ and $(1 + \beta_1L + \beta_2L^2 + \beta_3L^3)$ can each be factored as $(1 + cL)(1 + aL)$ and $(1 + cL)(1 + b_1L + b_2L^2)$, respectively. Since $(1 + cL)$ is a common factor to each, (2.43) has the equivalent, but more parsimonious, form:⁴

$$(1 + aL)y_t = (1 + b_1L + b_2L^2)\epsilon_t \quad (2.44)$$

In order to ensure that the model is parsimonious, the various a_i and β_i should all have t -statistics of 2.0 or greater (so that each coefficient is significantly different from zero at the 5% level). Moreover, the coefficients should not be strongly correlated with each other. Highly collinear coefficients are unstable; usually one or more can be eliminated from the model without reducing forecast performance.

Stationarity and Invertibility

The distribution theory underlying the use of the sample ACF and PACF as approximations to those of the true data-generating process assumes that the $\{y_t\}$ sequence is stationary. Moreover, t -statistics and Q -statistics also presume that the data are stationary. The estimated autoregressive coefficients should be consistent with this underlying assumption. Hence, we should be suspicious of an AR(1) model if the estimated value of a_1 is close to unity. For an ARMA(2, q) model, the characteristic roots of the estimated polynomial $(1 - a_1L - a_2L^2)$ should lie outside of the unit circle.

The Box-Jenkins approach also necessitates that the model be **invertible**. Formally, $\{y_t\}$ is invertible if it can be represented by a finite-order or convergent autoregressive process. Invertibility is important because the use of the ACF and PACF implicitly assumes that the $\{y_t\}$ sequence can be well approximated by an

autoregressive model. As a demonstration, consider the simple MA(1) model:

$$y_t = \epsilon_t - \beta_1\epsilon_{t-1} \quad (2.45)$$

so that if $|\beta_1| < 1$,

$$y_t/(1 - \beta_1L) = \epsilon_t$$

or

$$y_t + \beta_1y_{t-1} + \beta_1^2y_{t-2} + \beta_1^3y_{t-3} + \dots = \epsilon_t \quad (2.46)$$

If $|\beta_1| < 1$, (2.46) can be estimated using the Box-Jenkins method. However, if $|\beta_1| \geq 1$, the $\{y_t\}$ sequence cannot be represented by a finite-order AR process; as such, it is not invertible. More generally, for an ARMA model to have a convergent AR representation, the roots of the polynomial $(1 + \beta_1L + \beta_2L^2 + \dots + \beta_qL^q)$ must lie outside of the unit circle. Note that there is nothing “improper” about a non-invertible model. The $\{y_t\}$ sequence implied by $y_t = \epsilon_t - \epsilon_{t-1}$ is stationary in that it has a constant time-invariant mean ($Ey_t = Ey_{t-s} = 0$, a constant time-invariant variance [$\text{var}(y_t) = \text{var}(y_{t-s}) = \sigma^2(1 + \beta_1^2)$], and the autocovariances $\gamma_1 = -\beta_1\sigma^2$ and all other $\gamma_s = 0$). The problem is that the technique does not allow for the estimation of such models. If $\beta_1 = 1$, (2.46) becomes

$$y_t = -y_{t-1} + y_{t-2} - y_{t-3} + y_{t-4} + \dots$$

Clearly, the autocorrelations and partial autocorrelations between y_t and y_{t-s} will never decay.

Goodness of Fit

A good model will fit the data well. Obviously, R^2 and the average of the residual sum of squares are common “goodness-of-fit” measures in ordinary least squares. The problem with these measures is that the “fit” necessarily improves as more parameters are included in the model. Parsimony suggests using the AIC and/or SBC as more appropriate measures of the overall fit of the model. Also, be cautious of estimates that fail to converge rapidly. Most software packages estimate the parameters of an ARMA model using non-linear search procedures. If the search fails to converge rapidly, it is possible that the estimated parameters are unstable. In such circumstances, adding an additional observation or two can greatly alter the estimates.

The third stage in the Box-Jenkins methodology involves **diagnostic checking**. The standard practice is to plot the residuals to look for outliers and evidence of periods in which the model does not fit the data well. If all plausible ARMA models

show evidence of a poor fit during a reasonably long portion of the sample, it is wise to consider using intervention analysis, transfer function analysis, or any other of the multivariate estimation methods discussed in later chapters. If the variance of the residuals is increasing, a logarithmic transformation may be appropriate. Alternatively, you may wish to actually model any tendency of the variance to change using the ARCH techniques discussed in Chapter 3.

It is particularly important that the residuals from an estimated model be serially uncorrelated. Any evidence of serial correlation implies a systematic movement in the $\{y_t\}$ sequence that is not accounted for by the ARMA coefficients included in the model. Hence, any of the tentative models yielding nonrandom residuals should be eliminated from consideration. To check for correlation in the residuals, construct the ACF and PACF of the *residuals* of the estimated model. You can then use (2.41) and (2.42) to determine whether any or all of the residual autocorrelations or partial autocorrelations are statistically significant.⁵ Although there is no significance level that is deemed "most appropriate," be wary of any model yielding (1) several residual correlations that are marginally significant and (2) a *Q*-statistic that is barely significant at the 10% level. In such circumstances, it is usually possible to formulate a better performing model.

In the previous section, recall that the estimated AR(1) model had Box-Ljung *Q*-statistics indicating a possible MA term at lag 12. As a result, we also estimated the model $y_t = 0.7953y_{t-1} + \epsilon_t - 0.033\epsilon_{t-12}$. The procedure of adding another coefficient is called **overfitting**. Overfit a model if the initial ACF and PACF yield ambiguous implications concerning the proper form of the ARMA coefficients. In the first example, the AR(1) model (i.e., model 1) outperformed the ARMA(1, 1) model. Obviously, in other circumstances, the "overfitted" model may outperform the first model. As an additional diagnostic check, some researchers will overfit a model by including a coefficient at some randomly selected lag. If such overfitting greatly affects the model, the estimated model is likely to yield poor forecasts.

If there are sufficient observations, fitting the same ARMA model to each of two subsamples can provide useful information concerning the assumption that the data-generating process is unchanging. In the estimated AR(2) model in the last section, the sample was split in half. In general, suppose you estimated an ARMA(p, q) model using a sample size of T observations. Denote the sum of the squared residuals as SSR. Divide the T observations into two subsamples with t_m observations in the first and $t_n = T - t_m$ observations in the second. Use each subsample to estimate the two models:

$$\begin{aligned} y_t &= a_0(1) + a_1(1)y_{t-1} + \dots + a_p(1)y_{t-p} + \epsilon_t + \beta_1(1)\epsilon_{t-1} + \dots + \beta_q(1)\epsilon_{t-q} && \text{using } t_1, \dots, t_m \\ y_t &= a_0(2) + a_1(2)y_{t-1} + \dots + a_p(2)y_{t-p} + \epsilon_t + \beta_1(2)\epsilon_{t-1} + \dots + \beta_q(2)\epsilon_{t-q} && \text{using } t_{m+1}, \dots, t_T \end{aligned}$$

Let the sum of the squared residuals from each model be SSR_1 and SSR_2 , respectively. To test the restriction that all coefficients are equal [i.e., $a_0(1) = a_0(2)$ and

$a_1(1) = a_1(2)$ and $\dots a_p(1) = a_p(2)$ and $\beta_1(1) = \beta_1(2)$ and $\dots \beta_q(1) = \beta_q(2)$], use an *F*-test and form:⁶

$$F = \frac{(\text{SSR} - \text{SSR}_1 - \text{SSR}_2)/(n)}{(\text{SSR}_1 + \text{SSR}_2)/(T - 2n)} \quad (2.47)$$

where n = number of parameters estimated ($n = p + q + 1$ if an intercept is included and $p + q$ otherwise)
the number of degrees of freedom are $(n, T - 2n)$.

Intuitively, if the restriction that the two sets of coefficients is not binding, the total from the two models (i.e., $\text{SSR}_1 + \text{SSR}_2$) should equal the sum of the squared residuals from the entire sample estimation. Hence, *F* should equal zero. The larger the calculated value of *F*, the more restrictive is the assumption that the two sets of coefficients are equal.

Similarly, the model can be estimated over nearly all the sample period. If we use 20 years of quarterly data, for example, the model might be estimated using only the first 19 years of data. Then, the model can be used to make forecasts of the last year of data. For each period t , the forecast error is the difference between the forecast and known value of y_t . The sum of the squared forecast errors is a useful way to compare the adequacy of alternative models. Those models with poor *out-of-sample* forecasts should be eliminated. Some of the details in constructing out-of-sample forecasts are discussed in the next section.

9. THE FORECAST FUNCTION

Perhaps the most important use of an ARMA model is to forecast future values of the $\{y_t\}$ sequence.⁷ To simplify the discussion, it is assumed that the actual data-generating process and current and past realizations of the $\{\epsilon_t\}$ and $\{y_t\}$ sequences are known to the researcher. First, consider the forecasts from the AR(1) model $y_t = a_0 + a_1y_{t-1} + \epsilon_t$. Updating one period, we obtain

$$y_{t+1} = a_0 + a_1y_t + \epsilon_{t+1}$$

If you know the coefficients a_0 and a_1 , you can forecast y_{t+1} conditioned on the information available at period t as

$$E_t y_{t+1} = a_0 + a_1 y_t \quad (2.48)$$

where $E_t y_{t+j} =$ a short-hand way to write the conditional expectation of y_{t+j} given the information available at t

Formally, $E_t y_{t+j} = E(y_{t+j} | y_t, y_{t-1}, y_{t-2}, \dots, \epsilon_t, \epsilon_{t-1}, \dots)$.

In the same way, since $y_{t+2} = a_0 + a_1 y_{t+1} + \epsilon_{t+2}$, the forecast of y_{t+2} conditioned on the information available at period t is

$$E_t y_{t+2} = a_0 + a_1 E_t y_{t+1}$$

and using (2.48), we obtain

$$E_t y_{t+2} = a_0 + a_0 a_1 + a_1^2 y_t \quad (2.49)$$

It should not require too much effort to convince yourself that

$$E_t y_{t+3} = a_0 + a_0 a_1 + a_0 a_1^2 + a_1^3 y_t$$

and in general,

$$E_t y_{t+j} = a_0(1 + a_1 + a_1^2 + \dots + a_1^{j-1}) + a_1^j y_t \quad (2.50)$$

Equation (2.50)—called the **forecast function**—yields the j -step ahead forecasts for each value y_{t+j} . Since $|a_1| < 1$, (2.50) yields a convergent sequence of forecasts. If we take the limit of $E_t y_{t+j}$ as $j \rightarrow \infty$, we find that $E_t y_{t+j} \rightarrow a_0/(1 - a_1)$. This result is really quite general. *For any stationary ARMA model, the conditional forecast of y_{t+j} converges to the unconditional mean as $j \rightarrow \infty$.* Unfortunately, the forecasts from an ARMA model will not be perfectly accurate. Forecasting from time period t , we can define the j -step ahead forecast error, $f_t(j)$ —as the difference between the realized value of y_{t+j} and forecasted value:

$$f_t(j) \equiv y_{t+j} - E_t y_{t+j}$$

Hence, the one-step ahead forecast error is: $f_t(1) = y_{t+1} - E_t y_{t+1} = \epsilon_{t+1}$ (i.e., the “unforecastable” portion of y_{t+1} given the information available in t). To find the two-step ahead forecast error, we need to form $f_t(2) = y_{t+2} - E_t y_{t+2}$. Since $y_{t+2} = a_0 + a_1 a_0 + a_1^2 y_t + \epsilon_{t+2} + a_1 \epsilon_{t+1}$ and $E_t y_{t+2} = a_0 + a_1 a_0 + a_1^2 y_t$, it follows that

$$f_t(2) = \epsilon_{t+2} + a_1 \epsilon_{t+1}$$

You should take a few moments to demonstrate that for the AR(1) model, the j -step ahead forecast error is given by

$$f_t(j) = \epsilon_{t+j} + a_1 \epsilon_{t+j-1} + a_1^2 \epsilon_{t+j-2} + a_1^3 \epsilon_{t+j-3} + \dots + a_1^{j-1} \epsilon_{t+1} \quad (2.51)$$

Equation (2.51) shows that the forecasts from (2.50) yield unbiased estimates of each value y_{t+j} . The proof is trivial; since $E_t \epsilon_{t+j} = E_t \epsilon_{t+j-1} = \dots = E_t \epsilon_{t+1} = 0$, the conditional expectation of (2.51) is $E_t f_t(j) = 0$. Since the expected value of the forecast error is zero, the forecasts are unbiased.

Although unbiased, the forecasts from an ARMA model are necessarily inaccurate. To find the variance of the forecast error, continue to assume that the elements of the $\{\epsilon_t\}$ sequence are independent with variance σ^2 . Hence, from (2.51) the variance of the forecast error is

$$\text{Var}[f_t(j)] = \sigma^2 [1 + a_1^2 + a_1^4 + a_1^6 + \dots + a_1^{2(j-1)}] \quad (2.52)$$

Since the one-step forecast error variance is σ^2 , the two-step ahead forecast error variance is $\sigma^2(1 + a_1^2)$, etc. The essential point to note is that the variance of the forecast error is an increasing function of j . As such, you can have more confidence in short-term rather than long-term forecasts. In the limit as $j \rightarrow \infty$, the forecast error variance converges to $\sigma^2/(1 - a_1^2)$; hence, the forecast error variance converges to the unconditional variance of the $\{y_t\}$ sequence.

Moreover, assuming that the $\{\epsilon_t\}$ sequence is normally distributed, you can place confidence intervals around the forecasts. The one-step ahead forecast of y_{t+1} is $a_0 + a_1 y_t$ and the variance is σ^2 . As such, the 95% confidence interval for the one-step ahead forecast can be constructed as

$$a_0 + a_1 y_t \pm 1.96\sigma$$

In the same way, the two-step ahead forecast is $a_0(1 + a_1) + a_1^2 y_t$, and (2.52) indicates that $\text{var}[f_t(2)]$ is $\sigma^2(1 + a_1^2)$. Thus, the 95% confidence interval for the two-step ahead forecast is

$$a_0(1 + a_1) + a_1^2 y_t \pm 1.96\sigma(1 + a_1^2)^{1/2}$$

Of course, if there is any uncertainty concerning the parameters, the confidence intervals will be wider than those reported here.

Iterative Forecasts

The derivation of (2.50)—the forecast function for an AR(1) model—relied on forward iteration. To generalize the discussion, it is possible to use the iterative technique to derive the forecast function for any ARMA(p, q) model. To keep the algebra simple, consider the ARMA(2, 1) model:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \epsilon_t + \beta_1 \epsilon_{t-1} \quad (2.53)$$

Updating one period yields

$$y_{t+1} = a_0 + a_1 y_t + a_2 y_{t-1} + \epsilon_{t+1} + \beta_1 \epsilon_t$$

If we continue to assume that (1) all coefficients are known; (2) all variables subscripted $t, t-1, t-2$, etc. are known at period t ; and (3) $E_t \epsilon_{t+j} = 0$ for $j > 0$, the conditional expectation of y_{t+1} is

$$E_t y_{t+1} = a_0 + a_1 y_t + a_2 y_{t-1} + \beta_1 \epsilon_t \quad (2.54)$$

Equation (2.54) is the one-step ahead forecast of y_{t+1} . To find the two-step ahead forecast, update (2.53) by two periods:

$$y_{t+2} = a_0 + a_1 y_{t+1} + a_2 y_t + \epsilon_{t+2} + \beta_1 \epsilon_{t+1}$$

The conditional expectation of y_{t+2} is

$$E_t y_{t+2} = a_0 + a_1 E_t y_{t+1} + a_2 y_t \quad (2.55)$$

Equation (2.55) expresses the two-step ahead forecast in terms of the one-step ahead forecast and current value of y_t . Combining (2.54) and (2.55) yields

$$\begin{aligned} E_t y_{t+2} &= a_0 + a_1(a_0 + a_1 y_t + a_2 y_{t-1} + \beta_1 \epsilon_t) + a_2 y_t \\ &= a_0(1 + a_1) + (a_1^2 + a_2)y_t + a_1 a_2 y_{t-1} + a_1 \beta_1 \epsilon_t \end{aligned}$$

You should be able to demonstrate that the three-step ahead forecast is

$$\begin{aligned} E_t y_{t+3} &= a_0 + a_1 E_t y_{t+2} + a_2 E_t y_{t+1} \\ &= a_0 + a_1 \{a_0(1 + a_1) + [a_1^2 + a_2]y_t + a_1 a_2 y_{t-1} + a_1 \beta_1 \epsilon_t\} + \\ &\quad a_2(a_0 + a_1 y_t + a_2 y_{t-1} + \beta_1 \epsilon_t) \\ &= a_0(1 + a_1 + a_1^2 + a_2) + (a_1^3 + 2a_1 a_2)y_t + (a_1^2 a_2 + a_2^2)y_{t-1} + \beta_1(a_1^2 + a_2)\epsilon_t \quad (2.56) \end{aligned}$$

Finally, all j -step ahead forecasts can be obtained from

$$E_t y_{t+j} = a_0 + a_1 E_t y_{t+j-1} + a_2 E_t y_{t+j-2}, \quad j \geq 2 \quad (2.57)$$

Equations (2.56) and (2.57) suggest that the forecasts will satisfy a second-order difference equation. As long as the characteristic roots of (2.57) lie inside the unit circle, the forecasts will converge to the unconditional mean $a_0/(1 - a_1 - a_2)$.

An Alternative Derivation of the Forecast Function

Instead of using the iterative technique, it is often preferable to derive the forecast function using the solution methodology discussed in Section 4 of Chapter 1. For any ARMA(p, q) model, the solution technique entails (1) finding all homogeneous solutions; (2) finding the particular solution; (3) forming the general solution as the sum of the homogeneous and particular solutions; and (4) imposing the initial conditions. This solution methodology will express y_t in terms of the p initial conditions y_0, y_1, \dots, y_{p-1} and q initial values $\epsilon_0, \epsilon_1, \dots, \epsilon_{q-1}$. The only twist is that the forecast function expresses y_{t+j} in terms of $y_t, y_{t-1}, \dots, y_{t-p+1}$ and $\epsilon_t, \epsilon_{t-1}, \dots, \epsilon_{t-q+1}$. To illustrate the appropriate modification of the time subscripts, consider the AR(2) model:

$$y_t = 3 + 0.9y_{t-1} - 0.2y_{t-2} + \epsilon_t$$

In Section 8 of Chapter 1, it was shown that the solution is

$$y_t = 10 + (0.4)^t [5(y_0 - 10) - 10(y_1 - 10)] + (0.5)^t [10(y_1 - 10) - 4(y_0 - 10)] + \sum_{i=0}^{t-2} \alpha_i \epsilon_{t-i}$$

where the values of α_i satisfy $\alpha_i = 5(0.5)^i - 4(0.4)^i$.

The problem is to modify this equation so as to express y_{t+j} in terms of $y_t, y_{t-1}, \dots, y_{t-j+1}$, and the $\{\epsilon_t\}$ sequence. Updating by j periods, we find

$$\begin{aligned} y_{t+j} &= 10 + (0.4)^j [5(y_{t-1} - 10) - 10(y_t - 10)] \\ &\quad + (0.5)^j [10(y_t - 10) - 4(y_{t-1} - 10)] + \sum_{i=0}^{j-1} \alpha_i \epsilon_{t+j-i} \end{aligned}$$

Taking the conditional expectation of y_{t+j} yields the forecast function:

$$E_t y_{t+j} = 10 + (0.4)^j [5(y_{t-1} - 10) - 10(y_t - 10)] + (0.5)^j [10(y_t - 10) - 4(y_{t-1} - 10)]$$

Obviously, as j increases, the forecast approaches the unconditional mean of 10. For practice, try the ARMA(1, 1) model:

$$y_t = a_0 + a_1 y_{t-1} + \epsilon_t + \beta_1 \epsilon_{t-1}$$

where $\{\epsilon_t\}$ is a white-noise process, $|a_1| < 1$, and there is a given initial condition for y_0 .

You should recognize that the homogeneous equation $y_t - a_1 y_{t-1} = 0$ has the solution $A(a_1)^t$, where A is an arbitrary constant. Next, use lag operators to obtain the particular solution as

$$y_t = a_0/(1 - a_1) + \epsilon_t/(1 - a_1 L) + \beta_1 \epsilon_{t-1}/(1 - a_1 L) \quad (2.58)$$

so that the general solution is

$$y_t = a_0/(1 - a_1) + \sum_{i=0}^{\infty} a_1^i \epsilon_{t-i} + \beta_1 \sum_{i=0}^{\infty} a_1^i \epsilon_{t-1-i} + A a_1^t \quad (2.59)$$

Now impose the initial condition for y_0 . Since (2.59) must hold for all periods, including period zero, it follows that

$$y_0 = a_0/(1 - a_1) + \sum_{i=0}^{\infty} a_1^i \epsilon_{-i} + \beta_1 \sum_{i=0}^{\infty} a_1^i \epsilon_{-1-i} + A \quad (2.60)$$

Solving (2.60) for A eliminates the arbitrary constant. Combining (2.59) and (2.60), we get

$$y_t = a_0/(1-a_1) + \sum_{i=0}^{\infty} a_1^i \epsilon_{t-i} + \beta_1 \sum_{i=0}^{\infty} a_1^i \epsilon_{t-1-i} + \left[y_0 - a_0/(1-a_1) - \sum_{i=0}^{\infty} a_1^i \epsilon_{-i} - \beta_1 \sum_{i=0}^{\infty} a_1^i \epsilon_{-1-i} \right] a_1^t$$

so that

$$y_t = a_0/(1-a_1) + \sum_{i=0}^{t-1} a_1^i \epsilon_{t-i} + \beta_1 \sum_{i=0}^{t-1} a_1^i \epsilon_{t-1-i} + [y_0 - a_0/(1-a_1)] a_1^t \quad (2.61)$$

To this point, (2.61) is simply the general solution to the stochastic difference equation represented by an ARMA(1, 1) process. This solution expresses the current value of y_t in terms of the constants a_0 , a_1 , and β_1 , $\{\epsilon_t\}$ sequence, and initial value of y_0 .

The important point is that (2.61) can be used to forecast y_t conditioned on information available at period zero. Given $E_0 \epsilon_i = 0$ for $i > 0$, it follows that

$$E_0 y_t = a_0/(1-a_1) + \beta_1 a_1^{t-1} \epsilon_0 + [y_0 - a_0/(1-a_1)] a_1^t \quad (2.62)$$

Equation (2.62) can be viewed as the *t-step ahead* forecast function given information available in period zero. To form the *j-step ahead* forecasts conditioned on information available at t , first change the time subscript in (2.62) so that the *j-step ahead* forecasts are

$$\begin{aligned} E_0 y_j &= a_0/(1-a_1) + \beta_1 a_1^{j-1} \epsilon_0 + [y_0 - a_0/(1-a_1)] a_1^j \\ &= [a_0/(1-a_1)](1-a_1^j) + \beta_1 a_1^{j-1} \epsilon_0 + y_0 a_1^j \end{aligned} \quad (2.63)$$

Next, update (2.63) by t periods so that

$$E_t y_{t+j} = [a_0/(1-a_1)](1-a_1^j) + \beta_1 a_1^{j-1} \epsilon_t + y_t a_1^j \quad (2.64)$$

Equation (2.64) is in the desired form; (2.64) expressed the forecast of y_{t+j} conditioned on information available at period t . The various *j-step ahead* forecasts are

$$\begin{aligned} E_t y_{t+1} &= a_0 + \beta_1 \epsilon_t + a_1 y_t \\ E_t y_{t+2} &= [a_0/(1-a_1)](1-a_1^2) + \beta_1 a_1 \epsilon_t + y_t a_1^2 \\ E_t y_{t+3} &= [a_0/(1-a_1)](1-a_1^3) + \beta_1 a_1^2 \epsilon_t + y_t a_1^3 \\ &\dots \end{aligned}$$

Given that $|a_1| < 1$, the limiting value of the forecast as $j \rightarrow \infty$ is the unconditional mean: $\lim E_t y_{t+j} = a_0/(1-a_1)$.

As a check, you can compare (2.64) to (2.50); after all, the AR(1) and ARMA(1, 1) models are equivalent if $\beta_1 = 0$. If $\beta_1 = 0$, (2.64) becomes

$$E_t y_{t+j} = [a_0/(1-a_1)](1-a_1^j) + y_t a_1^j \quad (2.65)$$

Note that (2.65) is identical to (2.50); for $|a_1| < 1$,

$$a_0 \sum_{i=0}^{j-1} a_1^i = [a_0/(1-a_1)](1-a_1^j)$$

The example illustrates the basic point that for any ARMA(p, q) model, the forecast function for y_{t+j} will have the form

$$E_t y_{t+j} = \alpha_0(j) + \alpha_1(j)y_t + \alpha_2(j)y_{t-1} + \dots + \alpha_p(j)y_{t-p+1} + \gamma_1(j)\epsilon_t + \dots + \gamma_q\epsilon_{t-q+1} \quad (2.66)$$

where all values of $\alpha_i(j)$ and $\gamma_i(j)$ are undetermined coefficients.

The notation $\alpha_i(j)$ and $\gamma_i(j)$ is designed to stress the point that the coefficients are a function of j . Since we are working with stationary and invertible processes, we know the nature of the solution is such that as $j \rightarrow \infty$, $\alpha_0(j) \rightarrow a_0/(1-\sum a_i)$, $\alpha_i(j) \rightarrow 0$, and that $\sum |\gamma_i(j)|^2$ is finite.

In practice, you will not know the actual order of the ARMA process or coefficients of that process. Instead, to create out-of-sample forecasts, it is necessary to use the estimated coefficients from what you believe to be the most appropriate form of an ARMA model. The rule of thumb is that forecasts from an ARMA model should never be trusted if the model is estimated with fewer than 50 observations. Suppose you have T observations of the $\{y_t\}$ sequence and choose to fit an ARMA(2, 1) model to the data. Let a hat or caret (i.e.: a $\hat{\cdot}$) over a parameter denote the estimated value of a parameter and let $\{\hat{\epsilon}_t\}$ denote the residuals of the estimated model. Hence, the estimated AR(2, 1) model can be written as

$$y_t = \hat{a}_0 + \hat{a}_1 y_{t-1} + \hat{a}_2 y_{t-2} + \hat{\epsilon}_t + \hat{\beta}_1 \hat{\epsilon}_{t-1}$$

Given that the sample contains T observations, the out-of-sample forecasts are easily constructed. For example, you can use (2.54) to forecast the value of y_{T+1} as

$$E_T y_{T+1} = \hat{a}_0 + \hat{a}_1 y_T + \hat{a}_2 y_{T-1} + \hat{\beta}_1 \hat{\epsilon}_T \quad (2.67)$$

Given the estimated values of \hat{a}_0 , \hat{a}_1 , and \hat{a}_2 , (2.67) can easily be constructed using the actual values y_T , y_{T-1} , and $\hat{\epsilon}_T$ (i.e., the last residual of your estimated model). Similarly, the forecast of y_{T+2} can be constructed as

$$E_T y_{T+2} = \hat{a}_0 + \hat{a}_1 E_T y_{T+1} + \hat{a}_2 y_T$$

where $E_T y_{T+1}$ = the forecast from (2.67)

Given these two forecasts, all subsequent forecasts can be obtained from the difference equation:

$$E_T y_{T+j} = a_0 + a_1 E_T y_{T+j-1} + a_2 E_T y_{T+j-2} \quad \text{for } j \geq 2$$

10. A MODEL OF THE WPI

The ARMA estimations performed in Section 8 were almost too straightforward. In practice, we rarely find a data series precisely conforming to a theoretical ACF or PACF. This section is intended to illustrate some of the ambiguities frequently encountered in the Box-Jenkins technique. These ambiguities may lead two equally skilled econometricians to estimate and forecast a series using very different ARMA processes. Many view the necessity to rely on the researcher's judgment and experience as a serious weakness of a procedure that is designed to be scientific.

It is useful to illustrate the Box-Jenkins modeling procedure by estimating a quarterly model of the U.S. Wholesale Price Index (WPI). The file labeled WPI.WK1 on the data disk contains the data used in this section. Exercise 10 at the end of this chapter will help you to reproduce the results reported below.

The top graph of Figure 2.5 clearly reveals that there is little point in modeling the series as being stationary; there is a decidedly positive trend or drift throughout the period 1960:I to 1990:IV. The first difference of the series seems to have a constant mean, although inspection of the middle graph suggests that the variance is an increasing function of time. As shown in the bottom graph of the same figure, the first difference of the logarithm (denoted by $\Delta l w p i$) is the most likely candidate to be covariance stationary. The large volatility of the WPI accompanying the oil price shocks in the 1970s should make us somewhat wary of the assumption that the process is covariance stationary. At this point, some researchers would make additional transformations intended to reduce the volatility exhibited in the 1970s. However, it seems reasonable to estimate a model of the $\{\Delta l w p i_t\}$ sequence. As always, you should maintain a healthy skepticism of the accuracy of your model.

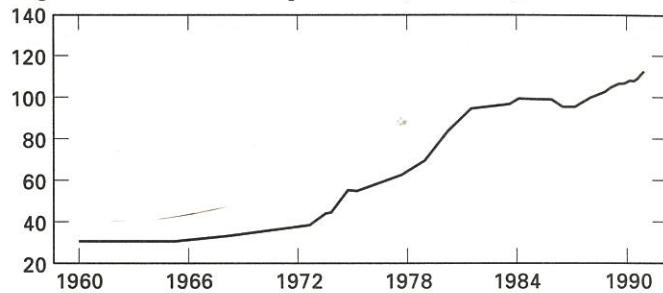
Before reading on, you should examine the autocorrelation and partial autocorrelation functions of the $\{\Delta l w p i_t\}$ sequence shown in Figure 2.6. Try to identify the tentative models that you would want to estimate. In making your decision, note the following:

1. The ACF and PACF converge to zero reasonably quickly. We do not want to *overdifference* the data and try to model the $\{\Delta^2 l w p i_t\}$ sequence.
2. The theoretical ACF of a pure MA(q) process cuts off to zero at lag q and the theoretical ACF of an AR(1) model decays geometrically. Examination of the

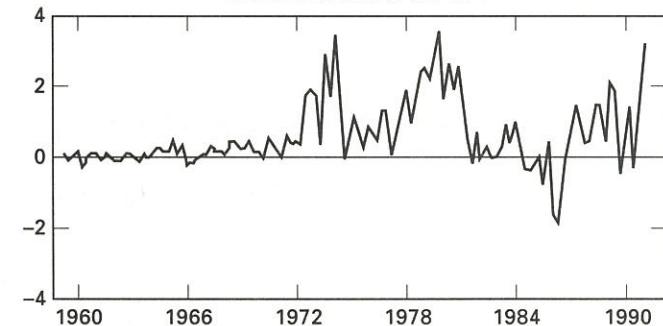
two graphs of Figure 2.6 suggests that neither of these specifications seems appropriate for the sample data.

3. The PACF is such that $\phi_{1,1} = 0.609$ and cuts off to 0.252 abruptly (i.e., $\phi_{2,2} = 0.252$). Overall, the PACF suggests that we should consider models such as $p = 1$ and $p = 2$. The ACF is suggestive of an AR(2) process or a process with both autoregressive and moving average components.
4. Note the jump in ACF at lag 4 and the small spike in the PACF at lag 4 ($\phi_{4,4} = 0.198$). Since we are using quarterly data, we might want to incorporate a seasonal factor at lag 4.

Figure 2.5 U.S. wholesale price index (1985 = 100).



First difference of the WPI



Logarithmic change in the WPI

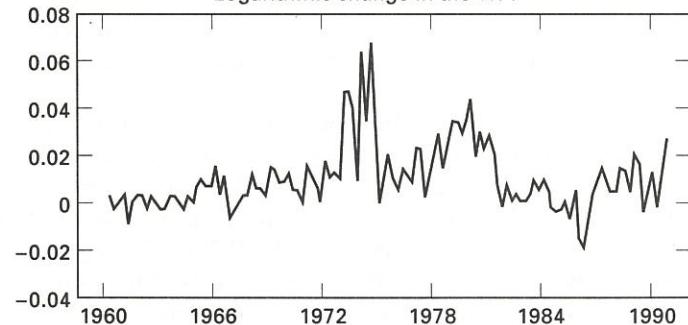
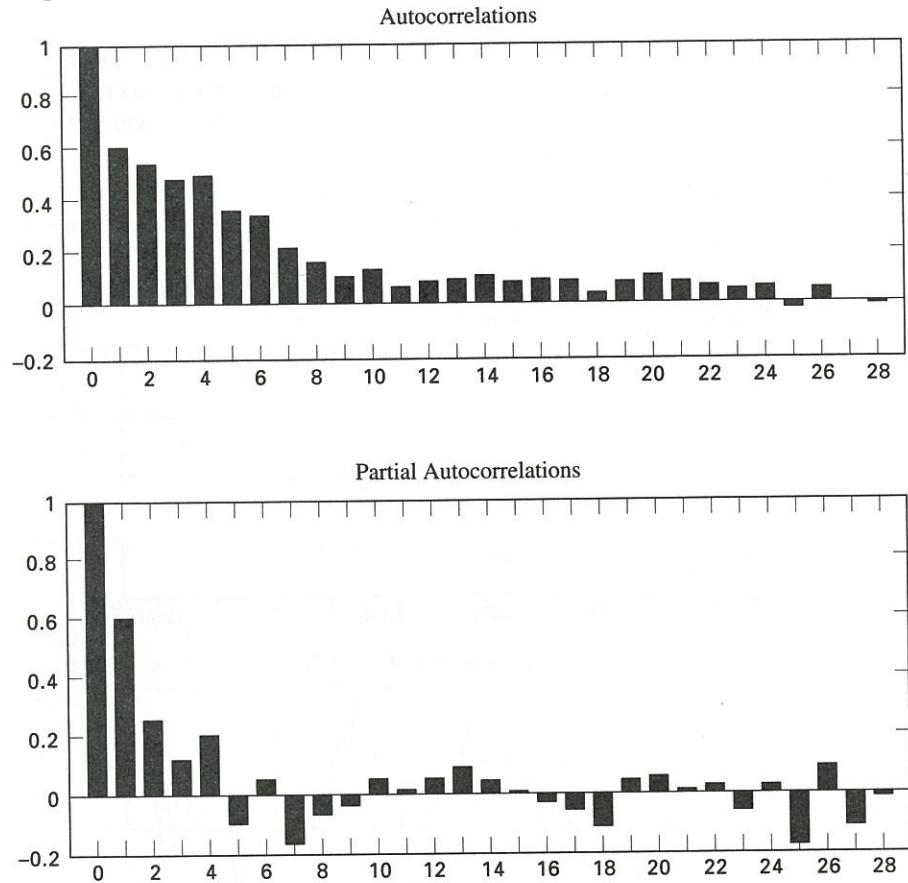


Figure 2.6 ACF and PACF for the logarithmic change in the WPI.



Points 1 to 4 suggest an ARMA(1, 1) or AR(2) model. In addition, we might want to consider models with a seasonal term at lag 4. Since computing time is inexpensive, we can estimate a variety of models and compare their results. Table 2.4 reports estimates of five tentative models; note the following points:

1. The estimated AR(1) model confirms our analysis in the identification stage. Although the estimated value of a_1 (0.618) is less than unity in absolute value and more than eight standard deviations from zero, the AR(1) specification is inadequate. Forming the Ljung–Box Q -statistic for 12 lags of the residuals yields a value of 23.6; we can reject the null that $Q = 0$ at the 1% significance level. Hence, the lagged residuals of this model exhibit substantial serial autocorrelation. Then we must eliminate this model from consideration.

2. The AR(2) model is an improvement over the AR(1) specification. The estimated coefficients ($a_1 = 0.456$ and $a_2 = 0.258$) are each significantly different from zero at the 1% level and imply characteristic roots in the unit circle. Q -statistics indicate that the autocorrelations of the residuals are not statistically significant. As measured by the AIC, the fit of the AR(2) model is superior to that of the AR(1); the SBC is the same for the two models. Overall, the AR(2) model dominates the AR(1) specification.
3. The ARMA(1, 1) specification dominates the AR(2) model. The estimated coefficients are of high quality (with t values of 14.9 and -4.22). The estimated value of a_1 is positive but less than unity, and the Q -statistics indicate that the autocorrelations of the residuals are not statistically significant. Moreover, all goodness-of-fit measures select the ARMA(1, 1) specification over the AR(2) model. Thus, there is little reason to maintain the AR(2) specification.

Table 2.4: Estimates of the WPI (Logarithmic First Differences)

	$p = 1$ $q = 0$	$p = 2$ $q = 0$	$p = 1$ $q = 1$	$p = 1$ $q = 1, 4$	$p = 1$ $q = 2$
a_0	0.011 (4.14)	0.011 (3.31)	0.012 (2.63)	0.011 (2.76)	0.012 (2.62)
a_1	0.618 (8.54)	0.456 (5.11)	0.887 (14.9)	0.791 (9.21)	0.887 (13.2)
a_2		0.258 (2.89)			
β_1			-0.484 (-4.22)	-0.409 (-3.62)	-0.483 (-4.19)
β_2					-0.002 (-0.019)
β_4				0.315 (3.36)	
SSR	0.0156	0.0145	0.0141	0.0134	0.0141
AIC	-503.3	-506.1	-513.1	-518.2	-511.1
SBC	-497.7	-497.7	-504.7	-507.0	-499.9
$Q(12)$	23.6 (0.008)	11.7 (0.302)	11.7 (0.301)	4.8 (0.898)	11.7 (0.301)
$Q(24)$	28.6 (0.157)	15.6 (0.833)	15.4 (0.842)	9.3 (0.991)	15.3 (0.841)
$Q(30)$	40.1 (0.082)	22.8 (0.742)	22.7 (0.749)	14.8 (0.972)	22.6 (0.749)

Notes: Each coefficient is reported with the associated t -statistic for the null hypothesis that the estimated value is equal to zero.

SSR is the sum of squared residuals.

$Q(n)$ reports the Ljung–Box Q -statistic for the autocorrelations of the n residuals of the estimated model. With 122 observations, $T/4$ is approximately equal to 30. Significance levels are in parentheses.

4. In order to account for the possibility of seasonality, we estimated the ARMA(1, 1) model with an additional moving average coefficient at lag 4, that is, we estimated a model of the form $y_t = a_0 + a_1 y_{t-1} + \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_4 \epsilon_{t-4}$. More sophisticated seasonal patterns are considered in the next section. For now, note that the additive expression $\beta_4 \epsilon_{t-4}$ is often preferable to an additive autoregressive term of the form $a_4 y_{t-4}$. For truly seasonal shocks, the expression $\beta_4 \epsilon_{t-4}$ best captures spikes—not decay—at the quarterly lags. The coefficients of the estimated ARMA[1, (1, 4)] model are all highly significant with t -statistics of 9.21, -3.62, and 3.36.⁸ The Q -statistics are all very low, implying that the autocorrelations of the residuals are statistically equal to zero. Moreover, the AIC and SBC strongly select this model over the ARMA(1, 1) model.
5. In contrast, the ARMA(1, 2) contains a superfluous coefficient. The t -statistic for β_2 is sufficiently low that we should eliminate this model.

Having identified and estimated a plausible model, we want to perform additional diagnostic checks of model adequacy. Due to the high volatility in the 1970s, the sample was split into the two subperiods: 1960:I to 1971:IV and 1972:I to 1990:IV. Model estimates for each subperiod are

$$\Delta lwp_i = 0.004 + 0.641 \Delta lwp_{i-1} + \epsilon_i - 0.351 \epsilon_{i-1} + 0.172 \epsilon_{i-4} \quad (1960:\text{I}-1971:\text{IV})$$

and

$$\Delta lwp_i = 0.016 + 0.753 \Delta lwp_{i-1} + \epsilon_i - 0.394 \epsilon_{i-1} + 0.335 \epsilon_{i-4} \quad (1972:\text{I}-1990:\text{IV})$$

The coefficients of the two models appear to be quite similar; we can formally test for the equality of coefficients using (2.47). Respectively, the sums of squared residuals for the two models are $SSR_1 = 0.001359$ and $SSR_2 = 0.011681$, and from Table 2.4 we can see that $SSR = 0.0134$. Since $T = 122$ and $n = 4$ (including the intercept means there are four estimated coefficients), (2.47) becomes

$$\begin{aligned} F &= [(0.0134 - 0.001359 - 0.011681)/4]/[0.001359 + 0.011681]/(122-8) \\ &= 0.78681 \end{aligned}$$

With 4 degrees of freedom in the numerator and 114 in the denominator, we cannot reject the null of no structural change in the coefficients (i.e., we accept the hypothesis that there is no change in the structural coefficients).

As a final check, out-of-sample forecasts were constructed for each of the two models. By using additional data through 1992:II, the variance of the out-of-sample forecast errors of the ARMA(1, 1) and ARMA[1, (1,4)] models were calculated to be 0.00011 and 0.00008, respectively. Clearly, all the diagnostics select the ARMA[1, (1,4)] model. Although the ARMA[1, (1,4)] model appears to be adequate, other researchers might have selected a decidedly different model. Consider some of the alternatives listed below:

- Trends:** Although the logarithmic change of the WPI wholesale appears to be stationary, the ACF converges to zero rather slowly. Moreover, both the ARMA(1, 1) and ARMA[1, (1,4)] models yield estimated values of a_1 (0.887 and 0.791, respectively) that are close to unity. Some researchers might have chosen to model the second difference of the series. Others might have detrended the data using a deterministic time trend. Chapter 4 discusses formal tests for the appropriate form of the trend.
- The seasonality of the data was modeled using a moving average term at lag 4. However, there are many other plausible ways to model the seasonality in the data, as discussed in the next section. For example, many computer programs are capable of estimating multiplicative seasonal coefficients. Consider the multiplicative seasonal model:

$$(1 - a_1 L)y_t = (1 + \beta_1 L)(1 + \beta_4 L^4)\epsilon_t$$

Here, the seasonal expression $\beta_4 \epsilon_{t-4}$ enters the model in a multiplicative, rather than a linear, fashion. Experimenting with various multiplicative seasonal coefficients might be a way to improve forecasting performance.

- Given the volatility of the $\{\Delta lwp_i\}$ sequence during the 1970s, the assumption of a constant variance might not be appropriate. Transforming the data using a square root, rather than the logarithm, might be more appropriate. A general class of transformations was proposed by Box and Cox (1964). Suppose that all values of $\{y_t\}$ are positive so that it is possible to construct the transformed $\{y_t^*\}$ sequence as

$$\begin{aligned} y_t^* &= (y_t^\lambda - 1)/\lambda, & \lambda \neq 0 \\ &= \ln(y_t), & \lambda = 0 \end{aligned}$$

The common practice is to transform the data using a preselected value of λ . Selecting a value of λ that is close to zero acts to “smooth” the sequence. As in the WPI example (which simply set $\lambda = 0$), an ARMA model can be fit to the transformed data. Although some software programs have the capacity to simultaneously estimate λ along with the other parameters of the ARMA model, this approach has fallen out of fashion. Instead, it is possible to actually model the variance using the methods discussed in Chapter 3.

11. SEASONALITY

Many economic processes exhibit some form of seasonality. The agricultural, construction, and travel sectors have obvious seasonal patterns resulting from their dependence on the weather. Similarly, the Thanksgiving–Christmas holiday season has a pronounced influence on the retail trade. In fact, the seasonal variation of some series may account for the preponderance of its total variance. Forecasts that

ignore important seasonal patterns will have a high variance. In the last section, we saw how the inclusion of a four-quarter seasonal factor could help improve the model of the WPI. This section expands that discussion by illustrating some of the techniques that can be used to identify seasonal patterns.

Too many people fall into the trap of ignoring seasonality if they are working with **deseasonalized** or **seasonally adjusted** data. Suppose you collect a data set that the U.S. Bureau of the Census has “seasonally adjusted” using its X-11 method.⁹ In principle, your seasonally adjusted data should have the seasonal pattern removed. However, caution is necessary. Although a standardized procedure may be necessary for a government agency reporting hundreds of series, the procedure might not be best for an individual wanting to model a single series. Even if you use seasonally adjusted data, a seasonal pattern might remain. This is particularly true if you do not use the entire span of data; the portion of the data used in your study can display more (or less) seasonality than the overall span. There is another important reason to be concerned about seasonality when using deseasonalized data. Implicit in any method of seasonal adjustment is a two-step procedure. First, the seasonality is removed, and second, the autoregressive and moving average coefficients are estimated using Box-Jenkins techniques. As surveyed in Bell and Hillmer (1984), often the seasonal and ARMA coefficients are best identified and estimated jointly. In such circumstances, it is wise to avoid using seasonally adjusted data.

Models of Seasonal Data

The Box-Jenkins technique for modeling seasonal data is no different from that of nonseasonal data. The twist introduced by seasonal data of period s is that the seasonal coefficients of the ACF and PACF appear at lags $s, 2s, 3s, \dots$, rather than at lags $1, 2, 3, \dots$. For example, two purely seasonal models for quarterly data might be

$$y_t = a_4 y_{t-4} + \epsilon_t, \quad |a_4| < 1 \quad (2.68)$$

and

$$y_t = \epsilon_t + \beta_4 \epsilon_{t-4} \quad (2.69)$$

You can easily convince yourself that the theoretical correlogram for (2.68) is such that $\rho_i = (a_4)^{i/4}$ if $i/4$ is an integer, and $\rho_i = 0$ otherwise; thus, the ACF exhibits decay at lags 4, 8, 12, \dots . For model (2.69), the ACF exhibits a single spike at lag 4 and all other correlations are zero.

In practice, identification will be complicated by the fact that the seasonal pattern will interact with the nonseasonal pattern in the data. The ACF and PACF for a combined seasonal/nonseasonal process will reflect both elements. Note that the final model of the wholesale price index estimated in the last section had the form

$$y_t = a_1 y_{t-1} + \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_4 \epsilon_{t-4} \quad (2.70)$$

Alternatively, an autoregressive coefficient at lag 4 might have been used to capture the seasonality:

$$y_t = a_1 y_{t-1} + a_4 y_{t-4} + \epsilon_t + \beta_1 \epsilon_{t-1} \quad (2.71)$$

Both these methods treat the seasonal coefficients additively; an AR or MA coefficient is added at the seasonal period. **Multiplicative seasonality** allows for the interaction of the ARMA and seasonal effects. Consider the multiplicative specifications:

$$(1 - a_1 L) y_t = (1 + \beta_1 L)(1 + \beta_4 L^4) \epsilon_t \quad (2.72)$$

$$(1 - a_1 L)(1 - a_4 L^4) y_t = (1 + \beta_1 L) \epsilon_t \quad (2.73)$$

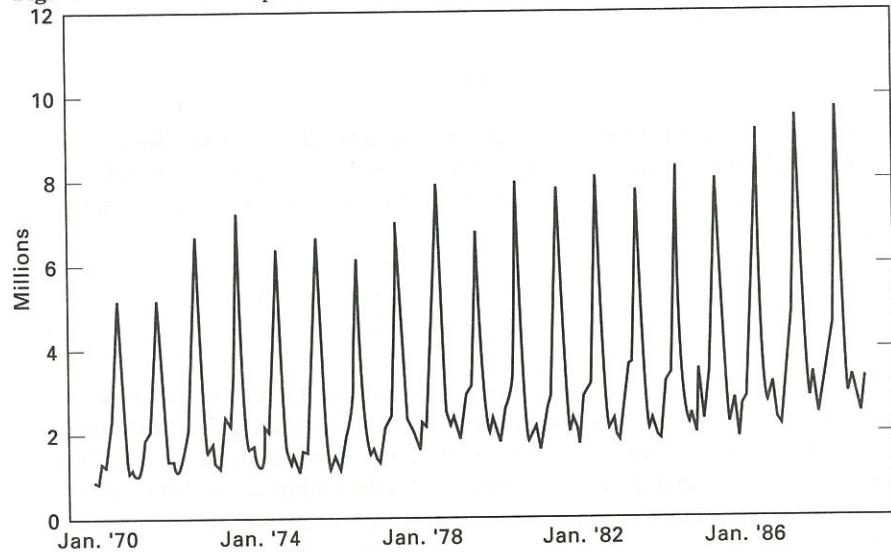
Equation (2.72) differs from (2.70) in that it allows the moving average term at lag 1 to interact with the seasonal moving average effect at lag 4. In the same way, (2.73) allows the autoregressive term at lag 1 to interact with the seasonal autoregressive effect at lag 4. Many researchers prefer the multiplicative form since a rich interaction pattern can be captured with a small number of coefficients. Rewrite (2.72) as

$$y_t = a_1 y_{t-1} + \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_4 \epsilon_{t-4} + \beta_1 \beta_4 \epsilon_{t-5} \quad (2.74)$$

Estimating only three coefficients (i.e., a_1 , β_1 , and β_4) allows us to capture the effects of an autoregressive term at lag 1 and the effects of moving average terms at lags 1, 4, and 5. Of course, you do not really get something for nothing. The estimates of the three moving average coefficients are interrelated. A researcher estimating the unconstrained model $y_t = a_1 y_{t-1} + \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_4 \epsilon_{t-4} + \beta_5 \epsilon_{t-5}$ would necessarily obtain a smaller residual sum of squares, since β_5 is not constrained to equal $\beta_1 \beta_4$. However, (2.72) is clearly the more parsimonious model. If the unconstrained value of β_5 approximates the product $\beta_1 \beta_4$, the multiplicative model will be preferable. For this reason, most software packages have routines capable of estimating multiplicative models. Otherwise, there are no theoretical grounds leading us to prefer one form of seasonality over another. As illustrated in the last section, experimentation and diagnostic checks are probably the best way to obtain the most appropriate model.

Seasonal Differencing

Spain is undoubtedly the most popular destination for European vacationers. During the months of July and August, the beaches along the Mediterranean coast swell with tourists basking in the sun. Figure 2.7 shows the monthly number of tourists visiting Spain between January 1970 and March 1989; the strong seasonal pattern dominates the movement in the series. You will also note that Spain's popularity has been growing; the series appears to be nonstationary in that the mean is increasing over time.

Figure 2.7 Tourism in Spain.

This combination of strong seasonality and nonstationarity is often found in economic data. The ACF for a nonstationary seasonal process is similar to that for a nonstationary nonseasonal process; with seasonal data the spikes at lags s , $2s$, $3s$, ... do not exhibit rapid decay. The other autocorrelations are dwarfed by the seasonal effects. Notice ACF for the Spanish tourism data shown in Figure 2.8. The autocorrelation coefficients at lags 12, 24, 36, and 48 are all close to unity and the seasonal peaks decay slowly. The coefficients at lags 6, 18, 30, and 42 are all negative since tourism is always low 6 months from the summer boom.

Let y_t denote the log of number of tourists visiting Spain each month; the first step in the Box-Jenkins method is to difference the $\{y_t\}$ sequence so as to make it stationary. In contrast to the other series we examined, the appropriate way to difference strongly seasonal data is at the seasonal period. Formal tests for seasonal differencing are examined in Chapter 4. For now, it is sufficient to note that the seasonal difference $(1 - L^{12})y_t = y_t - y_{t-12}$ will have a smaller variance than the first difference $y_t - y_{t-1}$. In the Spanish data, the strong seasonality means that January-to-January and July-to-July changes are not as pronounced as the changes between June and July. Figure 2.9 shows the first and twelfth differences of the data; clearly, the twelfth difference has less variation and should be easier to identify and estimate.

The logarithmic twelfth difference (i.e., $y_t - y_{t-12}$) displays a flat ACF showing little tendency to decay. The first 12 of the autocorrelations are

ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8	ρ_9	ρ_{10}	ρ_{11}	ρ_{12}
0.26	0.31	0.26	0.28	0.23	0.24	0.19	0.21	0.19	0.20	0.15	-0.17

There is no reasonable way to fit a low-order model to the seasonally differenced data; the seasonal differencing did not eliminate the time-varying mean. In order to impart stationarity into the series, the next step is to take the first difference of the already seasonally differenced data. The ACF and PACF for the series $(1 - L)(1 - L^{12})y_t$ are shown in Figure 2.10; the properties of this series are much more amenable to the Box-Jenkins methodology. For the first 10 coefficients, the single spike in the ACF and uniform decay of the PACF suggest an MA(1) model. The significant coefficients at lags 11, 12, and 13 might result from additive or multiplicative seasonal factors. The estimates of the following three models are reported in Table 2.5:

$$(1 - L^{12})(1 - L)(1 - a_{12}L^{12})y_t = (1 + \beta_1L)\epsilon_t \quad \text{Model 1: Autoregressive}$$

$$(1 - L^{12})(1 - L)y_t = (1 + \beta_1L)(1 + \beta_{12}L^{12})\epsilon_t \quad \text{Model 2: Multiplicative moving average}$$

$$(1 - L^{12})(1 - L)y_t = (1 + \beta_1L + \beta_{12}L^{12})\epsilon_t \quad \text{Model 3: Additive moving average}$$

The point estimates of the coefficients all imply stationarity and invertibility. Moreover, all are at least six standard deviations from zero. However, the diagnostic statistics all suggest that model 2 is preferred. Model 2 has the best fit in that it has the lowest sum of squared residuals (SSR). Moreover, the Q -statistics for lags

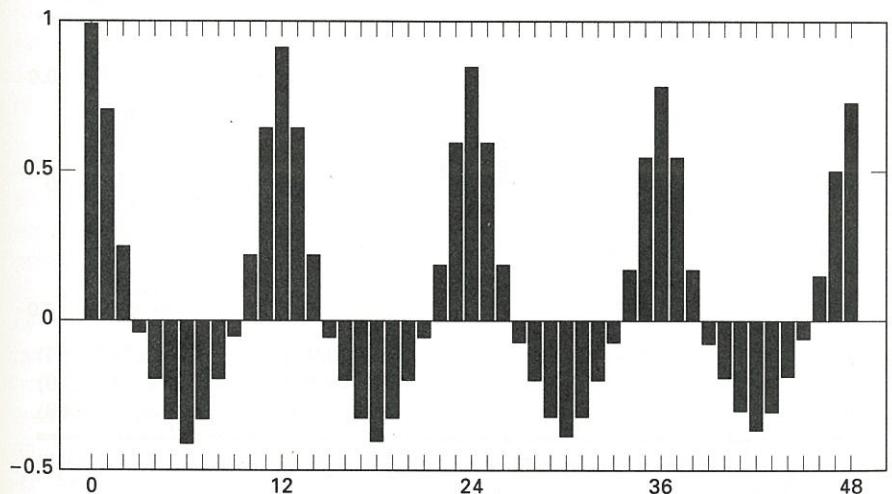
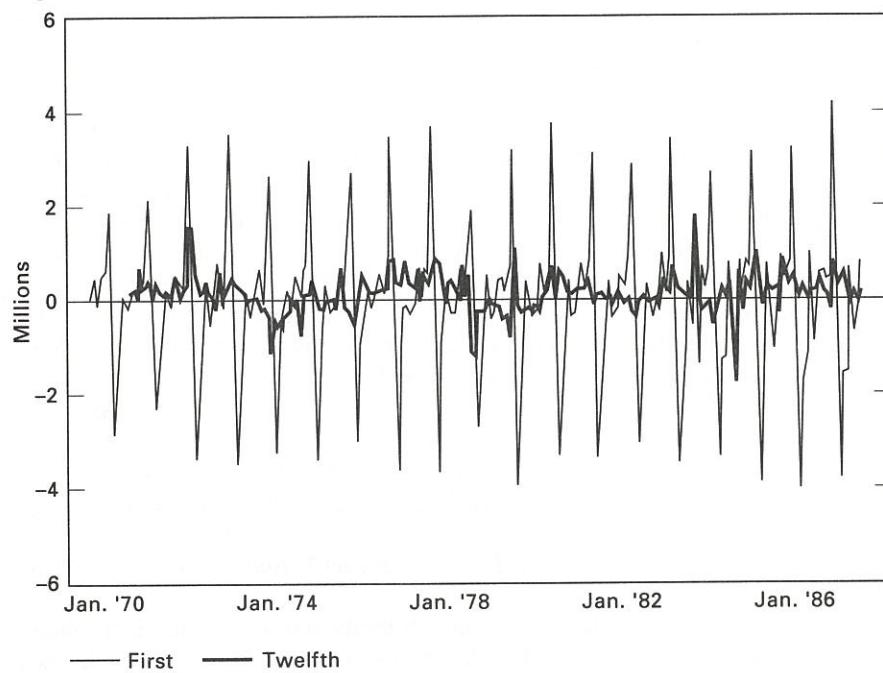
Figure 2.8 Correlogram of tourism in Spain.

Figure 2.9 First and twelfth differences.**Table 2.5:** Three Models of Spanish Tourism

	Model 1 ¹	Model 2	Model 3
a_{12}	-0.408 (-6.54)		
β_1	-0.738 (-15.56)	0.740 (-16.14)	-0.640 (-14.75)
β_{12}		-0.671 (-13.02)	-0.306 (-7.00)
SSR	2.823	2.608	3.367
AIC	217.8	212.98	268.70
SBC	224.5	219.75	275.47
$Q(12)$	8.59 (0.571)	4.38 (0.928)	25.54 (0.004)
$Q(24)$	41.11 (0.007)	15.71 (0.830)	66.58 (0.000)
$Q(48)$	67.91 (0.019)	37.61 (0.806)	99.31 (0.000)

Clearly, there is no difference between an additive seasonality and multiplicative seasonality when all other autoregressive coefficients are zero.

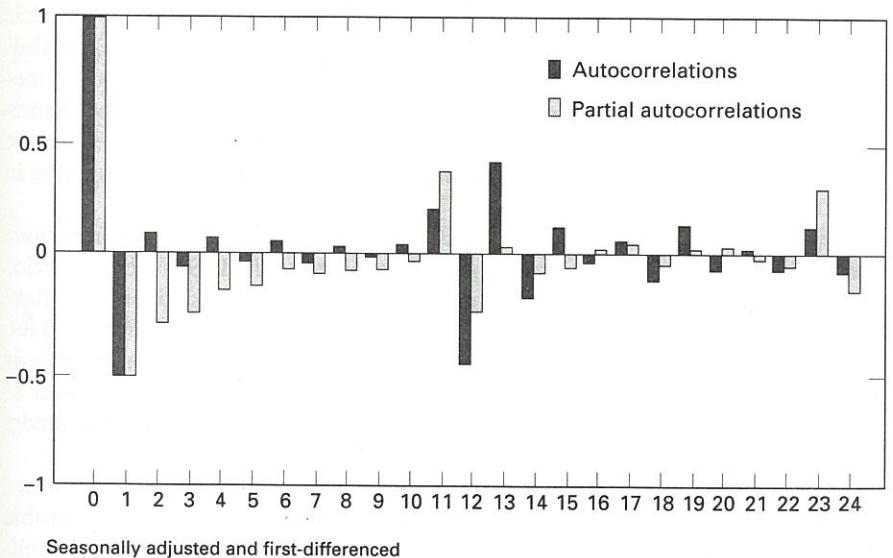
12, 24, and 48 indicate that the residual autocorrelations are insignificant. In contrast, the residual correlations for model 1 are significant at long lags [i.e., $Q(24)$ and $Q(48)$ are significant at the 0.007 and 0.019 levels] and the residual correlations for model 3 are significant for lags 12, 24, and 48. Other diagnostic methods including overfitting and splitting the sample suggest that model 2 is appropriate.

The procedures illustrated in this example of fitting a model to highly seasonal data are typical of many other series. With highly seasonal data, it is necessary to supplement the Box-Jenkins method:

1. In the identification stage, it is necessary to seasonally difference the data and check the ACF of the resultant series. Often, the seasonally differenced data will not be stationary. In such instances, the data may also need to be first-differenced.
2. Use the ACF and PACF to identify potential models. Try to estimate models with low-order nonseasonal ARMA coefficients. Consider both additive and multiplicative seasonality. Allow the appropriate form of seasonality to be determined by the various diagnostic statistics.

A compact notation has been developed that allows for the efficient representation of intricate models. As in previous sections, the d th difference of a series is denoted by Δ^d . For example,

$$\begin{aligned}\Delta^2 y_t &= \Delta(y_t - y_{t-1}) \\ &= y_t - 2y_{t-1} + y_{t-2}\end{aligned}$$

Figure 2.10 ACF and PACF for Spanish Tourism.

A seasonal difference is denoted by Δ_s , where s is the period of the data. The D th such seasonal difference is Δ_s^D . For example, if we wanted the second seasonal difference of the Spanish data, we could form

$$\begin{aligned}\Delta_{12}^2 y_t &= \Delta_{12}(y_t - y_{t-12}) \\ &= \Delta_{12}y_t - \Delta_{12}y_{t-12} \\ &= y_t - y_{t-12} - (y_{t-12} - y_{t-24}) \\ &= y_t - 2y_{t-12} + y_{t-24}\end{aligned}$$

Combining the two types of differencing yields $\Delta^d \Delta_s^D$. Multiplicative models are written in the form ARIMA(p, d, q)(P, D, Q) _{s} .

where p and q = the nonseasonal ARMA coefficients

d = number of nonseasonal differences

P = number of multiplicative autoregressive coefficients;

D = number of seasonal differences

Q = number of multiplicative moving average coefficients

s = seasonal period

Using this notation, we can say that the fitted model of Spanish tourism is an ARIMA(0, 1, 1)(0, 1, 1)₁₂ model. In applied work, the ARIMA(0, 1, 1)(0, 1, 1) _{s} model occurs routinely; it is called the “airline model” ever since Box and Jenkins (1976) used this model to analyze airline travel data.

SUMMARY AND CONCLUSIONS

The chapter focuses on the Box–Jenkins (1976) approach to identification, estimation, diagnostic checking, and forecasting a univariate time series. ARMA models can be viewed as a special class of linear stochastic difference equations. By definition, an ARMA model is covariance stationary in that it has a finite and time-invariant mean and covariances. For an ARMA model to be stationary, the characteristic roots of the difference equation must lie inside the unit circle. Moreover, the process must have started infinitely far in the past or the process must always be in equilibrium.

In the identification stage, the series is plotted and the sample autocorrelations and partial correlations are examined. As illustrated using the U.S. Wholesale Price Index, a slowly decaying autocorrelation function suggests nonstationarity behavior. In such circumstances, Box and Jenkins recommend differencing the data. Formal tests for nonstationarity are presented in Chapter 4. A common practice is to use a logarithmic or Box–Cox transformation if the variance does not appear to be constant. Chapter 3 presents some modern techniques that can be used to model the variance.

The sample autocorrelations and partial correlations of the suitably transformed data are compared to those of various theoretical ARMA processes. All plausible models are estimated and compared using a battery of diagnostic criteria. A well-

estimated model (1) is parsimonious; (2) has coefficients that imply stationarity and invertibility; (3) fits the data well; (4) has residuals that approximate a white-noise process; (5) has coefficients that do not change over the sample period; and (6) has good out-of-sample forecasts.

In utilizing the Box–Jenkins methodology, you will find yourself making many seemingly ad hoc choices. The most parsimonious model may not have the best fit or out-of-sample forecasts. You will find yourself addressing the following types of questions: What is the most appropriate data transformation? Is an ARMA(2, 1) model more appropriate than an ARMA(1, 2) specification? How to best model seasonality? Given this latitude, many view the Box–Jenkins methodology as an art rather than a science. Nevertheless, the technique is best learned through experience. The exercises at the end of this chapter are designed to guide you through the types of choices you will encounter in your own research.

QUESTIONS AND EXERCISES

1. In the coin-tossing example of Section 1, your winnings on the last four tosses (w_t) can be denoted by

$$w_t = 1/4\epsilon_t + 1/4\epsilon_{t-1} + 1/4\epsilon_{t-2} + 1/4\epsilon_{t-3}$$

A. Find the expected value of w_t . Find the expected value given that $\epsilon_{t-3} = \epsilon_{t-2} = 1$.

B. Find $\text{var}(w_t)$. Find $\text{var}(w_t)$ conditional on $\epsilon_{t-3} = \epsilon_{t-2} = 1$.

C. Find: i. $\text{Cov}(w_t, w_{t-1})$ ii. $\text{Cov}(w_t, w_{t-2})$ iii. $\text{Cov}(w_t, w_{t-5})$

2. Substitute (2.10) into $y_t = a_0 + a_1 y_{t-1} + \epsilon_t$. Show that the resulting equation is an identity.

A. Find the homogeneous solution to $y_t = a_0 + a_1 y_{t-1} + \epsilon_t$.

B. Find the particular solution given that $|a_1| < 1$.

C. Show how to obtain (2.10) by combining the homogeneous and particular solutions.

3. Consider the second-order autoregressive process $y_t = a_0 + a_2 y_{t-2} + \epsilon_t$, where $|a_2| < 1$.

A. Find: i. $E_{t-2}y_t$ ii. $E_{t-1}y_t$ iii. $E_t y_{t+2}$

iv. $\text{Cov}(y_t, y_{t-1})$ v. $\text{Cov}(y_t, y_{t-2})$ vi. The partial autocorrelations ϕ_{11} and ϕ_{22}

- B. Find the impulse response function. Given y_{t-2} , trace out the effects on an ϵ_t shock on the $\{y_t\}$ sequence.

- C. Determine the forecast function E_y_{t+s} . The forecast error f_s is the difference between y_{t+s} and E_y_{t+s} . Derive the correlogram of the $\{f_s\}$ sequence. [Hint: Find E_{tf_s} , $\text{var}(f_s)$, and $E(f_s f_{s-j})$ for $j = 0$ to s .]
4. Two different balls are drawn from a jar containing three balls numbered 1, 2, and 4. Let $x =$ number on the first ball drawn and $y =$ sum of the two balls drawn.
- Find the joint probability distribution for x and y ; that is, find $\text{prob}(x = 1, y = 3)$, $\text{prob}(x = 1, y = 5), \dots$, and $\text{prob}(x = 4, y = 6)$.
 - Find each of the following: $E(x)$, $E(y)$, $E(y|x = 1)$, $E(x|y = 5)$, $\text{var}(x|y = 5)$, and $E(y^2)$.
 - Consider the two functions $w_1 = 3x^2$ and $w^2 = x^{-1}$. Find $E(w_1 + w_2)$ and $E(w_1 + w_2|y = 3)$.
 - How would your answers change if the balls were drawn with replacement?
5. The general solution to an n th-order difference equation requires n arbitrary constants. Consider the second-order equation $y_t = a_0 + 0.75y_{t-1} - 0.125y_{t-2} + \epsilon_t$.
- Find the homogeneous and particular solutions. Discuss the shape of the impulse response function.
 - Find the values of the initial conditions (and A_1 and A_2) that ensure the $\{y_t\}$ sequence is stationary. (Note: A_1 and A_2 are the arbitrary constants in the homogeneous solution.)
 - Given your answer to part B, derive the correlogram for the $\{y_t\}$ sequence.
6. Consider the second-order stochastic difference equation $y_t = 1.5y_{t-1} - 0.5y_{t-2} + \epsilon_t$.
- Find the characteristic roots of the homogeneous equation.
 - Demonstrate that the roots of $1 - 1.5L + 0.5L^2$ are the reciprocals of your answer in part A.
 - Given initial conditions for y_0 and y_1 , find the solution for y_t in terms of the current and past values of the $\{\epsilon_t\}$ sequence. Explain why it is not possible to obtain the backward-looking solution for y_t unless such initial conditions are given.
 - Find the forecast function for y_{t+s} .
 - Find: E_y , Ey_{t+1} , $\text{var}(y_t)$, $\text{var}(y_{t+1})$, and $\text{cov}(y_{t+1}, y_t)$.
7. The file entitled SIM_2.WK1 contains the simulated data sets used in this chapter. The first column contains the 100 values of the simulated AR(1)

process used in Section 7. This first series is entitled Y1. Use this series to perform the following tasks. (Note: Due to differences in data handling and rounding, your answers need only approximate those presented here.)

- A. Plot the sequence against time. Do the data appear to be stationary? Show that the properties of the sequence are

Sample mean	-0.5707418062	Variance	1.939987
Skewness	-0.31011	Significance Level (Sk=0)	0.21239328

- B. Verify that the first 12 coefficients of the ACF and PACF are

ACF:

1:	0.7394472	0.5842742	0.4711050	0.3885974	0.3443779	0.3350913
7:	0.2972263	0.3251532	0.2689484	0.2007989	0.1886648	0.0824283

PACF:

1:	0.7394472	0.0827240	0.0302925	0.0255945	0.0601115	0.0889358
7:	-0.0165339	0.1438633	-0.1002335	-0.0653566	0.0699036	-0.2040202

Ljung–Box Q -statistics: $Q(8) = 177.5774$,
 $Q(16) = 197.8423$, $Q(24) = 201.2825$

- C. Use the data to verify the results given in Table 2.2.
D. Determine whether it is appropriate to include a constant in the AR(1) process. You should obtain the following estimates:

Coefficient	Estimate	Standard Error	t-Statistic	Significance Level
1. CONSTANT	-0.538045291	0.380434146	-1.41429	0.16044514
2. AR{1}	0.756861387	0.067241069	11.25594	0.00000000

- E. Estimate the series as an AR(2) process without an intercept. You should obtain:

Coefficient	Estimate	Standard Error	t-Statistic	Significance Level
1. AR{1}	0.7048671016	0.0993987373	7.09131	0.00000000
2. AR{2}	0.1094585628	0.0986680252	1.10936	0.26998889

Ljung–Box Q -statistics: $Q(8) = 5.1317$, $Q(16) = 15.8647$, $Q(24) = 21.0213$

- F. Estimate the series as an ARMA(1, 1) process without an intercept. You should obtain:

Coefficient	Estimate	Standard Error	t-Statistic	Significance Level
1. AR{1}	0.846376753	0.068533381	12.34985	0.00000000
2. MA{1}	-0.148770547	0.125784398	-1.18274	0.23977273

Verify that the first 12 coefficients of the ACF and PACF of the *residuals* are:

ACF:

1: -0.0069909	-0.0365955	-0.0375520	-0.0749124	-0.0683620	0.0546530
7: -0.0808082	0.1598166	0.0732022	-0.0080406	0.1686742	-0.0484844

PACF:

1: -0.0069909	-0.0366462	-0.0381264	-0.0770739	-0.0733243	0.0460005
7: -0.0923797	0.1542973	0.0630681	0.0027253	0.1917630	-0.0374165

Ljung–Box Q -statistics: $Q(8) = 5.2628$, significance level 0.51057476
 $Q(16) = 15.7449$, significance level 0.32919794
 $Q(24) = 21.0950$, significance level 0.51487365

G. Compare the AIC and SBC values from the models estimated in parts D, E, and F.

8. The second column in file entitled SIM_2.WK1 contains the 100 values of the simulated ARMA(1, 1) process used in Section 7. This series is entitled Y2. Use this series to perform the following tasks. (Note: Due to differences in data handling and rounding, your answers need only approximate those presented here.)

A. Plot the sequence against time. Do the data appear to be stationary? Show that the properties of the sequence are:

Sample mean	0.02254818000	Variance	5.743104
Skewness	-0.06175	Significance level (Sk = 0)	0.80390523

ACF:

1: -0.8343833	0.5965289	-0.4399659	0.3497724	-0.3187446	0.3316348
7: -0.3371782	0.3166057	-0.2761498	0.1789268	-0.0839171	0.0375968

PACF:

1: -0.8343833	-0.3280611	-0.1942907	-0.0145160	-0.1398293	0.0891764
7: 0.0004335	0.0143663	0.0166776	-0.1987829	-0.0462213	-0.0212410

B. Verify the results in Table 2.3.

C. Estimate the process using a pure MA(2) model. You should obtain:

Coefficient	Estimate	Standard Error	t-Statistic	Significance Level
1. MA{1}	-1.152648087	0.087208938	-13.21709	0.00000000
2. MA{2}	0.521919469	0.087336869	5.97594	0.00000004

D. Verify that the first 12 coefficients of the ACF and PACF of the *residuals* are

ACF:

1: -0.1281102	0.2841720	-0.2721070	0.0641308	-0.1690135	0.1591088
7: -0.1711865	0.1009624	-0.2300744	0.0202238	-0.0918914	-0.0507396

PACF:

1: -0.1281102	0.2722277	-0.2314021	-0.0521753	-0.0407344	0.0989550
7: -0.1253922	-0.0203505	-0.1278106	-0.0870339	0.0170745	-0.1709188

Ljung–Box Q -statistics: $Q(8) = 28.4771$, significance level 0.00007638
 $Q(16) = 37.4666$, significance level 0.00062675
 $Q(24) = 38.8424$, significance level 0.01470990

9. The third column in SIM_2.WK1 contains the 100 values of an AR(2) process; this series is entitled Y3. Use this series to perform the following tasks. (Note: Due to differences in data handling and rounding, your answers need only approximate those presented here.)

A. Plot the sequence against time. Verify the ACF and PACF coefficients reported in Section 7. Compare the sample ACF and PACF to those of a theoretical AR(2) process.

B. Estimate the series as an AR(1) process. You should find:

Coefficient	Estimate	Standard Error	t-Statistic	Significance Level
1. AR{1}	0.4676067905	0.0892951880	5.23664	0.00000093

ACF of the Residuals:

1: 0.2226399	-0.3349466	-0.3386407	0.0569540	0.0807033	-0.1656232
7: -0.1358947	0.1490039	0.1810292	-0.0022135	-0.0893884	-0.0245175

PACF of the Residuals:

1: 0.2226399	-0.4045690	-0.1809423	0.0803672	-0.1663664	-0.2353309
7: -0.0327129	0.0578083	-0.0587342	0.0005358	0.0422312	-0.0381843

Ljung–Box Q -statistics: $Q(8) = 36.9968$, significance level 0.00000470
 $Q(16) = 55.8708$, significance level 0.00000127
 $Q(24) = 69.0486$, significance level 0.00000170

C. Why is the AR(1) model inadequate?

D. Could an ARMA(1, 1) process generate the type of sample ACF and PACF found in part A? Estimate the series as an ARMA(1, 1) process. You should obtain:

Coefficient	Estimate	Standard Error	t-Statistic	Significance Level
1. AR{1}	0.1861328174	0.1592235925	1.16900	0.24526729
2. MA{1}	0.5057665581	0.1407905283	3.59233	0.00051680

ACF of the Residuals:

1: 0.0284101 -0.1131579 -0.3143993 0.0716440 0.0162748 -0.1298382
7: -0.1197985 0.1392267 0.1194444 0.0174992 -0.1155456 0.0427301

PACF of the Residuals:

1: 0.0284101 -0.1140571 -0.3118831 0.0757999 -0.0596767 -0.2396433
7: -0.0872039 0.1041284 -0.0272326 -0.0175071 -0.0164607 0.0486076

Ljung–Box Q -Statistics: $Q(8) = 17.7685$, significance level 0.00683766
 $Q(16) = 37.0556$, significance level 0.00072359
 $Q(24) = 44.9569$, significance level 0.00268747

Why is the ARMA(1, 1) model inadequate?

E. Estimate the series as an AR(2) process to verify the results reported in the text. Also show that

ACF of the Residuals:

1: 0.0050856 0.0167033 -0.1311013 0.0737802 -0.0183142 -0.1857531
7: -0.1223167 0.1169804 0.0827464 -0.0445903 -0.1014803 0.0879798
13: -0.1499004 0.0365971 -0.1062701 0.2608459 -0.0365855 -0.1119749
19: -0.0855518 0.0179101 0.0695385 -0.1661957 -0.0183144 0.0479631

PACF of the Residuals:

1: 0.0050856 0.0166779 -0.1313096 0.0764420 -0.0160463 -0.2098313
7: -0.1023138 0.1265615 0.0378627 -0.0653412 -0.0679885 0.0629571
13: -0.2287224 0.0563135 -0.0068239 0.2076758 -0.0936362 -0.1587757
19: -0.0419646 -0.0410407 0.0716762 -0.1014686 0.0384143 -0.0779761

Ljung–Box Q -Statistics: $Q(8) = 9.2697$, significance level 0.15896993
 $Q(16) = 24.6248$, significance level 0.03845761
 $Q(24) = 31.8487$, significance level 0.08001287

The Q -statistics indicate that the autocorrelations at longer lags are statistically different from zero at the usual significance levels. Why might you choose *not* to model such long lags when using actual economic data?

F. Now estimate the series as an AR(2) but also include a moving average term at lag 16. Show that the residuals are such that

ACF of the Residuals:

1: 0.0265736 0.0040771 -0.0933018 0.0858766 0.0225622 -0.1521287
7: -0.1643954 0.0947202 0.1447444 0.0017055 -0.0718022 0.0512581
13: -0.1023376 0.0151149 -0.1029252 0.0174225 -0.0629532 -0.1078434
19: -0.0754905 -0.0307818 0.0130560 -0.1275938 0.0223896 0.0338157

PACF of the Residuals:

1: 0.0265736 0.0033733 -0.0935665 0.0917077 0.0182999 -0.1663372
7: -0.1432380 0.1106009 0.1204167 -0.0169905 -0.0350092 0.0517180
13: -0.1887574 0.0078523 0.0014991 0.0232808 -0.0985569 -0.1417484
19: -0.0753388 -0.0797882 0.0086627 -0.1045587 0.0291697 -0.0227024

Ljung–Box Q -statistics: $Q(8) = 8.2222$, significance level 0.14440657
 $Q(16) = 13.9801$, significance level 0.37524746
 $Q(24) = 19.0856$, significance level 0.57964913

C. Compare the AIC and SBC values from the models estimated in parts B, D, E, and F.

10. The file called WPI.WK1 contains the U.S. Wholesale Price Index from 1960:Q1 to 1992:Q2. Make the data transformations indicated in the text.

- Use the sample from 1960:Q1 to 1990:Q4 in order to reproduce the results of Section 10.
- Use the fitted model to create “out-of-sample” forecasts for the 1991:Q1 to 1992:Q2 period.
- Consider some of the plausible alternative models suggested in the text.
 - Try to fit a model to the second-difference of the logarithm of the WPI.
 - Estimate the multiplicative seasonal model
- Compare these models to that of part B.

11. The file entitled US.WK1 contains quarterly value of the U.S. money supply ($M1$) from 60:Q1 to 91:Q4.

A. Plot the sequence against time. Verify that the properties of the sequence are

Sample mean	3.80169890625	Variance	5.260577E+22
Skewness	0.83949	Significance level (Sk = 0)	0.00012712

B. Detrend the data by estimating the regression:

$$\Delta \log(M1) = a_0 + b(\text{time}) + \epsilon_t$$

The ACF of the residuals is

1:	0.8835022	0.8752123	0.8064355	0.8334758	0.7165115	0.6968231
7:	0.6249026	0.6437679	0.5285896	0.5118881	0.4507793	0.4770092

Ljung–Box Q -statistics: $Q(8) = 630.0809$, significance level 0.000
 $Q(16) = 836.4612$, significance level 0.000

Does detrending seem to render the sequence stationary?

C. Calculate the ACF and PACF of the first difference of $\log(M1)$. You should obtain:

ACF:

1:	0.5394848	0.3234781	-0.5573607	0.8528067	-0.5168406	0.2986240
7:	-0.5523817	0.7950047	-0.5096188	0.2695013	-0.5425407	0.7549618

PACF:

1:	-0.5394848	0.0457493	-0.5175494	0.7167389	-0.0356317	-0.1396979
7:	-0.0457462	0.1998479	-0.0995162	-0.1475262	-0.0125845	0.0905883

Explain the observed pattern at lags 4, 8, and 12.

D. Seasonally difference the money supply as $\Delta_4 \log(M1) = \Delta \log(M1) - \Delta \log(M1)_{t-4}$. You should find that the ACF and PACF are

ACF:

1:	0.8585325	0.7148654	0.5452426	0.3963377	0.3401345	0.2636718
7:	0.1814409	0.0991204	0.0554050	0.0287039	0.0423198	0.0651970

PACF:

1:	0.8585325	-0.0844838	-0.1831526	-0.0283342	0.2688532	-0.1594976
7:	-0.1789985	-0.0055668	0.2312324	-0.0787959	-0.0015501	0.0736405

- E. For convenience, let ml_t denote $\Delta_4 \log(M1)$. Estimate the seasonally differenced log of the money supply as the AR(1) process:

$$ml_t = a_0 + a_1 ml_{t-1} + \epsilon_t$$

Coefficient	Estimate	Standard Error	t-Statistic	Significance Level
CONSTANT	0.06217	0.0090502490	6.86967	0.0000000
AR[1]	0.86241	0.0446622831	19.30970	0.0000000

Examine the diagnostic statistics to show that this model is inappropriate.

F. Estimate $\log(M1)$ using each of the following:

$$\begin{aligned} &\text{ARIMA}(1, 0, 0)(0, 1, 1) \\ &\text{ARIMA}[1, 0, (4)](0, 1, 0) \end{aligned}$$

Why is each inadequate?

G. Define $\Delta ml_t = ml_t - ml_{t-1}$ so that Δml_t is the first difference of the seasonal difference of the money supply. Estimate Δml_t as

$$\Delta ml_t = (1 + \beta_4 L^4)\epsilon_t$$

You should obtain:

Coefficient	Estimate	Standard Error	t-Statistic	Significance Level
MA[4]	-0.672328387	0.071121156	-9.45328	0.000000000

ACF of Residuals:

1:	0.0616653	0.1387445	-0.0388472	0.0720538	0.0875724	0.0110692
7:	-0.0622441	-0.0953258	-0.0131446	-0.1265891	-0.0802878	-0.0407282

PACF of Residuals:

1:	0.0616653	0.1354570	-0.0558297	0.0601665	0.0952727	-0.0207820
7:	-0.0826424	-0.0831404	0.0052625	-0.1232642	-0.0717116	0.0263945

Ljung–Box Q -statistics: $Q(8) = 6.5331$, significance level 0.479
 $Q(16) = 10.3813$, significance level 0.795
 $Q(24) = 14.0666$, significance level 0.925
 $Q(32) = 17.4491$, significance level 0.976

Explain why this model is superior to any of those in part F.

ENDNOTES

1. The appendix to this chapter provides a review of constructing joint probabilities, expected values, and variances.
2. Some authors let T equal the maximum number of observations that can be used in the estimation; hence, T changes with the number of parameters estimated. Since there is no underlying distributional theory associated with the AIC and SBC, this procedure cannot be said to be incorrect. Also be aware that there are several equivalent formulations of the AIC and SBC. Your software package may not yield the precise numbers reported in the text.
3. Nearly all econometric software packages contain a Box-Jenkins estimation procedure. Mechanics of the estimation usually entail nothing more than specifying the number of autoregressive and moving average coefficients to include in the estimated model.
4. Most software programs will not be able to estimate (2.43) since there is not a unique set of parameter values that minimizes the likelihood function.
5. Some software programs report the Durbin-Watson test statistic as a check for first-order serial correlation. This well-known test statistic is biased toward finding no serial correlation in the presence of lagged dependent variables. Hence, it is usually not used in ARMA models.
6. Estimation of an AR(p) model usually entails a loss of the number of usable observations. Hence, to estimate a sample using T observations, it will be necessary to have $(T+p)$ observations. Also note that the procedure outlined necessitates that the second subsample period incorporate the lagged values $t_m, t_{m-1}, \dots, t_{m-p+1}$.
7. Many of the details concerning optimal forecasts are contained in the appendix to Chapter 3.
8. In essence, the estimated equation is an ARMA(1, 4) model with the coefficients β_2 and β_3 constrained to be equal to zero. In order to distinguish between the two specifications, the notation ARMA[1, (1,4)] is used to indicate that only the moving average terms at lags 1 and 4 are included in the model.
9. The details of the X-11 procedure are not important for our purposes. The SAS statistical package can perform the X-11 procedure. The technical details of the procedure are explained in the Bureau of the Census report (1969).

APPENDIX Expected Values and Variance

1. Expected value of a discrete random variable

A random variable x is defined to be discrete if the range of x is countable. If x is discrete, there is a finite set of numbers x_1, x_2, \dots, x_n such that x takes on values only in that set. Let $f(x_j)$ = the probability that $x = x_j$. The mean or **expected value** of x is defined to be

$$E(x) = \sum_{j=1}^n x_j f(x_j)$$

Note the following:

1. We can let n go to infinity; the notion of a discrete variable is that the set be “denumerable” or a countable infinity. For example, the set of all positive integers is discrete.
2. If $\sum x_j f(x_j)$ does not converge, the mean is said not to exist.
3. $E(x)$ is an “average” of the possible values of x ; in the sum, each possible x_j is weighted by the probability that $x = x_j$, that is,

$$E(x) = w_1 x_1 + w_2 x_2 + \dots + w_n x_n$$

$$\text{where } \sum w_j = 1$$

2. Expected value of a continuous random variable

Now let x be a continuous random variable. Denote the probability that x is in the interval (x_0, x_1) be denoted by $f(x_0 \leq x \leq x_1)$. If the function $f(x)$ is depicted by Figure A2.1, it follows that

$$f(x_0 \leq x \leq x_1) = \int_{x_0}^{x_1} f(x) dx$$

The mean, or expected value, of x is

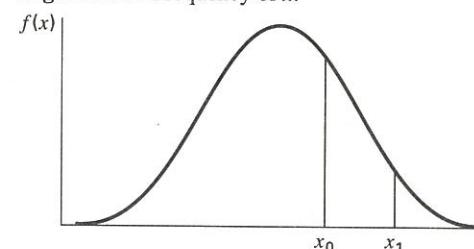
$$E(x) = \int_{-\infty}^{\infty} xf(x) dx$$

3. Expected value of a function

Let x be a random variable and $g(x)$ a function. The mean or expected value of $g(x)$ is

$$E[g(x)] = \sum_{j=1}^n g(x_j) f(x_j)$$

Figure A2.1 Frequency of x .



for discrete x or

$$E[g(x)] = \int_{-\infty}^{\infty} g(x)f(x) dx$$

for continuous x . Note: If $g(x_j) \equiv x_j$, we obtain the simple mean.

4. Properties of the expectations operator

- The expected value of a constant c is the value of the constant. That is, $E(c) = c$.

Proof:

$$g(x) = c \int_{-\infty}^{\infty} cf(x) dx = c \int_{-\infty}^{\infty} f(x) dx = c$$

- The expected value of a constant times a function is the constant times the expected value of the function.

Proof:

$$E[cg(x)] = \int_{-\infty}^{\infty} cg(x)f(x) dx = c \int_{-\infty}^{\infty} g(x)f(x) dx = cE[g(x)]$$

- The expected value of a sum is the sum of the expectations:

$$E[c_1g_1(x) \pm c_2g_2(x)] = c_1Eg_1(x) \pm c_2Eg_2(x)$$

Proof:

$$\begin{aligned} \int_{-\infty}^{\infty} [c_1g_1(x) \pm c_2g_2(x)]f(x) dx &= \int_{-\infty}^{\infty} c_1g_1(x)f(x) dx \pm \int_{-\infty}^{\infty} c_2g_2(x)f(x) dx \\ &= c_1E[g_1(x)] \pm c_2E[g_2(x)] \end{aligned}$$

5. Variance of a Random Variable

The variance of x is defined such that $\text{var}(x) = E\{[x - E(x)]^2\}$:

$$\text{Var}(x) = E[x^2 - 2x E(x) + E(x) E(x)]$$

Since $E(x)$ is a constant, $E[E(x)] = E(x)$ and $E[xE(x)] = [E(x)]^2$. Using these results and the property that the expectation of a sum is the sum of the expectations, we obtain

$$\begin{aligned} \text{Var}(x) &= E(x^2) - 2E[x E(x)] + E(x)^2 \\ &= E(x^2) - [E(x)]^2 \end{aligned}$$

6. Jointly Distributed Discrete Random Variables

Let x and y be random variables such that x takes on values x_1, x_2, \dots, x_n and y takes on values y_1, y_2, \dots, y_m . Also let f_{ij} denote the probability that $x = x_i$ and $y = y_j$. If $g(x, y)$ denotes a function of x and y , the expected value of the function is

$$E[g(x, y)] = \sum_{i=1}^n \sum_{j=1}^m f_{ij}g(x_i, y_j)$$

Expected value of a sum

Let the function $g(x, y)$ be $x + y$. The expected value of $x + y$ is:

$$\begin{aligned} E(x+y) &= \sum_i \sum_j f_{ij}(x_i + y_j) \\ &= \sum_i \sum_j f_{ij}x_i + \sum_i \sum_j f_{ij}y_j \\ &= \sum_j (f_{1j}x_1 + f_{2j}x_2 + \dots + f_{nj}x_n) + \sum_i (f_{i1}y_1 + f_{i2}y_2 + \dots + f_{im}y_m) \end{aligned}$$

Note that $(f_{11} + f_{12} + f_{13} + \dots + f_{1m})$ is the probability that x takes on the value x_1 denoted by f_{11} . More generally, $(f_{11} + f_{12} + f_{13} + \dots + f_{im})$ is the probability that x takes on the value x_i denoted by f_{ii} or $f(x_i)$. Since $(f_{11} + f_{21} + f_{31} + \dots + f_{n1})$ is the probability that $y = y_i$ denoted by $f(y_i)$, the two summations above can be written as

$$\begin{aligned} E(x+y) &= \sum_i x_i f(x_i) + \sum_j y_j f(y_j) \\ &= E(x) + E(y) \end{aligned}$$

Hence, we have generalized the result of 4.3 above to show that the expected value of a sum is the sum of the expectations.

7. Covariance and Correlation

The covariance between x and y , $\text{cov}(x, y)$, is defined to be

$$\text{Cov}(x, y) = E\{[x - E(x)][y - E(y)]\} \equiv \sigma_{xy}$$

Multiply $[x - E(x)]$ by $[y - E(y)]$ and use the property that the expected value of a sum is the sum of the expectations:

$$\begin{aligned}\text{Cov}(x, y) &= E(xy) - E[xE(y)] - E[yE(x)] + E[E(x)E(y)] \\ &= E(xy) - E(x)E(y)\end{aligned}$$

The **correlation coefficient** between x and y is defined to be

$$\rho_{xy} = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x)}\sqrt{\text{var}(y)}}$$

Since $\text{cov}(x, y) = E(xy) - E(x)E(y)$, we can express the expectation of the product of x and y , $E(xy)$, as

$$\begin{aligned}E(xy) &= E(x)E(y) + \text{cov}(x, y) \\ &= E(x)E(y) + \rho_{xy} \sigma_x \sigma_y\end{aligned}$$

where the standard deviation of variable z (denoted by σ_z) = the positive square root of z .

8. Conditional Expectation

Let x and y be jointly distributed random variables, where f_{ij} denotes the probability that $x = x_i$ and $y = y_j$. Each of the f_{ij} values is a **conditional probability**; each is the probability that x takes on the value x_i given that y takes on the specific value y_j .

The expected value of x conditional on y taking on the value y_j is:

$$E(x|y_j) = f_{1j}x_1 + f_{2j}x_2 + \dots + f_{nj}x_n$$

9. Statistical Independence

If x and y are **statistically independent**, the probability of $x = x_i$ and $y = y_j$ is the probability that $x = x_i$ multiplied by the probability that $y = y_j$. If we use the notation in number 6 above, *two events are statistically independent if and only if $f_{ij} = f(x_i)f(y_j)$* . For example, if we simultaneously toss a fair coin and roll a fair die, the probability of obtaining a head and a three is $1/12$; the probability of a head is $1/2$ and the probability of obtaining a three is $1/6$.

An extremely important implication follows directly from this definition. If x and y are independent events, the expected value of the product of the outcomes is the product of the expected outcomes:

$$E(xy) = E(x)E(y)$$

The proof is straightforward. Form $E(xy)$ as

$$\begin{aligned}E(xy) &= f_{11}x_1y_1 + f_{12}x_1y_2 + f_{13}x_1y_3 + \dots + f_{1m}x_1y_m + f_{21}x_2y_1 + f_{22}x_2y_2 \\ &\quad + f_{23}x_2y_3 + \dots + f_{2m}x_2y_m + \dots + f_{n1}x_ny_1 + f_{n2}x_ny_2 + f_{n3}x_ny_3 + \dots + f_{nm}x_ny_m\end{aligned}$$

Since x and y are independent, $f_{ij} = f(x_i)f(y_i)$. Hence

$$\begin{aligned}E(xy) &= \sum_{i=1}^n f_{i1}x_iy_1 + \dots + \sum_{i=1}^n f_{im}x_iy_m \\ &= \sum_{i=1}^n f(x_i)f(y_1)x_iy_1 + \sum_{i=1}^n f(x_i)f(y_2)x_iy_2 + \dots + \sum_{i=1}^n f(x_i)f(y_m)x_iy_m \\ &= f(y_1)y_1 \sum_{i=1}^n f(x_i)x_i + \dots + f(y_m)y_m \sum_{i=1}^n f(x_i)x_i\end{aligned}$$

Recall that $\sum f(x_i)x_i = E(x)$. Thus

$$E(xy) = E(x)[f(y_1)y_1 + f(y_2)y_2 + \dots + f(y_m)y_m]$$

so that $E(xy) = E(x)E(y)$. Since $\text{cov}(x, y) = E(xy) - E(x)E(y)$, it immediately follows that the covariance and correlation coefficient of two independent events is zero.

10. An Example of Conditional Expectation

Since the concept of conditional expectation plays such an important role in modern macroeconomics, it is worthwhile to consider the specific example of tossing dice. Let x denote the number of spots showing on die 1, y the number of spots on die 2, and S the sum of the spots ($S = x + y$). Each die is fair so that the probability of any face turning up is $1/6$. Since the outcomes on die 1 and die 2 are independent events, the probability of any specific values for x and y is the product of the probabilities. The possible outcomes and the probability associated with each outcome S are

S	2	3	4	5	6	7	8	9	10	11	12
$f(S)$	$1/36$	$2/36$	$3/36$	$4/36$	$5/36$	$6/36$	$5/36$	$4/36$	$3/36$	$2/36$	$1/36$

To find the expected value of the sum S , multiply each possible outcome by the probability associated with that outcome. As you well know if you have been to Las Vegas, the expected value is 7. Suppose that you roll the dice sequentially and that the first turns up 3 spots. What is the expected value of the sum given that $x = 3$? We know that y can take on values 1 through 6 each with

a probability of $1/6$. Given $x = 3$, the possible outcomes for S are 4 through 9, each with a probability of $1/6$. Hence, the conditional probability of S given three spots on die 1 is $E(S | x = 3) = (1/6)4 + (1/6)5 + (1/6)6 + (1/6)7 + (1/6)8 + (1/6)9 = 6.5$.