

CHAPTER 1

Fundamental Concepts of Time-Series Econometrics

Many of the principles and properties that we studied in cross-section econometrics carry over when our data are collected over time. However, time-series data present important challenges that are not present with cross sections and that warrant detailed attention.

Random variables that are measured over time are often called “time series.” We define the simplest kind of time series, “white noise,” then we discuss how variables with more complex properties can be derived from an underlying white-noise variable. After studying basic kinds of time-series variables and the rules, or “time-series processes,” that relate them to a white-noise variable, we then make the critical distinction between stationary and non-stationary time-series processes.

1.1 Time Series and White Noise

1.1.1 Time-series processes

A *time series* is a sequence of observations on a variable taken at discrete intervals in time.¹ We index the time periods as 1, 2, ..., T and denote the set of observations as

$$(y_1, y_2, \dots, y_T).$$

We often think of these observations as being a finite sample from a *time-series stochastic process* that began infinitely far back in time and will continue into the indefinite future:

$$\left\{ \overbrace{\dots, y_{-3}, y_{-2}, y_{-1}, y_0}^{\text{pre-sample}}, \overbrace{y_1, y_2, \dots, y_{T-1}, y_T}^{\text{sample}}, \overbrace{y_{T+1}, y_{T+2}, \dots}^{\text{post-sample}} \right\}.$$

Each element of the time series is treated as a random variable with a probability distribution. As with the cross-section variables of our earlier analysis, we assume that the distributions of the individual elements of the series have parameters in common. For example,

¹ The theory of discrete-time stochastic processes can be extended to continuous time, but we need not consider this here because econometricians typically have data only at discrete intervals.

we may assume that the variance of each y_t is the same and that the covariance between each adjacent pair of elements $\text{cov}(y_t, y_{t-1})$ is the same. If the distribution of y_t is the same for all values of t , then we say that the series y is *stationary*, which we define more precisely below. The aim of our statistical analysis is to use the information contained in the sample to infer properties of the underlying distribution of the time-series process (such as the covariances) from the sample of available observations.

1.1.2 White noise

The simplest kind of time-series process corresponds to the classical, normal error term of the Gauss-Markov Theorem. We call this kind of variable *white noise*. If a variable is white noise, then each element has an identical, independent, mean-zero distribution. Each period's observation in a white-noise time series is a complete "surprise": nothing in the previous history of the series gives us a clue whether the new value will be positive or negative, large or small.

Formally, we say that ε is a white-noise process if

$$\begin{aligned} E(\varepsilon_t) &= 0, \forall t, \\ \text{var}(\varepsilon_t) &= \sigma^2, \forall t, \\ \text{cov}(\varepsilon_t, \varepsilon_{t-s}) &= 0, \forall s \neq 0. \end{aligned} \tag{1.1}$$

Some authors define white noise to include the assumption of normality, but although we will usually assume that a white-noise process ε_t follows a normal distribution we do not include that as part of the definition. The covariances in the third line of equation (1.1) have a special name: they are called the *autocovariances* of the time series. The s -order autocovariance is the covariance between the value at time t and the value s periods earlier at time $t - s$.

Fluctuations in most economic time series tend to persist over time, so elements near each other in time are correlated. These series are *serially correlated* and therefore cannot be white-noise processes. However, even though most variables we observe are not simple white noise, we shall see that the concept of a white-noise process is extremely useful as a building block for modeling the time-series behavior of serially correlated processes.

1.1.3 White noise as a building block

We model serially correlated time series by breaking them into two additive components:

$$y_t = \overbrace{g(y_{t-1}, y_{t-2}, \dots, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots)}^{\text{effects of past on } y_t} + \varepsilon_t. \tag{1.2}$$

The function g is the part of the current y_t that can be predicted based on the past. The variable ε_t , which is assumed to be white noise, is the fundamental *innovation* or *shock* to the series at time t —the part that cannot be predicted based on the past history of the series.

From equation (1.2) and the definition of white noise, we see that the best possible forecast of y_t based on all information observed through period $t - 1$ (which we denote I_{t-1}) is

$$E(y_t | I_{t-1}) = g(y_{t-1}, y_{t-2}, \dots, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$$

and the unavoidable forecast error is the innovation $y_t - E(y_t | I_{t-1}) = \varepsilon_t$.

Every stationary time-series process and many useful non-stationary ones can be described by equation (1.2). Thus, although most economic time series are not white noise, any series can be decomposed into predictable and unpredictable components, where the latter is the fundamental underlying white-noise process of the series.

Characterizing the behavior of a particular time series means describing two things: (1) the function g that describes the part of y_t that is predictable based on past values and (2) the variance of the innovation ε_t . The most common specifications we shall use are *linear stochastic processes*, where the function g is linear and the number of lagged values of y and ε appearing in g is finite.

1.2 Linear Stochastic Processes and the Lag Operator

1.2.1 ARMA processes

Linear stochastic processes can be written in the form

$$y_t = \alpha + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad (1.3)$$

where

$$g(y_{t-1}, y_{t-2}, \dots, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = \alpha + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}. \quad (1.4)$$

The process y_t described by equation (1.3) has two sets of terms in the g function in addition to the constant. There are p *autoregressive* terms involving p lagged values of the variable and q *moving-average* terms with q lagged values of the innovation ε . We often refer to such a stochastic process as an autoregressive-moving-average process of order (p, q) , or an *AR-MA*(p, q) process.² If $q = 0$, so there are no moving-average terms, then the process is a pure autoregressive process: AR(p). Similarly, if $p = 0$ and there are no autoregressive terms, the process is a pure moving-average: MA(q). We will use autoregressive processes extensively in these chapters, but moving-average processes will appear relatively rarely.

² The analysis of ARMA processes was pioneered by Box and Jenkins (1976).

1.2.2 Lag operator

It is convenient to use a time-series “operator” called the *lag operator* when writing equations such as (1.3). The lag operator $L(\cdot)$ is a mathematical operator or function, just like the negation operator $-(\cdot)$ that turns a number or expression into its negative or the inverse operator $(\cdot)^{-1}$ that takes the reciprocal. Just as with these other operators, we shall often omit the parentheses when the argument to which the operator is applied is clear without them.

The lag operator’s argument is an element of a time series; when we apply the lag operator to an element y_t we get its predecessor y_{t-1} :

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

We can apply the lag operator iteratively to get lags longer than one period. When we do this, it is convenient to use an exponent on the L operator to indicate the number of lags:

$$L^2(y_t) \equiv L[L(y_t)] = L[y_{t-1}] = y_{t-2},$$

and, by extension, $L^n(y_t) = L^n y_t \equiv y_{t-n}$.

Using the lag operator, equation (1.3) can be written as

$$y_t - \phi_1 L(y_t) - \phi_2 L^2(y_t) \dots - \phi_p L^p(y_t) = \alpha + \varepsilon_t + \theta_1 L(\varepsilon_t) + \theta_2 L^2(\varepsilon_t) + \dots + \theta_q L^q(\varepsilon_t). \quad (1.5)$$

Working with the left-hand side of equation (1.5),

$$\begin{aligned} & y_t - \phi_1 L(y_t) - \phi_2 L^2(y_t) \dots - \phi_p L^p(y_t) \\ &= y_t - \phi_1 Ly_t - \phi_2 L^2 y_t - \dots - \phi_p L^p y_t \\ &= (1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) y_t \\ &\equiv \phi(L) y_t. \end{aligned} \quad (1.6)$$

The expression in parentheses in the third line of equation (1.6), which the fourth line defines as $\phi(L)$, is a *polynomial in the lag operator*. Moving from the second line to the third line involves factoring y_t out of the expression, which would be entirely transparent if L were a number. Even though L is an operator, we can perform this factorization as shown, writing $\phi(L)$ as a composite lag function to be applied to y_t .

Similarly, we can write the right-hand side of (1.5) as

$$\begin{aligned} & \alpha + \varepsilon_t + \theta_1 L(\varepsilon_t) + \theta_2 L^2(\varepsilon_t) + \dots + \theta_q L^q(\varepsilon_t) \\ &= \alpha + (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t \equiv \alpha + \theta(L) \varepsilon_t, \end{aligned}$$

with $\theta(L)$ defined by the second line as the moving-average polynomial in the lag operator. Using lag operator notation, we can rewrite the ARMA(p, q) process in equation (1.5) compactly as

$$\phi(L)y_t = \alpha + \theta(L)\varepsilon_t. \quad (1.7)$$

The left-hand side of (1.7) is the autoregressive part of the process and the right-hand side is the moving-average part (plus a constant that allows the mean of y to be non-zero).

1.2.3 Infinite moving-average representation of an ARMA process

Any finite ARMA process can be expressed as a (possibly) infinite moving average process. We can see intuitively how this works by recursively substituting for lags of y_t in equation (1.3). For simplicity, consider the ARMA(1, 1) process with zero mean:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}. \quad (1.8)$$

We assume that all observations t are generated according to (1.8). Lagging equation (1.8) one period yields $y_{t-1} = \phi_1 y_{t-2} + \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2}$, which we can substitute into (1.8) to get

$$\begin{aligned} y_t &= \phi_1 (\phi_1 y_{t-2} + \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2}) + \varepsilon_t + \theta_1 \varepsilon_{t-1} \\ &= \phi_1^2 y_{t-2} + \varepsilon_t + (\phi_1 + \theta_1) \varepsilon_{t-1} + \phi_1 \theta_1 \varepsilon_{t-2}. \end{aligned}$$

Further substituting $y_{t-2} = \phi_1 y_{t-3} + \varepsilon_{t-2} + \theta_1 \varepsilon_{t-3}$ yields

$$\begin{aligned} y_t &= \phi_1^2 (\phi_1 y_{t-3} + \varepsilon_{t-2} + \theta_1 \varepsilon_{t-3}) + \varepsilon_t + (\phi_1 + \theta_1) \varepsilon_{t-1} + \phi_1 \theta_1 \varepsilon_{t-2} \\ &= \phi_1^3 y_{t-3} + \varepsilon_t + (\phi_1 + \theta_1) \varepsilon_{t-1} + (\phi_1^2 + \phi_1 \theta_1) \varepsilon_{t-2} + \phi_1^2 \theta_1 \varepsilon_{t-3}. \end{aligned}$$

If we continue substituting in this manner we can push the lagged y term on the right-hand side further and further into the past. As long as $|\phi_1| < 1$, the coefficient on the lagged y term gets smaller and smaller as we continue substituting, approaching zero in the limit.³ However, each time we substitute we add another lagged ε term to the expression.

In the limit, if we were to hypothetically substitute infinitely many times, the lagged y term would converge to zero and there would be infinitely many lagged ε terms on the right-hand side:

$$y_t = \varepsilon_t + \sum_{s=1}^{\infty} \phi_1^{s-1} (\phi_1 + \theta_1) \varepsilon_{t-s}, \quad (1.9)$$

³ We shall see presently that the condition $|\phi_1| < 1$ is necessary for the ARMA(1, 1) process to be stationary.

which is a moving-average process with (if $\phi_1 \neq 0$) infinitely many terms. Equation (1.9) is referred to as the *infinite-moving-average representation* of the ARMA(1, 1) process; it exists provided that the process is stationary, which in turn requires $|\phi_1| < 1$ so that the lagged y term converges to zero after infinitely many substitutions.

The lag-operator notation can be useful in deriving the infinite-moving-average representation of an ARMA process. Starting from equation (1.7), we can “divide by” the autoregressive lag polynomial on both sides to get

$$\begin{aligned} y_t &= \frac{\alpha}{\phi(L)} + \frac{\theta(L)}{\phi(L)} \varepsilon_t \\ &= \frac{\alpha}{1 - \phi_1 - \phi_2 - \dots - \phi_p} + \frac{1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q}{1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p} \varepsilon_t. \end{aligned} \quad (1.10)$$

In the first term, there is no time-dependent variable on which the lag operator can operate, so it just operates on the constant one, which gives the expression in the denominator. The first term is the unconditional mean of y_t because the expected value of all of the ε variables in the second term is zero.

The quotient in front of ε_t in the second term is the ratio of two polynomials, which, in general, will be an infinite polynomial in L . The terms in this quotient can be (laboriously) evaluated for any particular ARMA model by polynomial long division.

For the zero-mean ARMA(1, 1) process we considered earlier, the first term is zero (because $\alpha = 0$). We can use a shortcut to evaluate the second term. We know from basic algebra that

$$\sum_{i=0}^{\infty} a^i = \frac{1}{1-a}$$

if $|a| < 1$. Assuming that $|\phi_1| < 1$, we can similarly write

$$\sum_{i=0}^{\infty} (\phi_1 L)^i = \frac{1}{1 - \phi_1 L} \quad (1.11)$$

(we can pretend that L is one for purposes of assessing whether this expression converges). From (1.10), we get

$$y_t = \left(\frac{1 + \theta_1 L}{1 - \phi_1 L} \right) \varepsilon_t = \left[\sum_{i=0}^{\infty} (\phi_1 L)^i (1 + \theta_1 L) \right] \varepsilon_t,$$

which is equivalent to equation (1.9).

1.3 Stationary and nonstationary time-series processes

The most important property of any time-series process is whether or not it is *stationary*. We shall define stationarity formally below. Intuitively, a process is stationary if a time series following that process has the same probability distribution at every point in time. For a stationary process, the effects of any shock must eventually die out as the shock recedes into the infinite past.

Variables that have a trend component are one example of nonstationary time series. A trended variable grows (or shrinks, if the trend is negative) over time, so the mean of its distribution is not the same at all dates.

In recent decades, econometricians have discovered that the traditional regression methods we studied earlier are poorly suited to the estimation of models with nonstationary variables. So before we begin our analysis of time-series regressions we must define stationarity clearly. We then proceed in subsequent chapters to examine regression techniques that are suitable for stationary and nonstationary time series.

1.3.1 Stationarity and ergodicity

A time-series process y is *strictly stationary* if the joint probability distribution of any pair of observations from the process (y_t, y_{t-s}) depends only on s , the distance in time between the observations, and not on t , the time in the sample from which the pair is drawn.⁴ For a time series that we observe for 1900 through 2000, this means that the joint distribution of the observations for 1920 and 1925 must be identical to the joint distribution of the observations for 1980 and 1985.

We shall work almost exclusively with normally distributed time series, so we can rely on the weaker concept of *weak stationarity* or *covariance stationarity*. For normally distributed time series (but not for non-normal series in general), covariance stationarity implies strict stationarity. A time-series process is covariance stationary if the means, variances, and covariance of any pair of observations (y_t, y_{t-s}) depends only on s and not on t .

Formally, a time-series y is covariance-stationary if

$$\begin{aligned} E(y_t) &= \mu, \forall t, \\ \text{var}(y_t) &= \sigma_y^2, \forall t, \\ \text{cov}(y_t, y_{t-s}) &= \sigma_s, \forall t, s. \end{aligned}$$

⁴ A truly formal definition would consider more than two observations, but the idea is the same. See Hamilton (1994, 46).

In words, this definition says that μ , σ_y^2 , and σ_s do *not* depend on t , so that the moments of the distribution are the same at every date in the sample.

The *autocovariances* σ_s of a series measure its degree of persistence. A shock to a series with large (positive) autocovariances will die out more slowly than if the same shock happened to series with smaller autocovariances. Because autocovariances (like all covariances and variances) depend on the units in which the variable is measured, we more commonly express persistence in terms of *autocorrelations* ρ_s . We define the s -order autocorrelation of y as

$$\rho_s \equiv \text{corr}(y_t, y_{t-s}) = \frac{\sigma_s}{\sigma_y^2}.$$

By construction, $\rho_0 = 1$ is the correlation of y_t with itself.

A concept that is closely related to stationarity is *ergodicity*. Hamilton (1994) shows that stationary, normally distributed time-series processes are ergodic if the limit of the sum of its absolute autocorrelations $\sum_{s=0}^{\infty} |\rho_s|$ is finite. This condition is stronger than $\lim_{s \rightarrow \infty} \rho_s = 0$; The autocorrelations must not only go to zero, they must go to zero “sufficiently quickly.” In our analysis, we shall routinely assume that stationary processes are also ergodic.

1.3.2 Kinds of non-stationarity

There are numerous ways in which a time-series process can fail to be stationary. A simple one is *breakpoint non-stationarity*, in which the parameters of the data-generating process change at a particular date. For example, there is evidence that many macroeconomic time series behaved differently after the oil shock of 1973 than they did before it. Another example would be an explicit change in a law or policy that would lead to a change in the behavior of a series at the time the new regime comes into effect.

A second common example is *trend non-stationarity*, in which a series has a deterministic trend. An example of such a series would be $y_t = \alpha + \beta t + x_t$, where β is the trend rate of increase in y and x is a stationary time series that measures the deviation of y from its fixed trend line. For a trend non-stationary series, the mean of the series varies linearly with time: in the example, $E(y_t) = \alpha + \beta t$, assuming that $E(x_t) = 0$.

Most of the non-stationary processes we shall examine have neither breakpoints nor trends. Instead, they are highly persistent, “integrated” processes that are sometimes called *stochastic trends*. We can think of these series as having a trend in which the change from period to period (β in the deterministic trend above) is a stationary random variable. Integrated processes can be made stationary by taking differences over time. We shall examine integrated processes in considerable detail below.

1.3.3 Stationarity in MA(q) models

Consider first a model that has no autoregressive terms, the MA(q) model

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q},$$

where ε is white noise with variance σ^2 . (We assume that the mean of the series is zero in all of our stationarity examples. Adding a constant α to the right-hand side does not change the variance or autocovariances and simply adds a time-invariant constant to the mean, so it does not affect stationarity.)

$$E(y_t) = E(\varepsilon_t) + \theta_1 E(\varepsilon_{t-1}) + \dots + \theta_q E(\varepsilon_{t-q}) = 0$$

because the mean of all of the white-noise errors is zero. The variance of y_t is

$$\begin{aligned} \text{var}(y_t) &= \text{var}(\varepsilon_t) + \theta_1^2 \text{var}(\varepsilon_{t-1}) + \dots + \theta_q^2 \text{var}(\varepsilon_{t-q}) \\ &= (1 + \theta_1^2 + \dots + \theta_q^2) \sigma^2, \end{aligned}$$

where we can ignore the covariances among the ε terms of various dates because they are zero for a white-noise process.

Finally, the s -order autocovariance of y , the covariance between values of y that are s periods apart, is

$$\text{cov}(y_t, y_{t-s}) = E\left[\left(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}\right)\left(\varepsilon_{t-s} + \theta_1 \varepsilon_{t-s-1} + \dots + \theta_q \varepsilon_{t-s-q}\right)\right]. \quad (1.12)$$

The only terms in this expression that will be non-zero are those for which the subscripts match. Thus,

$$\begin{aligned} \text{cov}(y_t, y_{t-1}) &= E\left[\left(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}\right)\left(\varepsilon_{t-1} + \theta_1 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-1-q}\right)\right] \\ &= (\theta_1 + \theta_1 \theta_2 + \theta_2 \theta_3 + \dots + \theta_{q-1} \theta_q) \sigma^2 \equiv \sigma_1, \end{aligned}$$

$$\begin{aligned} \text{cov}(y_t, y_{t-2}) &= E\left[\left(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}\right)\left(\varepsilon_{t-2} + \theta_1 \varepsilon_{t-3} + \dots + \theta_q \varepsilon_{t-2-q}\right)\right] \\ &= (\theta_2 + \theta_1 \theta_3 + \theta_2 \theta_4 + \dots + \theta_{q-2} \theta_q) \sigma^2 \equiv \sigma_2, \end{aligned}$$

and so on. The mean, variance, and autocovariances derived above do not depend on t , so the MA(q) process is stationary, regardless of the values of the θ parameters.

Moreover, for any $s > q$, the time interval t through $t - q$ in the first expression in (1.12) does not overlap the time interval $t - s$ through $t - s - q$ of the second expression. Thus, there are no contemporaneous cross-product terms and $\sigma_s = 0$ for $s > q$. This implies that all finite moving-average processes are ergodic.

1.3.4 Stationarity in AR(p) processes

While all finite MA processes are stationary, this is not true of autoregressive processes, so we now consider stationarity of pure autoregressive models. The zero-mean AR(p) process is written as

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t.$$

We consider first the AR(1) process

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, \quad (1.13)$$

then generalize. Equation (1.11) shows that the infinite-moving-average representation of the AR(1) process is

$$y_t = \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}. \quad (1.14)$$

Taking the expected value of both sides shows that $E(y_t) = 0$ because all of the white noise ε terms in the summation have zero expectation.

Taking the variance of both sides of equation (1.14) gives us

$$\sigma_y^2 = \text{var}(y_t) = \sum_{i=0}^{\infty} (\phi_1^2)^i \sigma^2 \quad (1.15)$$

because the covariances of the white-noise innovations at different points in time are all zero. If $|\phi_1| < 1$, then $\phi_1^2 < 1$ and the infinite series in equation (1.15) converges to

$$\frac{\sigma^2}{1 - \phi_1^2}.$$

If $|\phi_1| \geq 1$, then the variance in equation (1.15) is infinite and the AR(1) process is nonstationary. Thus, y_t has finite variance only if $|\phi_1| < 1$, which is a necessary condition for stationarity.

It is straightforward to show that

$$\begin{aligned} \sigma_s &= \text{cov}(y_t, y_{t-s}) = \phi_1^s \sigma_y^2, \text{ and} \\ \rho_s &= \phi_1^s, \quad s = 1, 2, \dots \end{aligned}$$

These covariances are finite and independent of t as long as $|\phi_1| < 1$. Thus, $|\phi_1| < 1$ is a necessary and sufficient condition for covariance stationarity of the AR(1) process. The condition $|\phi_1| < 1$ also assures that the AR process is ergodic because

$$\sum_{s=0}^{\infty} |\rho_s| = \sum_{s=0}^{\infty} |\phi_1^s| = \sum_{s=0}^{\infty} |\phi_1|^s = \frac{1}{1-|\phi_1|} < \infty.$$

Intuitively, the AR(1) process is stationary and ergodic if the effect of a past value of y dies out as time passes. The same intuition holds for the AR(p) process, but the mathematical analysis is more challenging. For the general case, we must rely on the concept of polynomial roots.

In algebra, we define the *roots of a polynomial* $P(x)$ to be the values of x for which $P(x) = 0$. To assess the stationarity of an autoregressive time-series process $\phi(L)y_t = \varepsilon_t$, we apply the concept of polynomial roots to the autoregressive lag polynomial $\phi(L)$. Of course, L is an operator rather than a variable, but the “roots” of $\phi(L)$, the values of “ L ” for which $\phi(L) = 0$, tell us whether or not the autoregressive process is stationary.

In the case of the AR(1) process we analyzed above, $\phi(L) = 1 - \phi_1 L$. We calculate the root(s) r by setting $1 - \phi_1 L = 0$. The only value of L that satisfies $1 - \phi_1 L = 0$ is $r_1 = 1/\phi_1$, which is the single root of this autoregressive lag polynomial. We demonstrated above that the AR(1) process is stationary if and only if $|\phi_1| < 1$, which is equivalent to the condition on the root $|r_1| = |1/\phi_1| > 1$. Thus, the AR(1) process is stationary if and only if the root of $\phi(L)$ is greater than one in absolute value.

For the AR(p) process, the polynomial $\phi(L)$ is a p -order polynomial, involving powers of L ranging from zero up to p . Polynomials of order p have p roots that satisfy $\phi(L) = 0$, some or all of which may be complex numbers with imaginary parts. The AR(p) process is stationary if and only if *all* of the roots satisfy $|r_j| > 1$, where the absolute value notation is interpreted as a modulus if r_j has an imaginary part. If a complex number is written $a + bi$, then the modulus is $\sqrt{a^2 + b^2}$, which (by the Pythagorean Theorem) is the distance of the point $a + bi$ from the origin in the complex plane. Therefore, stationarity requires that all of the roots of $\phi(L)$ lie more than one unit from the origin, or “outside the unit circle,” in the complex plane.

To see how this works, consider the AR(2) process $y_t = 0.75y_{t-1} - 0.125y_{t-2} + \varepsilon_t$. The lag polynomial for this process is $1 - 0.75L + 0.125L^2$. We can find that the (two) roots are $r_1 = 1/0.5 = 2$ and $r_2 = 1/0.25 = 4$ either by using the quadratic formula or by factoring the polynomial into $(1 - 0.5L)(1 - 0.25L)$. Both roots are real and greater in absolute value than one, so this AR process is stationary.

As a second example, we examine $y_t = 1.25y_{t-1} - 0.25y_{t-2} + \varepsilon_t$ with lag polynomial $1 - 1.25L + 0.25L^2$. This polynomial factors into $(1 - L)(1 - 0.25L)$, so the roots are $r_1 = 4$ and $r_2 = 1$. One root (4) is stable, but the second (1) lies on, not outside, the unit circle. We shall

see shortly that nonstationary autoregressive processes that have *unit roots* are called *integrated processes* and have special properties.

1.3.5 Stationarity in ARMA(p, q) processes

We have so far established that

- all finite moving-average are stationary, and
- autoregressive processes may be stationary or nonstationary depending on their parameters.

It should come as no surprise, then, that the stationarity of the ARMA(p, q) process depends *only* on the parameters of the autoregressive part, and not at all on the moving-average part. The ARMA(p, q) process $\phi(L)y_t = \theta(L)\varepsilon_t$ is stationary if and only if the p roots of the order- p polynomial $\phi(L)$ all lie outside the unit circle.

1.4 Random walks and other “unit-root” processes

An interesting case of nonstationary processes are those that have roots on, but not inside, the unit circle. The simplest case of a unit-root process is the AR(1) process with $\phi_1 = 1$, which is called a *random walk*.

Plugging $\phi_1 = 1$ into equation (1.13), the random-walk process is

$$y_t = y_{t-1} + \varepsilon_t,$$

which, in words, says that the value of y in period t is equal to its value in $t - 1$, plus a random “step” due to the white-noise shock ε_t . Bringing y_{t-1} to the left-hand side,

$$\Delta y_t \equiv y_t - y_{t-1} = \varepsilon_t,$$

so the *first difference* of y —the change in y from one period to the next—is white noise.

Writing the random walk in the form of a lag polynomial gives us

$$(1 - L)y_t = \varepsilon_t. \tag{1.16}$$

The term $(1 - L)$ is the *difference operator* (Δ) expressed in terms of lag notation: the current value minus last period’s value. The lone root of the polynomial $(1 - L)$ is 1, so this process has a single, unitary root.

Define the time series z given by

$$z_t = \Delta y_t = (1 - L)y_t = y_t - y_{t-1} \tag{1.17}$$

to be the first difference of y . If y follows a random walk, then by substituting from (1.17) into (1.16), $z_t = \varepsilon_t$ is white noise, which is a stationary process.

Eliminating the non-stationarity of a random walk by taking the difference is an example of a more general proposition: We can always make a unit-root process stationary by differencing. We now digress to prove this proposition, though the proof is not essential to our further analysis and can be skipped by those who find it confusing.

Suppose that y is an AR(p) process $\phi(L)y_t = \varepsilon_t$ and that d of the p roots of $\phi(L)$ are equal to one, with the other $p - d$ roots lying outside the unit circle (in the region of stationarity). (It is possible, as in the random-walk case, that $d = p$.)

Algebra tells us that we can factor any polynomial $\phi(L)$ as follows:

$$\phi(L) = \left(1 - \frac{1}{r_1}L\right) \left(1 - \frac{1}{r_2}L\right) \cdots \left(1 - \frac{1}{r_p}L\right),$$

where r_1, r_2, \dots, r_p are the p roots of the polynomial. The factor $(1 - L)$ will appear d times in this expression, once for each of the d roots that equal one. Suppose that we arbitrarily order the roots so that the first d roots are equal to one, then

$$\phi(L) = (1 - L)^d \left(1 - \frac{1}{r_{d+1}}L\right) \cdots \left(1 - \frac{1}{r_p}L\right). \quad (1.18)$$

To achieve stationarity in the presence of d unit roots, we must apply d differences to y_t . Let

$$z_t \equiv \Delta^d y_t = (1 - L)^d y_t. \quad (1.19)$$

Substituting z_t from (1.19) into (1.18) shows that we can write

$$\phi(L)y_t = \left(1 - \frac{1}{r_{d+1}}L\right) \cdots \left(1 - \frac{1}{r_p}L\right) \Delta^d y_t = \left(1 - \frac{1}{r_{d+1}}L\right) \cdots \left(1 - \frac{1}{r_p}L\right) z_t \equiv \phi^*(L)z_t = \varepsilon_t. \quad (1.20)$$

Because r_{d+1}, \dots, r_p are all (by assumption) outside the unit circle, the d th difference series z_t is stationary, described by the stationary $(p - d)$ -order autoregressive process $\phi^*(L)z_t = \varepsilon_t$ defined in (1.20).

To summarize, we can make an autoregressive process with d unit roots stationary by differencing the series d times—in other words, the d th difference of such a series is stationary. A nonstationary series that can be made stationary by differencing d times is said to be *integrated of order d* , abbreviated by $I(d)$. The term “integrated” comes from the calculus property that integration and differentiation are inverse operations. In the above example, z

equals the non-stationary series y differenced d times, and therefore y is equal to the stationary series z “integrated” d times.

The above examples have examined only pure autoregressive processes, but the analysis generalizes to ARMA processes as well. The general, integrated time-series process can be written as

$$\phi(L)\Delta^d y_t = \theta(L)\varepsilon_t,$$

where $\phi(L)$ is an order- p polynomial with roots outside the unit circle and $\theta(L)$ is an order- q polynomial. This process is called an *autoregressive integrated moving-average process*, or *ARIMA(p, d, q) process*.

1.5 Time Series in Stata

1.5.1 Defining the data set

To take advantage of time-series operations and commands in Stata, you must first define the time-series structure of your dataset. In order to do so, there must be a “time variable” that tells where each observation fits in the time sequence. The simplest time variable would just be a counter that starts at 1 with the first observation. If no time variable exists in your dataset, you can create such a counter with the following command:

```
generate timevar = _n
```

Once you have created a time variable (which be called anything), you must use the Stata command `tsset` to tell Stata about it. The simplest form of the command is

```
tsset timevar
```

where `timevar` is the name of the variable containing the time information.

Although the simple time variable is a satisfactory option for time-series data of any frequency, Stata is capable of customizing the time variable with a variety of time formats, with monthly, quarterly, and yearly being the ones most used by economists.

If you have monthly time-series data, the dataset have two numeric variables, one indicating the year and one the month. You can use Stata functions to combine these into a single time variable, then assign an appropriate format so that your output will be intuitive. Suppose that you have a dataset in which the variable `mon` contains the month number and `yr` contains the year. The `ym` function in Stata will translate these into a single variable, which can then be used as the time variable:

```
generate t = ym(yr, mon)
tsset t, monthly
```

(There are similar Stata functions for quarterly data and other frequencies, as well as functions that translate alphanumeric data strings in case your month variable has the values “January,” “February,” etc.) The `generate` command above assigns an integer value to the variable `t` on a scale in which January 1960 has the value of 0, February 1960 = 1, etc. The `monthly` option in the `tsset` command tells Stata explicitly that your dataset has a monthly time unit.

To get `t` to be displayed in a form that is easily interpreted as a date, set the variable’s format to `%tm` with the command:

```
format t %tm
```

(Again, there are formats for quarterly data and other frequencies as well as monthly.) After setting the format in this way, a value of zero for variable `t` will display as `1960m1` rather than as `0`.

One purpose of declaring a time variable is that it allows you to inform Stata about any gaps in your data without having “empty observations” in your dataset. Suppose that you have annual data from 1929 to 1940 and 1946 to 2010, with the World War II years missing from your dataset. The variable `year` contains the year number and jumps right from a value of 1940 for the 12th observation to 1946 for the 13th, with no intervening observations. Declaring the dataset with

```
tsset year, yearly
```

will inform Stata that observations corresponding to 1941–1945 are missing from your dataset. If you ask for the lagged value one period before the 1946 observation, Stata will correctly return a missing value rather than incorrectly using the value for 1940. If for any reason you need to add these missing observations to your dataset for the gap years, the Stata command `tsfill` will do so, filling them with missing values.

1.5.2 Lag and difference operators

The lag operator in Stata is `L.` or `l.` (capital or lower-case L followed by a period), so to get create a new variable that contains the lag of `y`, we could type

```
generate lagy = l.y
```

Lags of more than one period can be obtained by including a number after the `l` in the lag operator:

```
generate ytwicelagged = l2.y
```

In order to use the lag function, the time-series structure of the dataset must have already been defined using the `tsset` command.

You can use the difference operator `D.` or `d.` to take the difference of a variable. The following two commands produce identical results:

```
generate changeiny = d.y  
generate changeiny = y - l.y
```

As with the lag operator, a higher-order difference can be obtained by adding a number after the `d`:

```
generate seconddiffofy = d2.y
```

If data for the observation prior to period t is missing, either explicitly in the dataset or because there is no observation in the dataset for period $t - 1$, then both the lag operator and the difference operator return a missing value.

If you do not want to create extra variables for lags and differences, you can use the lag and difference operators directly in many Stata commands. For example, to regress a variable on its once-lagged value, you could type

```
regress y l.y
```

1.5.3 Descriptive statistics for time series

In addition to the standard Stata commands (such as `summarize`) that produce descriptive statistics, there are some additional descriptive commands specifically for time-series data. Most of these are available under the Graphics menu entry Time Series Graphs.

We can create a simple plot of one or more variables against the time variable using `tsline`. To calculate and graph the autocorrelations of a variable, we can use the `ac` command. The `xcorr` command will calculate the “cross correlations” between two series, which are defined as $\text{corr}(y_t, x_{t-s})$, $s = \dots, -2, -1, 0, 1, 2, \dots$.

References

- Box, George E. P., and Gwilym M. Jenkins. 1976. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- Hamilton, James D. 1994. *Time Series Analysis*. Princeton, N.J.: Princeton University Press.