# 15 Models of Nonstationary Time Series

Up to this point our analysis has typically been confined to stationary processes. This chapter introduces several approaches to modeling nonstationary time series and analyzes the dynamic properties of different models of nonstationarity. Consequences of nonstationarity for statistical inference are investigated in subsequent chapters.

## 15.1. *Introduction*

Chapters 3 and 4 discussed univariate time series models that can be written in the form

$$y_t = \mu + \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots = \mu + \psi(L)\varepsilon_t, \qquad [15.1.1]$$

where $\Sigma_{j=0}^{\infty}|\psi_j| < \infty$, roots of $\psi(z) = 0$ are outside the unit circle, and $\{\varepsilon_t\}$ is a white noise sequence with mean zero and variance $\sigma^2$. Two features of such processes merit repeating here. First, the unconditional expectation of the variable is a constant, independent of the date of the observation:

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

Second, as one tries to forecast the series farther into the future, the forecast $\hat{y}_{t+s|t} \equiv \hat{E}(y_{t+s}|y_t, y_{t-1}, \ldots)$ converges to the unconditional mean:

$$\lim_{s\to\infty} \hat{y}_{t+s|t} = \mu.$$

These can be quite unappealing assumptions for many of the economic and financial time series encountered in practice. For example, Figure 15.1 plots the level of nominal gross national product for the United States since World War II. There is no doubt that this series has trended upward over time, and this upward trend should be incorporated in any forecasts of this series.

There are two popular approaches to describing such trends. The first is to include a *deterministic time trend*:

$$y_t = \alpha + \delta t + \psi(L)\varepsilon_t. \qquad [15.1.2]$$

Thus, the mean $\mu$ of the stationary[1] process [15.1.1] is replaced by a linear function of the date $t$. Such a process is sometimes described as *trend-stationary*, because if one subtracts the trend $\alpha + \delta t$ from [15.1.2], the result is a stationary process.

The second specification is a *unit root* process,

$$(1 - L)y_t = \delta + \psi(L)\varepsilon_t, \qquad [15.1.3]$$

[1] Recall that "stationary" is taken to mean "covariance-stationary."

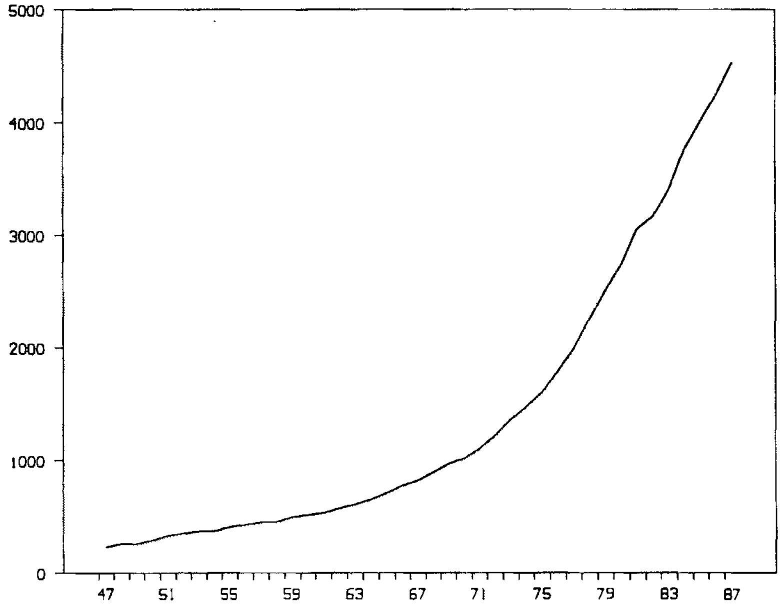


**FIGURE 15.1** U.S. nominal GNP, 1947–87.

where $\psi(1) \neq 0$. For a unit root process, a stationary representation of the form of [15.1.1] describes changes in the series. For reasons that will become clear shortly, the mean of $(1 - L)y_t$ is denoted $\delta$ rather than $\mu$.

The *first-difference* operator $(1 - L)$ will come up sufficiently often that a special symbol (the Greek letter $\Delta$) is reserved for it:

$$\Delta y_t \equiv y_t - y_{t-1}.$$

The prototypical example of a unit root process is obtained by setting $\psi(L)$ equal to 1 in [15.1.3]:

$$y_t = y_{t-1} + \delta + \varepsilon_t. \qquad [15.1.4]$$

This process is known as a *random walk with drift* $\delta$.

In the definition of the unit root process in [15.1.3], it was assumed that $\psi(1)$ is nonzero, where $\psi(1)$ denotes the polynomial

$$\psi(z) = 1 + \psi_1 z^1 + \psi_2 z^2 + \cdots$$

evaluated at $z = 1$. To see why such a restriction must be part of the definition of a unit root process, suppose that the original series $y_t$ is in fact stationary with a representation of the form

$$y_t = \mu + \chi(L)\varepsilon_t.$$

If such a stationary series is differenced, the result is

$$(1 - L)y_t = (1 - L)\chi(L)\varepsilon_t \equiv \psi(L)\varepsilon_t,$$

where $\psi(L) \equiv (1 - L)\chi(L)$. This representation is in the form of [15.1.3]—if the original series $y_t$ is stationary, then so is $\Delta y_t$. However, the moving average operator $\psi(L)$ that characterizes $\Delta y_t$ has the property that $\psi(1) = (1 - 1)\cdot\chi(1) = 0$. When we stipulated that $\psi(1) \neq 0$ in [15.1.3], we were thus ruling out the possibility that the original series $y_t$ is stationary.

It is sometimes convenient to work with a slightly different representation of the unit root process [15.1.3]. Consider the following specification:

$$y_t = \alpha + \delta t + u_t, \tag{15.1.5}$$

where $u_t$ follows a zero-mean *ARMA* process:

$$\begin{aligned}(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)u_t \\ = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)\varepsilon_t\end{aligned} \tag{15.1.6}$$

and where the moving average operator $(1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)$ is invertible. Suppose that the autoregressive operator in [15.1.6] is factored as in equation [2.4.3]:

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) = (1 - \lambda_1 L)(1 - \lambda_2 L) \cdots (1 - \lambda_p L).$$

If all of the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ are inside the unit circle, then [15.1.6] can be expressed as

$$u_t = \frac{1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q}{(1 - \lambda_1 L)(1 - \lambda_2 L) \cdots (1 - \lambda_p L)} \varepsilon_t \equiv \psi(L)\varepsilon_t,$$

with $\Sigma_{j=0}^{\infty}|\psi_j| < \infty$ and roots of $\psi(z) = 0$ outside the unit circle. Thus, when $|\lambda_i| < 1$ for all $i$, the process [15.1.5] would just be a special case of the trend-stationary process of [15.1.2].

Suppose instead that $\lambda_1 = 1$ and $|\lambda_i| < 1$ for $i = 2, 3, \ldots, p$. Then [15.1.6] would state that

$$\begin{aligned}(1 - L)(1 - \lambda_2 L)(1 - \lambda_3 L) \cdots (1 - \lambda_p L)u_t \\ = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)\varepsilon_t,\end{aligned} \tag{15.1.7}$$

implying that

$$(1 - L)u_t = \frac{1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q}{(1 - \lambda_2 L)(1 - \lambda_3 L) \cdots (1 - \lambda_p L)} \varepsilon_t \equiv \psi^*(L)\varepsilon_t,$$

with $\Sigma_{j=0}^{\infty}|\psi_j^*| < \infty$ and roots of $\psi^*(z) = 0$ outside the unit circle. Thus, if [15.1.5] is first-differenced, the result is

$$(1 - L)y_t = (1 - L)\alpha + [\delta t - \delta(t - 1)] + (1 - L)u_t = 0 + \delta + \psi^*(L)\varepsilon_t,$$

which is of the form of the unit root process [15.1.3].

The representation in [15.1.5] explains the use of the term "unit root process." One of the roots or eigenvalues ($\lambda_1$) of the autoregressive polynomial in [15.1.6] is unity, and all other eigenvalues are inside the unit circle.

Another expression that is sometimes used is that the process [15.1.3] is *integrated* of order 1. This is indicated as $y_t \sim I(1)$. The term "integrated" comes from calculus; if $dy/dt = x$, then $y$ is the integral of $x$. In discrete time series, if $\Delta y_t = x_t$, then $y$ might also be viewed as the integral, or sum over $t$, of $x$.

If a process written in the form of [15.1.5] and [15.1.6] has two eigenvalues $\lambda_1$ and $\lambda_2$ that are both equal to unity with the others all inside the unit circle, then second differences of the data have to be taken before arriving at a stationary time series:

$$(1 - L)^2 y_t = \kappa + \psi(L)\varepsilon_t.$$

Such a process is said to be integrated of order 2, denoted $y_t \sim I(2)$.

A general process written in the form of [15.1.5] and [15.1.6] is called an *autoregressive integrated moving average process*, denoted *ARIMA*$(p, d, q)$. The first parameter ($p$) refers to the number of autoregressive lags (not counting the unit roots), the second parameter ($d$) refers to the order of integration, and the third parameter ($q$) gives the number of moving average lags. Taking $d$th differences of an *ARIMA*$(p, d, q)$ produces a stationary *ARMA*$(p, q)$ process.

## 15.2. *Why Linear Time Trends and Unit Roots?*

One might wonder why, for the trend-stationary specification [15.1.2], the trend is specified to be a linear function of time ($\delta t$) rather than a quadratic function ($\delta t + \gamma t^2$) or exponential ($e^{\delta t}$). Indeed, the GNP series in Figure 15.1, like many economic and financial time series, seems better characterized by an exponential trend than a linear trend. An exponential trend exhibits constant proportional growth; that is, if

$$y_t = e^{\delta t}, \qquad [15.2.1]$$

then $dy/dt = \delta \cdot y_t$. Proportional growth in the population would arise if the number of children born were a constant fraction of the current population. Proportional growth in prices (or constant inflation) would arise if the government were trying to collect a constant level of real revenues from printing money. Such stories are often an appealing starting point for thinking about the sources of time trends, and exponential growth is often confirmed by the visual appearance of the series as in Figure 15.1. For this reason, many economists simply assume that growth is of the exponential form.

Notice that if we take the natural log of the exponential trend [15.2.1], the result is a linear trend,

$$\log(y_t) = \delta t.$$

Thus, it is common to take logs of the data before attempting to describe them with the model in [15.1.2].

Similar arguments suggest taking natural logs before applying [15.1.3]. For small changes, the first difference of the log of a variable is approximately the same as the percentage change in the variable:

$$\begin{aligned}(1 - L)\log(y_t) &= \log(y_t/y_{t-1}) \\ &= \log\{1 + [(y_t - y_{t-1})/y_{t-1}]\} \\ &\cong (y_t - y_{t-1})/y_{t-1},\end{aligned}$$

where we have used the fact that for $x$ close to zero, $\log(1 + x) \cong x$.[2] Thus, if the logs of a variable are specified to follow a unit root process, the assumption is that the rate of growth of the series is a stationary stochastic process. The same arguments used to justify taking logs before applying [15.1.2] also suggest taking logs before applying [15.1.3].

Often the units are slightly more convenient if $\log(y_t)$ is multiplied by 100. Then changes are measured directly in units of percentage change. For example, if $(1 - L)[100 \times \log(y_t)] = 1.0$, then $y_t$ is 1% higher than $y_{t-1}$.

## 15.3. *Comparison of Trend-Stationary and Unit Root Processes*

This section compares a trend-stationary process [15.1.2] with a unit root process [15.1.3] in terms of forecasts of the series, variance of the forecast error, dynamic multipliers, and transformations needed to achieve stationarity.

[2]See result [A.3.36] in the Mathematical Review (Appendix A) at the end of the book.

## Comparison of Forecasts

To forecast a trend-stationary process [15.1.2], the known deterministic component $(\alpha + \delta t)$ is simply added to the forecast of the stationary stochastic component:

$$\hat{y}_{t+s|t} = \alpha + \delta(t + s) + \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + \psi_{s+2}\varepsilon_{t-2} + \cdots. \qquad [15.3.1]$$

Here $\hat{y}_{t+s|t}$ denotes the linear projection of $y_{t+s}$ on a constant and $y_t, y_{t-1}, \ldots$. Note that for nonstationary processes, we will follow the convention that the "constant" term in a linear projection, in this case $\alpha + \delta(t + s)$, can be different for each date $t + s$. As the forecast horizon $(s)$ grows large, absolute summability of $\{\psi_j\}$ implies that this forecast converges in mean square to the time trend:

$$E[\hat{y}_{t+s|t} - \alpha - \delta(t + s)]^2 \to 0 \qquad \text{as} \quad s \to \infty.$$

To forecast the unit root process [15.1.3], recall that the change $\Delta y_t$ is a stationary process that can be forecast using the standard formula:

$$\begin{aligned} \Delta\hat{y}_{t+s|t} &\equiv \hat{E}[(y_{t+s} - y_{t+s-1})|y_t, y_{t-1}, \ldots] \\ &= \delta + \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + \psi_{s+2}\varepsilon_{t-2} + \cdots. \end{aligned} \qquad [15.3.2]$$

The level of the variable at date $t + s$ is simply the sum of the changes between $t$ and $t + s$:

$$\begin{aligned} y_{t+s} &= (y_{t+s} - y_{t+s-1}) + (y_{t+s-1} - y_{t+s-2}) + \cdots \\ &\quad + (y_{t+1} - y_t) + y_t \\ &= \Delta y_{t+s} + \Delta y_{t+s-1} + \cdots + \Delta y_{t+1} + y_t. \end{aligned} \qquad [15.3.3]$$

Taking the linear projection of [15.3.3] on a constant and $y_t, y_{t-1}, \ldots$ and substituting from [15.3.2] gives

$$\begin{aligned} \hat{y}_{t+s|t} &= \Delta\hat{y}_{t+s|t} + \Delta\hat{y}_{t+s-1|t} + \cdots + \Delta\hat{y}_{t+1|t} + y_t \\ &= \{\delta + \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + \psi_{s+2}\varepsilon_{t-2} + \cdots\} \\ &\quad + \{\delta + \psi_{s-1}\varepsilon_t + \psi_s\varepsilon_{t-1} + \psi_{s+1}\varepsilon_{t-2} + \cdots\} \\ &\quad + \cdots + \{\delta + \psi_1\varepsilon_t + \psi_2\varepsilon_{t-1} + \psi_3\varepsilon_{t-2} + \cdots\} + y_t \end{aligned}$$

or

$$\begin{aligned} \hat{y}_{t+s|t} &= s\delta + y_t + (\psi_s + \psi_{s-1} + \cdots + \psi_1)\varepsilon_t \\ &\quad + (\psi_{s+1} + \psi_s + \cdots + \psi_2)\varepsilon_{t-1} + \cdots. \end{aligned} \qquad [15.3.4]$$

Further insight into the forecast of a unit root process is obtained by analyzing some special cases. Consider first the random walk with drift [15.1.4], in which $\psi_1 = \psi_2 = \cdots = 0$. Then [15.3.4] becomes

$$\hat{y}_{t+s|t} = s\delta + y_t.$$

A random walk with drift $\delta$ is expected to grow at the constant rate of $\delta$ per period from whatever its current value $y_t$ happens to be.

Consider next an $ARIMA(0, 1, 1)$ specification $(\psi_1 = \theta, \psi_2 = \psi_3 = \cdots = 0)$. Then

$$\hat{y}_{t+s|t} = s\delta + y_t + \theta\varepsilon_t. \qquad [15.3.5]$$

Here, the current level of the series $y_t$ along with the current innovation $\varepsilon_t$ again defines a base from which the variable is expected to grow at the constant rate $\delta$.

Notice that $\varepsilon_t$ is the one-period-ahead forecast error:

$$\varepsilon_t = y_t - \hat{y}_{t|t-1}.$$

It follows from [15.3.5] that for $\delta = 0$ and $s = 1$,

$$\hat{y}_{t+1|t} = y_t + \theta(y_t - \hat{y}_{t|t-1}) \qquad [15.3.6]$$

or

$$\hat{y}_{t+1|t} = (1 + \theta)y_t - \theta\hat{y}_{t|t-1}. \qquad [15.3.7]$$

Equation [15.3.7] takes the form of a simple first-order difference equation, relating $\hat{y}_{t+1|t}$ to its own lagged value and to the input variable $(1 + \theta)y_t$. Provided that $|\theta| < 1$, expression [15.3.7] can be written using result [2.2.9] as

$$\begin{aligned}\hat{y}_{t+1|t} &= [(1 + \theta)y_t] + (-\theta)[(1 + \theta)y_{t-1}] \\ &\quad + (-\theta)^2[(1 + \theta)y_{t-2}] + (-\theta)^3[(1 + \theta)y_{t-3}] + \cdots \\ &= (1 + \theta)\sum_{j=0}^{\infty}(-\theta)^j y_{t-j}.\end{aligned} \qquad [15.3.8]$$

Expression [15.3.7] is sometimes described as *adaptive expectations*, and its implication [15.3.8] is referred to as *exponential smoothing*; typical applications assume that $-1 < \theta < 0$. Letting $y_t$ denote income, Friedman (1957) used exponential smoothing to construct one of his measures of *permanent income.* Muth (1960) noted that adaptive expectations or exponential smoothing corresponds to a rational forecast of future income only if $y_t$ follows an $ARIMA(0, 1, 1)$ process and the smoothing weight $(-\theta)$ is chosen to equal the negative of the moving average coefficient of the differenced data $(\theta)$.

For an $ARIMA(0, 1, q)$ process, the value of $y_t$ and the $q$ most recent values of $\varepsilon_t$ influence the forecasts $\hat{y}_{t+1|t}, \hat{y}_{t+2|t}, \ldots, \hat{y}_{t+q|t}$, but thereafter the series is expected to grow at the rate $\delta$. For an $ARIMA(p, 1, q)$, the forecast growth rate asymptotically approaches $\delta$.

Thus, the parameter $\delta$ in the unit root process [15.1.3] plays a similar role to that of $\delta$ in the deterministic time trend [15.1.2]. With either specification, the forecast $\hat{y}_{t+s|t}$ in [15.3.1] or [15.3.4] converges to a linear function of the forecast horizon $s$ with slope $\delta$; see Figure 15.2. The key difference is in the intercept of the line. For a trend-stationary process, the forecast converges to a line whose intercept is the same regardless of the value of $y_t$. By contrast, the intercept of the limiting forecast for a unit root process is continually changing with each new observation on $y$.

## *Comparison of Forecast Errors*

The trend-stationary and unit root specifications are also very different in their implications for the variance of the forecast error. For the trend-stationary process [15.1.2], the $s$-period-ahead forecast error is

$$\begin{aligned}y_{t+s} - \hat{y}_{t+s|t} &= \{\alpha + \delta(t + s) + \varepsilon_{t+s} + \psi_1\varepsilon_{t+s-1} + \psi_2\varepsilon_{t+s-2} + \cdots \\ &\qquad + \psi_{s-1}\varepsilon_{t+1} + \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + \cdots\} \\ &\qquad - \{\alpha + \delta(t + s) + \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + \psi_{s+2}\varepsilon_{t-2} + \cdots\} \\ &= \varepsilon_{t+s} + \psi_1\varepsilon_{t+s-1} + \psi_2\varepsilon_{t+s-2} + \cdots + \psi_{s-1}\varepsilon_{t+1}.\end{aligned}$$

The mean squared error ($MSE$) of this forecast is

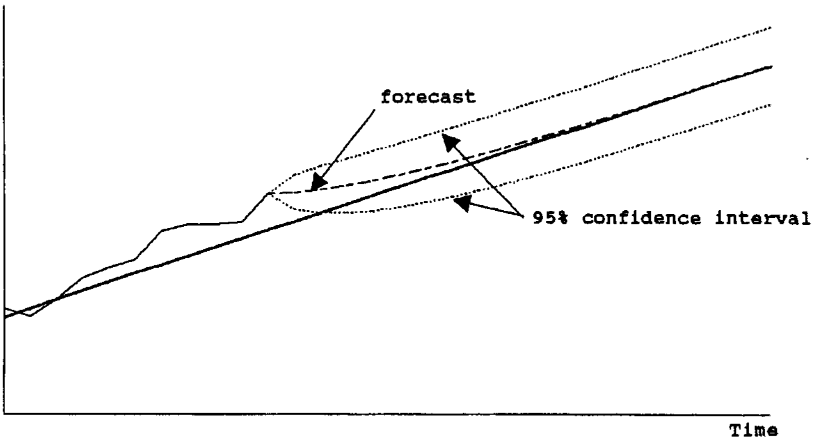
$$E[y_{t+s} - \hat{y}_{t+s|t}]^2 = \{1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{s-1}^2\}\sigma^2.$$

The $MSE$ increases with the forecasting horizon $s$, though as $s$ becomes large, the added uncertainty from forecasting farther into the future becomes negligible:

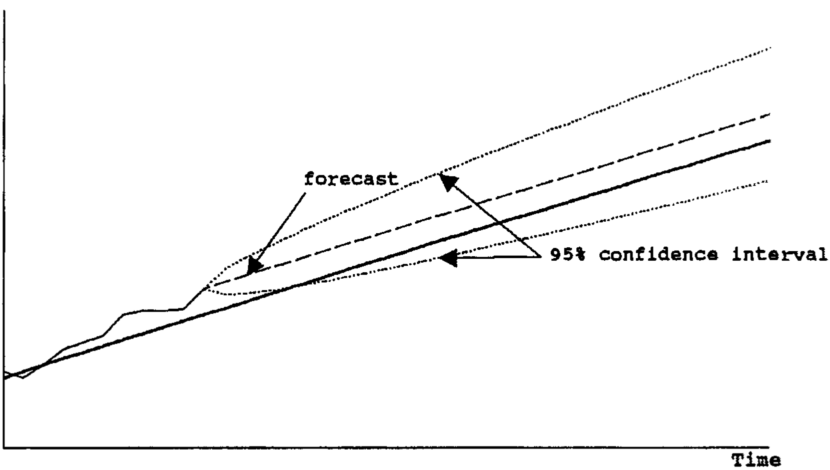
$$\lim_{s \to \infty} E[y_{t+s} - \hat{y}_{t+s|t}]^2 = \{1 + \psi_1^2 + \psi_2^2 + \cdots\}\sigma^2.$$

Note that the limiting $MSE$ is just the unconditional variance of the stationary component $\psi(L)\varepsilon_t$.

By contrast, for the unit root process [15.1.3], the $s$-period-ahead forecast error is




(a) Trend-stationary process




(b) Unit root process

**FIGURE 15.2** Forecasts and 95% confidence intervals.

$$\begin{aligned}
y_{t+s} - \hat{y}_{t+s|t} &= \{\Delta y_{t+s} + \Delta y_{t+s-1} + \cdots + \Delta y_{t+1} + y_t\} \\
&\quad - \{\Delta \hat{y}_{t+s|t} + \Delta \hat{y}_{t+s-1|t} + \cdots + \Delta \hat{y}_{t+1|t} + y_t\} \\
&= \{\varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + \cdots + \psi_{s-1}\varepsilon_{t+1}\} \\
&\quad + \{\varepsilon_{t+s-1} + \psi_1 \varepsilon_{t+s-2} + \cdots + \psi_{s-2}\varepsilon_{t+1}\} + \cdots + \{\varepsilon_{t+1}\} \\
&= \varepsilon_{t+s} + \{1 + \psi_1\}\varepsilon_{t+s-1} + \{1 + \psi_1 + \psi_2\}\varepsilon_{t+s-2} + \cdots \\
&\quad + \{1 + \psi_1 + \psi_2 + \cdots + \psi_{s-1}\}\varepsilon_{t+1},
\end{aligned}$$

with *MSE*

$$\begin{aligned}
E[y_{t+s} - \hat{y}_{t+s|t}]^2 = \{1 &+ (1 + \psi_1)^2 + (1 + \psi_1 + \psi_2)^2 + \cdots \\
&+ (1 + \psi_1 + \psi_2 + \cdots + \psi_{s-1})^2\}\sigma^2.
\end{aligned}$$

The *MSE* again increases with the length of the forecasting horizon $s$, though in contrast to the trend-stationary case, the *MSE* does not converge to any fixed value as $s$ goes to infinity. Instead, it asymptotically approaches a linear function of $s$ with slope $(1 + \psi_1 + \psi_2 + \cdots)^2\sigma^2$. For example, for an $ARIMA(0, 1, 1)$ process,

$$E[y_{t+s} - \hat{y}_{t+s|t}]^2 = \{1 + (s - 1)(1 + \theta)^2\}\sigma^2. \qquad [15.3.9]$$

To summarize, for a trend-stationary process the *MSE* reaches a finite bound as the forecast horizon becomes large, whereas for a unit root process the *MSE* eventually grows linearly with the forecast horizon. This result is again illustrated in Figure 15.2.

Note that since the *MSE* grows linearly with the forecast horizon $s$, the standard deviation of the forecast error grows with the square root of $s$. On the other hand, if $\delta > 0$, then the forecast itself grows linearly in $s$. Thus, a 95% confidence interval for $y_{t+s}$ expands more slowly than the level of the series, meaning that data from a unit root process with positive drift are certain to exhibit an upward trend if observed for a sufficiently long period. In this sense the trend introduced by a nonzero drift $\delta$ asymptotically dominates the increasing variability arising over time due to the unit root component. This result is very important for understanding the asymptotic statistical results to be presented in Chapters 17 and 18.

Figure 15.3 plots realizations of a Gaussian random walk without drift and with drift. The random walk without drift, shown in panel (a), shows no tendency to return to its starting value or any unconditional mean. The random walk with drift, shown in panel (b), shows no tendency to return to a fixed deterministic trend line, though the series is asymptotically dominated by the positive drift term.

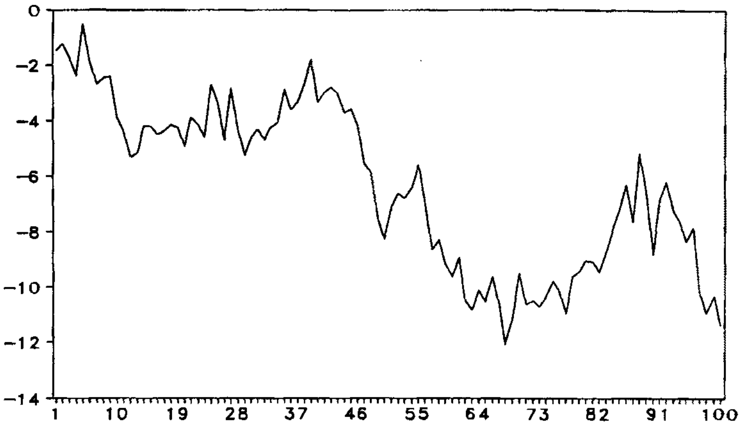
## *Comparison of Dynamic Multipliers*

Another difference between trend-stationary and unit root processes is the persistence of innovations. Consider the consequences for $y_{t+s}$ if $\varepsilon_t$ were to increase by one unit with $\varepsilon$'s for all other dates unaffected. For the trend-stationary process [15.1.2], this dynamic multiplier is given by

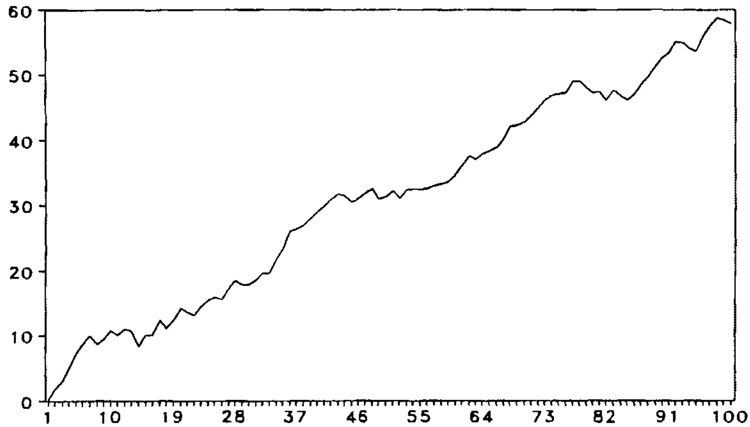
$$\frac{\partial y_{t+s}}{\partial \varepsilon_t} = \psi_s.$$

For a trend-stationary process, then, the effect of any stochastic disturbance eventually wears off:

$$\lim_{s\to\infty} \frac{\partial y_{t+s}}{\partial \varepsilon_t} = 0.$$




(a) Random walk without drift




(b) Random walk with drift

**FIGURE 15.3** Sample realizations from Gaussian unit root processes.

By contrast, for a unit root process, the effect of $\varepsilon_t$ on $y_{t+s}$ is seen from [15.3.4] to be[3]

$$\frac{\partial y_{t+s}}{\partial \varepsilon_t} = \frac{\partial y_t}{\partial \varepsilon_t} + \psi_s + \psi_{s-1} + \cdots + \psi_1 = 1 + \psi_1 + \psi_2 + \cdots + \psi_s.$$

An innovation $\varepsilon_t$ has a permanent effect on the level of $y$ that is captured by

$$\lim_{s\to\infty} \frac{\partial y_{t+s}}{\partial \varepsilon_t} = 1 + \psi_1 + \psi_2 + \cdots = \psi(1). \qquad [15.3.10]$$

[3]This, of course, contrasts with the multiplier that describes the effect of $\varepsilon_t$ on the *change* between $y_{t+s}$ and $y_{t+s-1}$, which is given by

$$\frac{\partial \Delta y_{t+s}}{\partial \varepsilon_t} = \psi_s.$$

As an illustration of calculating such a multiplier, the following $ARIMA(4, 1, 0)$ model was estimated for $y_t$ equal to 100 times the log of quarterly U.S. real GNP ($t$ = 1952:II to 1984:IV):

$$\Delta y_t = 0.555 + 0.312\,\Delta y_{t-1} + 0.122\,\Delta y_{t-2} - 0.116\,\Delta y_{t-3} - 0.081\,\Delta y_{t-4} + \varepsilon_t.$$

For this specification, the permanent effect of a one-unit change in $\varepsilon_t$ on the level of real GNP is estimated to be

$$\psi(1) = 1/\phi(1) = 1/(1 - 0.312 - 0.122 + 0.116 + 0.081) = 1.31.$$

### *Transformations to Achieve Stationarity*

A final difference between trend-stationary and unit root processes that deserves comment is the transformation of the data needed to generate a stationary time series. If the process is really trend stationary as in [15.1.2], the appropriate treatment is to subtract $\delta t$ from $y_t$ to produce a stationary representation of the form of [15.1.1]. By contrast, if the data were really generated by the unit root process [15.1.3], subtracting $\delta t$ from $y_t$ would succeed in removing the time-dependence of the mean but not the variance. For example, if the data were generated by [15.1.4], the random walk with drift, then

$$y_t - \delta t = y_0 + (\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t) \equiv y_0 + u_t.$$

The variance of the residual $u_t$ is $t\sigma^2$; it grows with the date of the observation. Thus, subtracting a time trend from a unit root process is not sufficient to produce a stationary time series.

The correct treatment for a unit root process is to difference the series, and for this reason a process described by [15.1.3] is sometimes called a *difference-stationary* process. Note, however, that if one were to try to difference a trend-stationary process [15.1.2], the result would be

$$\Delta y_t = \delta + (1 - L)\psi(L)\varepsilon_t.$$

This is a stationary time series, but a unit root has been introduced into the moving average representation. Thus, the result would be a noninvertible process subject to the potential difficulties discussed in Chapters 3 through 5.

## 15.4. *The Meaning of Tests for Unit Roots*

Knowing whether nonstationarity in the data is due to a deterministic time trend or a unit root would seem to be a very important question. For example, macroeconomists are very interested in knowing whether economic recessions have permanent consequences for the level of future GNP, or instead represent temporary downturns with the lost output eventually made up during the recovery. Nelson and Plosser (1982) argued that many economic series are better characterized by unit roots than by deterministic time trends. A number of economists have tried to measure the size of the permanent consequences by estimating $\psi(1)$ for various time series representations of GNP growth.[4]

Although it might be very interesting to know whether a time series has a unit root, several recent papers have argued that the question is inherently un-

[4] See, for example, Watson (1986), Clark (1987), Campbell and Mankiw (1987a, b), Cochrane (1988), Gagnon (1988), Stock and Watson (1988), Durlauf (1989), and Hamilton (1989).

answerable on the basis of a finite sample of observations.[5] The argument takes the form of two observations.

The first observation is that for any unit root process there exists a stationary process that will be impossible to distinguish from the unit root representation for any given sample size $T$. Such a stationary process is found easily enough by setting one of the eigenvalues close to but not quite equal to unity. For example, suppose the sample consists of $T = 10{,}000$ observations that were really generated by a driftless random walk:

$$y_t = y_{t-1} + \varepsilon_t \qquad \textit{true model} \text{ (unit root).} \tag{15.4.1}$$

Consider trying to distinguish this from the following stationary process:

$$y_t = \phi y_{t-1} + \varepsilon_t \qquad |\phi| < 1 \qquad \textit{false model} \text{ (stationary).} \tag{15.4.2}$$

The $s$-period-ahead forecast of [15.4.1] is

$$\hat{y}_{t+s|t} = y_t \tag{15.4.3}$$

with *MSE*

$$E(y_{t+s} - \hat{y}_{t+s|t})^2 = s\sigma^2. \tag{15.4.4}$$

The corresponding forecast of [15.4.2] is

$$\hat{y}_{t+s|t} = \phi^s y_t \tag{15.4.5}$$

with *MSE*

$$E(y_{t+s} - \hat{y}_{t+s|t})^2 = (1 + \phi^2 + \phi^4 + \cdots + \phi^{2(s-1)})\cdot\sigma^2. \tag{15.4.6}$$

Clearly there exists a value of $\phi$ sufficiently close to unity such that the observable implications of the stationary representation ([15.4.5] and [15.4.6]) are arbitrarily close to those of the unit root process ([15.4.3] and [15.4.4]) in a sample of size 10,000.

More formally, the conditional likelihood function for a Gaussian process characterized by [15.1.7] is continuous in the parameter $\lambda_1$. Hence, given any fixed sample size $T$, any small numbers $\eta$ and $\varepsilon$, and any unit root specification with $\lambda_1 = 1$, there exists a stationary specification with $\lambda_1 < 1$ with the property that the probability is less than $\varepsilon$ that one would observe a sample of size $T$ for which the value of the likelihood implied by the unit root representation differs by more than $\eta$ from the value of the likelihood implied by the stationary representation.

The converse proposition is also true—for any stationary process and a given sample size $T$, there exists a unit root process that will be impossible to distinguish from the unit root representation. Again, consider a simple example. Suppose the true process is white noise:

$$y_t = \varepsilon_t \qquad \textit{true model} \text{ (stationary).} \tag{15.4.7}$$

Consider trying to distinguish this from

$$(1 - L)y_t = (1 + \theta L)\varepsilon_t \qquad |\theta| < 1 \qquad \textit{false model} \text{ (unit root)} \tag{15.4.8}$$
$$y_0 = \varepsilon_0 = 0.$$

The $s$-period-ahead forecast of [15.4.7] is

$$\hat{y}_{t+s|t} = 0$$

with *MSE*

$$E(y_{t+s} - \hat{y}_{t+s|t})^2 = \sigma^2.$$

[5]See Blough (1992a, b), Cochrane (1991), Christiano and Eichenbaum (1990), Stock (1990), and Sims (1989). The sharpest statement of this view, and the perspective on which the remarks in the text are based, is that of Blough.

The forecast of [15.4.8] is obtained from [15.3.5]:

$$\begin{aligned}\hat{y}_{t+s|t} &= y_t + \theta\varepsilon_t \\ &= \{\Delta y_t + \Delta y_{t-1} + \cdots + \Delta y_2 + y_1\} + \theta\varepsilon_t \\ &= \{(\varepsilon_t + \theta\varepsilon_{t-1}) + (\varepsilon_{t-1} + \theta\varepsilon_{t-2}) + \cdots + (\varepsilon_2 + \theta\varepsilon_1) + (\varepsilon_1)\} + \theta\varepsilon_t \\ &= (1 + \theta)\{\varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_1\}.\end{aligned}$$

From [15.3.9], the *MSE* of the $s$-period-ahead forecast is

$$E(y_{t+s} - \hat{y}_{t+s|t})^2 = \{1 + (s - 1)(1 + \theta)^2\}\sigma^2.$$

Again, clearly, given any fixed sample size $T$, there exists a value of $\theta$ sufficiently close to $-1$ that the unit root process [15.4.8] will have virtually the identical observable implications to those of the stationary process [15.4.7].

Unit root and stationary processes differ in their implications at infinite time horizons, but for any given finite number of observations on the time series, there is a representative from either class of models that could account for all the observed features of the data. We therefore need to be careful with our choice of wording—testing whether a particular time series "contains a unit root," or testing whether innovations "have a permanent effect on the level of the series," however interesting, is simply impossible to do.

Another way to express this is as follows. For a unit root process given by [15.1.3], the autocovariance-generating function of $(1 - L)y_t$ is

$$g_{\Delta Y}(z) = \psi(z)\sigma^2\psi(z^{-1}).$$

The autocovariance-generating function evaluated at $z = 1$ is then

$$g_{\Delta Y}(1) = [\psi(1)]^2\sigma^2. \qquad [15.4.9]$$

Recalling that the population spectrum of $\Delta y$ at frequency $\omega$ is defined by

$$s_{\Delta Y}(\omega) = \frac{1}{2\pi}\, g_{\Delta Y}(e^{-i\omega}),$$

expression [15.4.9] can alternatively be described as $2\pi$ times the spectrum at frequency zero:

$$s_{\Delta Y}(0) = \frac{1}{2\pi}\,[\psi(1)]^2\sigma^2.$$

By contrast, if the true process is the trend-stationary specification [15.1.2], the autocovariance-generating function of $\Delta y$ can be calculated from [3.6.15] as

$$g_{\Delta Y}(z) = (1 - z)\psi(z)\sigma^2\psi(z^{-1})(1 - z^{-1}),$$

which evaluated at $z = 1$ is zero. Thus, if the true process is trend-stationary, the population spectrum of $\Delta y$ at frequency zero is zero, whereas if the true process is characterized by a unit root, the population spectrum of $\Delta y$ at frequency zero is positive.

The question of whether $y_t$ follows a unit root process can thus equivalently be expressed as a question of whether the population spectrum of $\Delta y$ at frequency zero is zero. However, there is no information in a sample of size $T$ about cycles with period greater than $T$, just as there is no information in a sample of size $T$ about the dynamic multiplier for a horizon $s > T$.

These observations notwithstanding, there are several closely related and very interesting questions that are answerable. Given enough data, we certainly can ask whether innovations have a significant effect on the level of the series over a specified finite horizon. For a fixed time horizon (say, $s = 3$ years), there exists a sample size (say, the half century of observations since World War II) such that we can meaningfully inquire whether $\partial y_{t+s}/\partial \varepsilon_t$ is close to zero. We cannot tell whether the data were really generated by [15.4.1] or a close relative of the form of [15.4.2], but we can measure whether innovations have much persistence over a fixed interval (as in [15.4.1] or [15.4.2]) or very little persistence over that interval (as in [15.4.7] or [15.4.8]).

We can also arrive at a testable hypothesis if we are willing to restrict further the class of processes considered. Suppose the dynamics of a given sample $\{y_1, \ldots, y_T\}$ are to be modeled using an autoregression of fixed, known order $p$. For example, suppose we are committed to using an $AR(1)$ process:

$$y_t = \phi y_{t-1} + \varepsilon_t. \qquad [15.4.10]$$

Within this class of models, the restriction

$$H_0\colon \phi = 1$$

is certainly testable. While it is true that there exist local alternatives (such as $\phi = 0.99999$) against which a test would have essentially no power, this is true of most hypothesis tests. There are also alternatives (such as $\phi = 0.3$) that would lead to certain rejection of $H_0$ given enough observations. The hypothesis "$\{y_t\}$ is an $AR(1)$ process with a unit root" is potentially refutable; the hypothesis "$\{y_t\}$ is a general unit root process of the form [15.1.3]" is not.

There may be good reasons to restrict ourselves to consider only low-order autoregressive representations. Parsimonious models often perform best, and autoregressions are much easier to estimate and forecast than moving average processes, particularly moving average processes with a root near unity.

If we are indeed committed to describing the data with a low-order autoregression, knowing whether the further restriction of a unit root should be imposed can clearly be important for two reasons. The first involves a familiar trade-off between efficiency and consistency. If a restriction (in this case, a unit root) is true, more efficient estimates result from imposing it. Estimates of the other coefficients and dynamic multipliers will be more accurate, and forecasts will be better. If the restriction is false, the estimates are unreliable no matter how large the sample. Researchers differ in their advice on how to deal with this trade-off. One practical guide is to estimate the model both with and without the unit root imposed. If the key inferences are similar, so much the better. If the inferences differ, some attempt at explaining the conflicting findings (as in Christiano and Ljungqvist, 1988, or Stock and Watson, 1989) may be desirable.

In addition to the familiar trade-off between efficiency and consistency, the decision whether or not to impose unit roots on an autoregression also raises issues involving the asymptotic distribution theory one uses to test hypotheses about the process. This issue is explored in detail in later chapters.

## 15.5. *Other Approaches to Trended Time Series*

Although most of the analysis of nonstationarity in this book will be devoted to unit roots and time trends, this section briefly discusses two alternative approaches to modeling nonstationarity: fractionally integrated processes and processes with occasional, discrete shifts in the time trend.

## *Fractional Integration*

Recall that an integrated process of order $d$ can be represented in the form

$$(1 - L)^d y_t = \psi(L)\varepsilon_t, \qquad [15.5.1]$$

with $\Sigma_{j=0}^{\infty}|\psi_j| < \infty$. The normal assumption is that $d = 1$, or that the first difference of the series is stationary. Occasionally one finds a series for which $d = 2$ might be a better choice.

Granger and Joyeux (1980) and Hosking (1981) suggested that noninteger values of $d$ in [15.5.1] might also be useful. To understand the meaning of [15.5.1] for noninteger $d$, consider the $MA(\infty)$ representation implied by [15.5.1]. It will be shown shortly that the inverse of the operator $(1 - L)^d$ exists provided that $d < \frac{1}{2}$. Multiplying both sides of [15.5.1] by $(1 - L)^{-d}$ results in

$$y_t = (1 - L)^{-d}\psi(L)\varepsilon_t. \qquad [15.5.2]$$

For $z$ a scalar, define the function

$$f(z) \equiv (1 - z)^{-d}.$$

This function has derivatives given by

$$\begin{aligned}
\frac{\partial f}{\partial z} &= d\cdot(1 - z)^{-d-1} \\
\frac{\partial^2 f}{\partial z^2} &= (d + 1)\cdot d\cdot(1 - z)^{-d-2} \\
\frac{\partial^3 f}{\partial z^3} &= (d + 2)\cdot(d + 1)\cdot d\cdot(1 - z)^{-d-3} \\
&\vdots \\
\frac{\partial^j f}{\partial z^j} &= (d + j - 1)\cdot(d + j - 2)\cdots(d + 1)\cdot d\cdot(1 - z)^{-d-j}.
\end{aligned}$$

A power series expansion for $f(z)$ around $z = 0$ is thus given by

$$\begin{aligned}
(1 - z)^{-d} &= f(0) + \left.\frac{\partial f}{\partial z}\right|_{z=0}\cdot z + \frac{1}{2!}\left.\frac{\partial^2 f}{\partial z^2}\right|_{z=0}\cdot z^2 + \frac{1}{3!}\left.\frac{\partial^3 f}{\partial z^3}\right|_{z=0}\cdot z^3 + \cdots \\
&= 1 + dz + (1/2!)(d + 1)dz^2 + (1/3!)(d + 2)(d + 1)dz^3 + \cdots.
\end{aligned}$$

This suggests that the operator $(1 - L)^{-d}$ might be represented by the filter

$$\begin{aligned}
(1 - L)^{-d} &= 1 + dL + (1/2!)(d + 1)dL^2 \\
&\quad + (1/3!)(d + 2)(d + 1)dL^3 + \cdots \qquad [15.5.3] \\
&= \sum_{j=0}^{\infty} h_j L^j,
\end{aligned}$$

where $h_0 \equiv 1$ and

$$h_j \equiv (1/j!)(d + j - 1)(d + j - 2)(d + j - 3)\cdots(d + 1)(d). \qquad [15.5.4]$$

Appendix 15.A to this chapter establishes that if $d < 1$, $h_j$ can be approximated for large $j$ by

$$h_j \cong (j + 1)^{d-1}. \qquad [15.5.5]$$

Thus, the time series model

$$y_t = (1 - L)^{-d}\varepsilon_t = h_0\varepsilon_t + h_1\varepsilon_{t-1} + h_2\varepsilon_{t-2} + \cdots \qquad [15.5.6]$$

describes an $MA(\infty)$ representation in which the impulse-response coefficient $h_j$ behaves for large $j$ like $(j + 1)^{d-1}$. For comparison, recall that the impulse-response coefficient associated with the $AR(1)$ process $y_t = (1 - \phi L)^{-1}\varepsilon_t$ is given by $\phi^j$. The impulse-response coefficients for a stationary $ARMA$ process decay geometrically, in contrast to the slower decay implied by [15.5.5]. Because of this slower rate of decay, Granger and Joyeux proposed the fractionally integrated process as an approach to modeling long memories in a time series.

In a finite sample, this long memory could be approximated arbitrarily well with a suitably large-order $ARMA$ representation. The goal of the fractional-difference specification is to capture parsimoniously long-run multipliers that decay very slowly.

The sequence of limiting moving average coefficients $\{h_j\}_{j=0}^{\infty}$ given in [15.5.4] can be shown to be square-summable provided that $d < \frac{1}{2}$:[6]

$$\sum_{j=0}^{\infty} h_j^2 < \infty \qquad \text{for } d < \tfrac{1}{2}.$$

Thus, [15.5.6] defines a covariance-stationary process provided that $d < \frac{1}{2}$. If $d > \frac{1}{2}$, the proposal is to difference the process before describing it by [15.5.2]. For example, if $d = 0.7$, the process of [15.5.1] implies

$$(1 - L)^{-0.3}(1 - L)y_t = \psi(L)\varepsilon_t;$$

that is, $\Delta y_t$ is fractionally integrated with parameter $d = -0.3 < \frac{1}{2}$.

Conditions under which fractional integration could arise from aggregation of other processes were described by Granger (1980). Geweke and Porter-Hudak (1983) and Sowell (1992) proposed techniques for estimating $d$. Diebold and Rudebusch (1989) analyzed GNP data and the persistence of business cycle fluctuations using this approach, while Lo (1991) provided an interesting investigation of the persistence of movements in stock prices.

## *Occasional Breaks in Trend*

According to the unit root specification [15.1.3], events are occurring all the time that permanently affect the course of $y$. Perron (1989) and Rappoport and Reichlin (1989) have argued that economic events that have large permanent effects

[6]Reasoning as in Appendix 3.A to Chapter 3.

$$\begin{aligned}\sum_{j=0}^{N-1} (j + 1)^{2(d-1)} &= \sum_{j=1}^{N} j^{2(d-1)} \\ &< 1 + \int_1^N x^{2(d-1)}\,dx \\ &= 1 + [1/(2d - 1)]x^{2d-1}\big|_{x=1}^{N} \\ &= 1 + [1/(2d - 1)]\cdot[N^{2d-1} - 1],\end{aligned}$$

which converges to $1 - [1/(2d - 1)]$ as $N \to \infty$, provided that $d < \frac{1}{2}$.

are relatively rare. The idea can be illustrated with the following model, in which $y_t$ is stationary around a trend with a single break:

$$y_t = \begin{cases} \alpha_1 + \delta t + \varepsilon_t & \text{for } t < T_0 \\ \alpha_2 + \delta t + \varepsilon_t & \text{for } t \geq T_0. \end{cases} \qquad [15.5.7]$$

The finding is that such series would appear to exhibit unit root nonstationarity on the basis of the tests to be discussed in Chapter 17.

Another way of thinking about the process in [15.5.7] is as follows:

$$\Delta y_t = \xi_t + \delta + \varepsilon_t - \varepsilon_{t-1}, \qquad [15.5.8]$$

where $\xi_t = (\alpha_2 - \alpha_1)$ when $t = T_0$ and is zero otherwise. Suppose $\xi_t$ is viewed as a random variable with some probability distribution—say,

$$\xi_t = \begin{cases} \alpha_2 - \alpha_1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p. \end{cases}$$

Evidently, $p$ must be quite small to represent the idea that this is a relatively rare event. Equation [15.5.8] could then be rewritten as

$$\Delta y_t = \mu + \eta_t, \qquad [15.5.9]$$

where

$$\mu = p(\alpha_2 - \alpha_1) + \delta$$
$$\eta_t = \xi_t - p(\alpha_2 - \alpha_1) + \varepsilon_t - \varepsilon_{t-1}.$$

But $\eta_t$ is the sum of a zero-mean white noise process $[\xi_t - p(\alpha_2 - \alpha_1)]$ and an independent $MA(1)$ process $[\varepsilon_t - \varepsilon_{t-1}]$. Therefore, an $MA(1)$ representation for $\eta_t$ exists. From this perspective, [15.5.9] could be viewed as an $ARIMA(0, 1, 1)$ process,

$$\Delta y_t = \mu + v_t + \theta v_{t-1},$$

with a non-Gaussian distribution for the innovations $v_t$:

$$v_t = y_t - \hat{E}(y_t | y_{t-1}, y_{t-2}, \ldots).$$

The optimal linear forecasting rule,

$$\hat{E}(y_{t+s} | y_t, y_{t-1}, \ldots) = \mu s + y_t + \theta v_t,$$

puts a nonvanishing weight on each date's innovation. This weight does not disappear as $s \to \infty$, because each period essentially provides a new observation on the variable $\xi_t$ and the realization of $\xi_t$ has permanent consequences for the level of the series. From this perspective, a time series satisfying [15.5.7] could be described as a unit root process with non-Gaussian innovations.

Lam (1990) estimated a model closely related to [15.5.7] where shifts in the slope of the trend line were assumed to follow a Markov chain and where U.S. real GNP was allowed to follow a stationary third-order autoregression around this trend. Results of his maximum likelihood estimation are reported in Figure 15.4. These findings are very interesting for the question of the long-run consequences of economic recessions. According to this specification, events that permanently changed the level of GNP coincided with the recessions of 1957, 1973, and 1980.

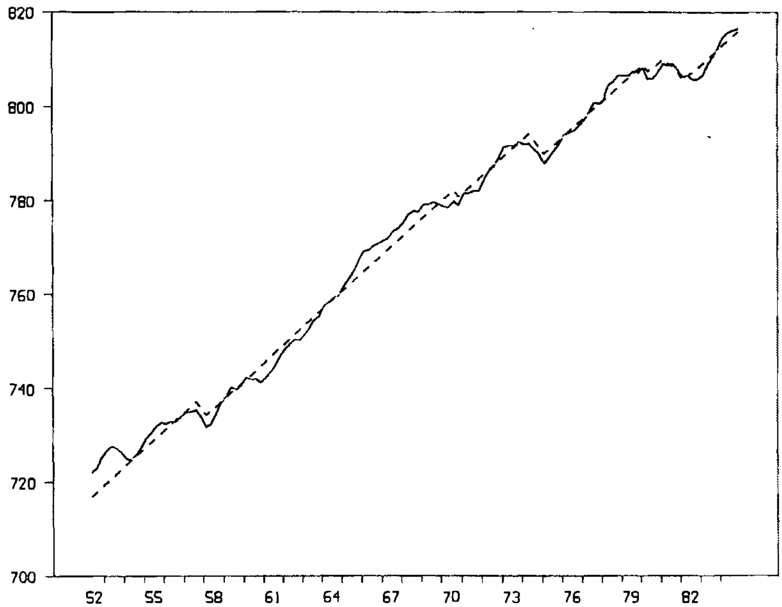



**FIGURE 15.4** Discrete trend shifts estimated for U.S. real GNP, 1952–84 (Lam, 1990).

## APPENDIX 15.A. *Derivation of Selected Equations for Chapter 15*

■ **Derivation of Equation [15.5.5].** Write [15.5.4] as

$$
\begin{aligned}
h_j &\equiv (1/j!)(d+j-1)(d+j-2)(d+j-3)\cdots(d+1)(d) \\
&= \left[\frac{d+j-1}{j}\right]\left[\frac{d+j-2}{j-1}\right]\left[\frac{d+j-3}{j-2}\right]\cdots\left[\frac{d+1}{2}\right]\left[\frac{d}{1}\right] \\
&= \left[\frac{j+d-1}{j}\right]\left[\frac{j-1+d-1}{j-1}\right]\left[\frac{j-2+d-1}{j-2}\right]\times\cdots \\
&\quad\times\left[\frac{j-(j-2)+d-1}{j-(j-2)}\right]\left[\frac{j-(j-1)+d-1}{j-(j-1)}\right] \\
&= \left[1+\frac{d-1}{j}\right]\left[1+\frac{d-1}{j-1}\right]\left[1+\frac{d-1}{j-2}\right]\times\cdots \\
&\quad\times\left[1+\frac{d-1}{j-(j-2)}\right]\left[1+\frac{d-1}{j-(j-1)}\right].
\end{aligned}
\tag{15.A.1}
$$

For large $j$, we have the approximation

$$\left[1+\frac{d-1}{j}\right]\cong\left[1+\frac{1}{j}\right]^{d-1}. \tag{15.A.2}$$

To justify this formally, consider the function $g(x)\equiv(1+x)^{d-1}$. Taylor's theorem states that

$$
\begin{aligned}
(1+x)^{d-1} &= g(0)+\left.\frac{\partial g}{\partial x}\right|_{x=0}\cdot x+\frac{1}{2}\left.\frac{\partial^2 g}{\partial x^2}\right|_{x=\delta}\cdot x^2 \\
&= 1+(d-1)x+\frac{1}{2}(d-1)(d-2)(1+\delta)^{d-3}x^2
\end{aligned}
\tag{15.A.3}
$$

for some $\delta$ between zero and $x$. For $x > -1$ and $d < 1$, equation [15.A.3] implies that

$$(1 + x)^{d-1} \geq 1 + (d - 1)x.$$

Letting $x = 1/j$ gives

$$1 + \frac{d - 1}{j} \leq \left[1 + \frac{1}{j}\right]^{d-1} = \left[\frac{j + 1}{j}\right]^{d-1} \qquad [15.A.4]$$

for all $j > 0$ and $d < 1$, with the approximation [15.A.2] improving as $j \to \infty$. Substituting [15.A.4] into [15.A.1] implies that

$$h_j \cong \left[\frac{j + 1}{j}\right]^{d-1}\left[\frac{j}{j - 1}\right]^{d-1}\left[\frac{j - 1}{j - 2}\right]^{d-1} \cdots \left[\frac{3}{2}\right]^{d-1}\left[\frac{2}{1}\right]^{d-1} = (j + 1)^{d-1}. \qquad [15.A.5]$$

■

## Chapter 15 References


Blough, Stephen R. 1992a. "The Relationship between Power and Level for Generic Unit Root Tests in Finite Samples." *Journal of Applied Econometrics* 7:295–308.

———. 1992b. "Near Observational Equivalence of Unit Root and Stationary Processes: Theory and Implications." Johns Hopkins University. Mimeo.

Box, G. E. P., and Gwilym M. Jenkins. 1976. *Time Series Analysis: Forecasting and Control*, rev. ed. San Francisco: Holden-Day.

Campbell, John Y., and N. Gregory Mankiw. 1987a. "Permanent and Transitory Components in Macroeconomic Fluctuations." *American Economic Review Papers and Proceedings* 77:111–17.

——— and ———. 1987b. "Are Output Fluctuations Transitory?" *Quarterly Journal of Economics* 102:857–80.

Christiano, Lawrence J., and Martin Eichenbaum. 1990. "Unit Roots in Real GNP: Do We Know and Do We Care?" in Allan H. Meltzer, ed., *Unit Roots, Investment Measures, and Other Essays*, 7–61. Carnegie-Rochester Conference Series on Public Policy, Vol. 32. Amsterdam: North-Holland.

——— and Lars Ljungqvist. 1988. "Money Does Granger-Cause Output in the Bivariate Money-Output Relation." *Journal of Monetary Economics* 22:217–35.

Clark, Peter K. 1987. "The Cyclical Component of U.S. Economic Activity." *Quarterly Journal of Economics* 102:797–814.

Cochrane, John H. 1988. "How Big Is the Random Walk in GNP?" *Journal of Political Economy* 96:893–920.

———. 1991. "A Critique of the Application of Unit Root Tests." *Journal of Economic Dynamics and Control* 15:275–84.

Diebold, Francis X., and Glenn D. Rudebusch. 1989. "Long Memory and Persistence in Aggregate Output." *Journal of Monetary Economics* 24:189–209.

Durlauf, Steven N. 1989. "Output Persistence, Economic Structure, and Choice of Stabilization Policy." *Brookings Papers on Economic Activity* 2:1989, 69–116.

Friedman, Milton. 1957. *A Theory of the Consumption Function*. Princeton, N.J.: Princeton University Press.

Gagnon, Joseph E. 1988. "Short-Run Models and Long-Run Forecasts: A Note on the Permanence of Output Fluctuations." *Quarterly Journal of Economics* 103:415–24.

Geweke, John, and Susan Porter-Hudak. 1983. "The Estimation and Application of Long Memory Time Series Models." *Journal of Time Series Analysis* 4:221–38.

Granger, C. W. J. 1980. "Long Memory Relationships and the Aggregation of Dynamic Models." *Journal of Econometrics* 14:227–38.

——— and Roselyne Joyeux. 1980. "An Introduction to Long-Memory Time Series Models and Fractional Differencing." *Journal of Time Series Analysis* 1:15–29.

Hamilton, James D. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57:357–84.

Hosking, J. R. M. 1981. "Fractional Differencing." *Biometrika* 68:165–76.

Lam, Pok-sang. 1990. "The Hamilton Model with a General Autoregressive Component: Estimation and Comparison with Other Models of Economic Time Series." *Journal of Monetary Economics* 26:409–32.

Lo, Andrew W. 1991. "Long-Term Memory in Stock Market Prices." *Econometrica* 59:1279–1313.

Muth, John F. 1960. "Optimal Properties of Exponentially Weighted Forecasts." *Journal of the American Statistical Association* 55:299–306.

Nelson, Charles R., and Charles I. Plosser. 1982. "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications." *Journal of Monetary Economics* 10:139–62.

Perron, Pierre. 1989. "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis." *Econometrica* 57:1361–1401.

Rappoport, Peter, and Lucrezia Reichlin. 1989. "Segmented Trends and Nonstationary Time Series." *Economic Journal* supplement 99:168–77.

Sims, Christopher A. 1989. "Modeling Trends." Yale University. Mimeo.

Sowell, Fallaw. 1992. "Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models." *Journal of Econometrics* 53:165–88.

Stock, James H. 1990. "'Unit Roots in Real GNP: Do We Know and Do We Care?' A Comment," in Allan H. Meltzer, ed., *Unit Roots, Investment Measures, and Other Essays*, 63–82. Carnegie-Rochester Conference Series on Public Policy, Vol. 32. Amsterdam: North-Holland.

——— and Mark W. Watson. 1988. "Variable Trends in Economic Time Series." *Journal of Economic Perspectives* vol. 2, no. 3, 147–74.

——— and ———. 1989. "Interpreting the Evidence on Money-Income Causality." *Journal of Econometrics* 40:161–81.

Watson, Mark W. 1986. "Univariate Detrending Methods with Stochastic Trends." *Journal of Monetary Economics* 18:49–75.