Fundamental Concepts in Time Series Analysis

In: Introduction to Time Series Analysis

By: Mark Pickup
Pub. Date: 2015
Access Date: September 23, 2019
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9781452282015
Online ISBN: 9781483390857
DOI: https://dx.doi.org/10.4135/9781483390857
Print pages: 19-50

Fundamental Concepts in Time Series Analysis

In this chapter, you will be introduced to some of the fundamental concepts of time series data and time series analysis: autoregression, autocorrelation, serial correlation, stationarity, exogeneity, weak dependence, trending, seasonality, structural breaks, and stability. Some of these topics we touched on in Chapter 1. Other topics will be new to you. We will revisit and extend each of these concepts in later chapters, as they become particularly pertinent.

2.1 Autoregression

If Y_t is a time series process, we say that it is an autoregressive process of order 1 if Y_t is a function of its previous value (the lag of Y_t, denoted Y_{t-1}) and a stochastic error (ε_t):

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t, \qquad (2.1.1)$$

with ε_t ~NID $(0, \sigma_\varepsilon^2)$—that is, the ε_t values are independent and normally distributed with mean 0 and variance σ_ε^2. We use the following notation for an autoregressive process of order 1: AR(1). Without a constant, Equation 2.1.1 assumes that the long-run equilibrium for the process (if one exists) is 0. The equilibrium can be thought of as the expected value (average) of y_t over a very long period of time. This is discussed further in the next chapter, but for now it suffices to say that this assumption can be relaxed by including a constant: $y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t$.

Equation 2.1.1 describes the data-generating process for an autoregressive process. Such a time series process is modelled by including lags (or just a single lag) of the dependent variable on the right-hand side (as explanatory/independent variables). The unknown parameters α_1 and σ_ε^2 are estimated.

The subjective interpretation of a process that is autoregressive is that it has "memory." Assume that $|\alpha_1| < 1$. This is an assumption to which we will return. If something occurs to move the process out of its long-run equilibrium (an intervention), it will, in the absence of any other intervention, return to its long-run equilibrium. However, it will not return immediately. In the periods following the intervention, the process will exhibit the decaying effect of the intervention until the process returns to equilibrium.

In the public responsiveness model of Chapter 1, we noted that it might be prudent to model relative preference as a function of past relative preference in order to avoid a violation of the zero conditional mean assumption. The model based on Equation 2.1.1 is just the sort of model we might use to do this.

2.2 Autocorrelation

A concept related to autoregression is autocorrelation. We define the correlation between the elements *(Y_t)*

of a time series Y_t and the same time series lagged once, Y_{t-1}, as follows:

$$\operatorname{Corr}\left(y_t, y_{t-1}\right) \equiv \rho \equiv \frac{E\left(\left(y_t-\mu_y\right)\left(y_{t-1}-\mu_y\right)\right)}{E\left(y_t-\mu_y\right)^2}. \qquad (2.2.1)$$

Note: E() denotes the expected value $\mu_y = E(y_t)$ and μ_y denotes the theoretical mean of y_t. This is the covariance between y_t and its own lag y_{t-1} divided by the variance of y_t. This is called autocorrelation. The covariance of the element of a variable with a lag of itself is called autocovariance, although if its meaning is clear from the context, it will often simply be called covariance.

Under the assumption of stationarity, discussed below in Section 2.4 we can estimate the autocorrelation as follows:

$$\hat{\rho} = \frac{\sum_{t=2}^{T}\left(y_t-\bar{y}\right)\left(y_{t-1}-\bar{y}\right)}{\sum_{t=2}^{T}\left(y_t-\bar{y}\right)^2}. \qquad (2.2.2)$$

Note: $\bar{y}$ denotes the mean of the observed values of Y_t. Also, note that we begin the summation at *t* - 2 to reflect the fact that in practice we lose the first data point when we lag a variable. The relationship between autocorrelation and autoregression can be seen in the fact that when we model Y_t as an AR(1) process, as depicted in Equation 2.1.1, the estimate of the α_1 coefficient is the estimated autocorrelation $\hat{\rho}$.

This autocorrelation is called first-order autocorrelation, as it is the correlation between a time series and the *first lag* of itself. An extension is higher-order autocorrelation. For example, the elements of a time series $Y_t$may be correlated with the elements of the second lag of itself, Y_{t-2}. We would call this second-order autocorrelation and define it as follows:

$$\rho_2 \equiv \frac{E\left(\left(y_t-\mu_y\right)\left(y_{t-2}-\mu_y\right)\right)}{E\left(y_t-\mu_y\right)^2}. \qquad (2.2.3)$$

Generalizing this to lag *s*,

$$\rho_s \equiv \frac{E\left(\left(y_t-\mu_y\right)\left(y_{t-s}-\mu_y\right)\right)}{E\left(y_t-\mu_y\right)^2}. \qquad (2.2.4)$$

These are unknown for the true data-generating process, but they can be estimated from the sample data, assuming that the series is stationary:

$$\hat{\rho}_s = \frac{\sum_{t=s+1}^{T} \left(y_t - \bar{y}\right)\left(y_{t-s} - \bar{y}\right)}{\sum_{t=1}^{T} \left(y_t - \bar{y}\right)^2}. \qquad (2.2.5)$$

As an example, consider the second example described in Chapter 1. We can examine the monthly vote intention data (the percentage of the population indicating that they would vote for the governing party in an election) for the German government of Helmut Kohl from October 1982 till September 1998 (Figure 2.1). The first-order autocorrelation is estimated to be 0.62, and the second-order autocorrelation is estimated to be 0.57. We can plot the first-, second-, and higher-order autocorrelations in a single figure, as in Figure 2.2. This is called an autocorrelation function and will be discussed further in Chapter 5.

Figure 2.1 West German Government's Popularity, 1982 to 1998

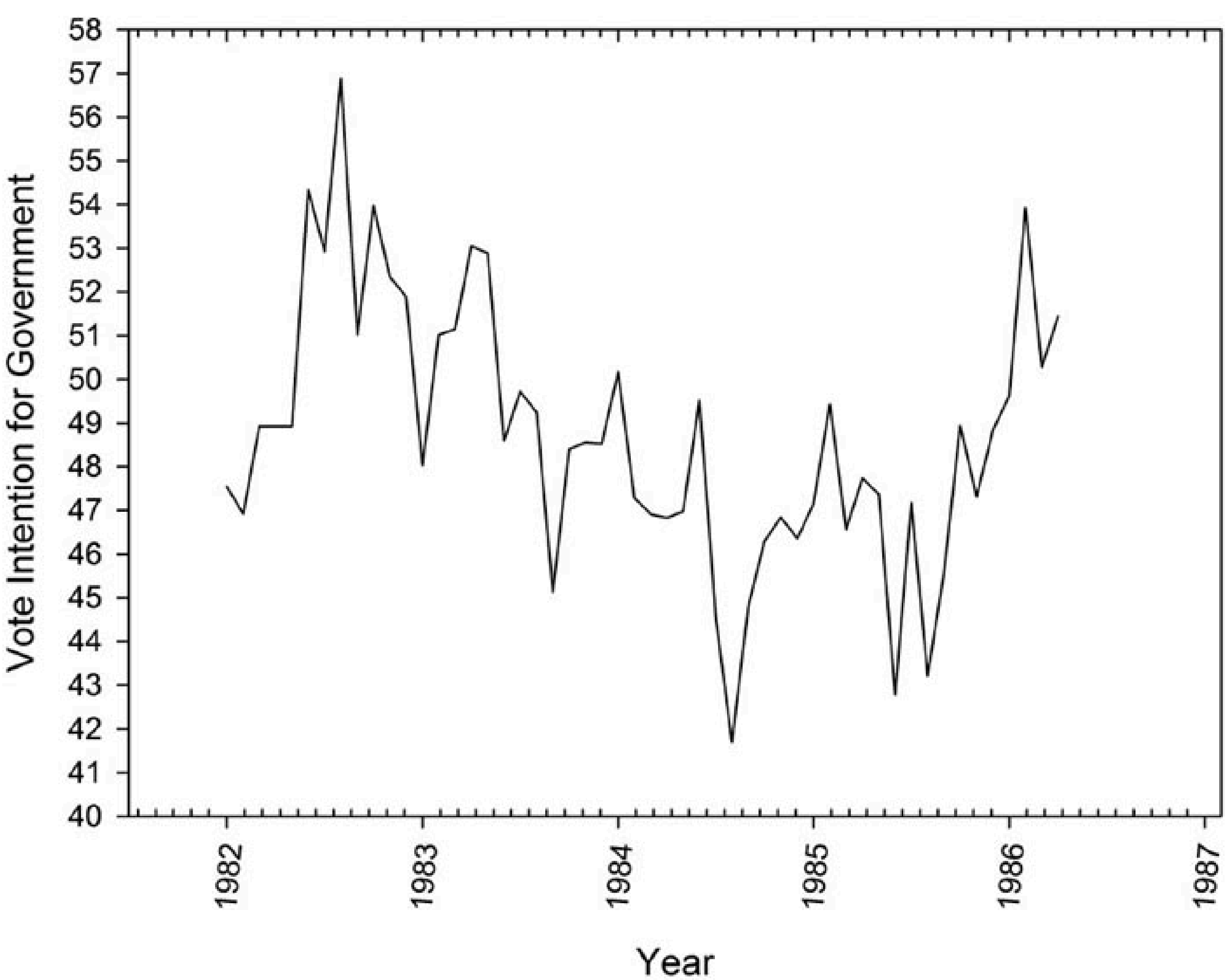
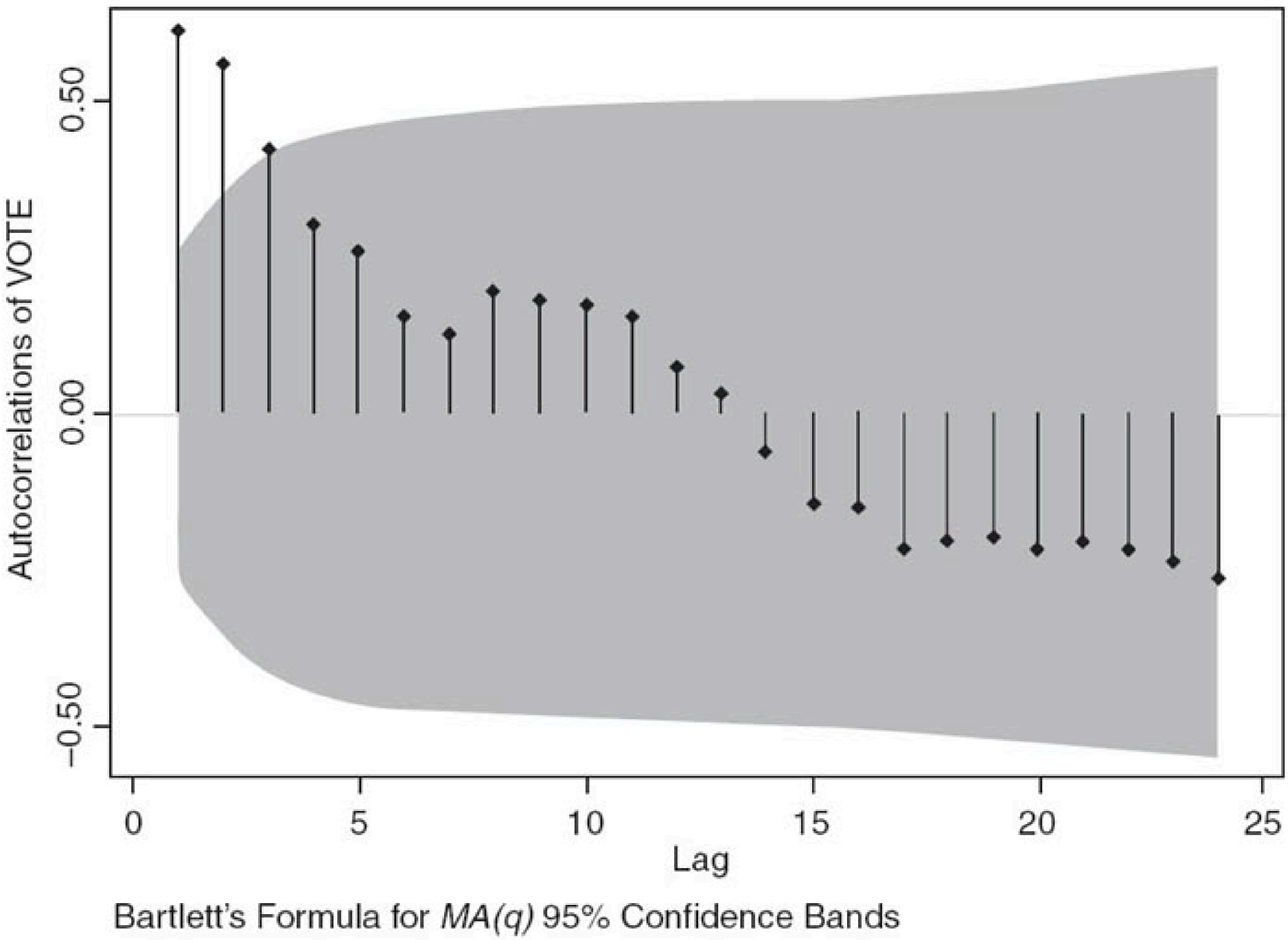


Figure 2.2 Autocorrelation Function for Vote Intention



We can also conduct a test of whether there is any evidence that the time series data depart from a white noise process. This process is defined as follows:

$$y_t = \varepsilon_t, \ \varepsilon_t \sim \mathrm{NID}(0, \sigma_\varepsilon^2). \qquad (2.2.6)$$

A white noise process (Equation 2.2.6) is equal to a stochastic error with constant variance. Such a process has no autocorrelation of any order.

As with the AR(1) process, this can be given a nonzero mean by including a constant, β_0. One of the most common tests for a white noise process is the Portmanteau *(Q)* test (Ljung & Box, 1978):

$$Q = T(T+2)\sum_{s=1}^{p}\frac{\hat{\rho}_s^2}{T-s}. \qquad (2.2.7)$$

This tests that the error autocorrelations are jointly zero, based on the first *p* autocorrelations:

$$H_0\colon \rho_1, \rho_2, \ldots, \rho_p = 0.$$

The choice of *p* is somewhat arbitrary. High *p* will capture autocorrelations at high lags but reduces the power of the test. Statistical packages have a default setting. For example, Stata uses either a *p* of 40 or *(T/2)* - 2, whichever is smallest. The *Q*-test statistic is chi-squared distributed with *p* degrees of freedom, so the *P* value can be calculated for the purpose of hypothesis testing. The null hypothesis can be interpreted as "The process is white noise."

In Chapter 3, we will look at additional ways of testing whether a process is autocorrelated. Such tests will be useful as many of our time series models depend on the assumption that the error term is a white noise process—that is, it is not autocorrelated. In Chapter 1, we discussed how correlated errors will result from a violation of the assumption of independence of observations. We call this serial correlation. We will want to test the assumption of no serial correlation by applying such white noise tests to the residuals from our models.

2.3 Serial Correlation

If the error term in a statistical model contains autocorrelation (see Section 2.2), this is often referred to as serial correlation. For example, consider the time series process Y_t modelled as a function of a constant, a single independent variable X_t, and an error term:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t. \qquad (2.3.1)$$

An example of serial correlation would be if ε_t was correlated with a lag of itself. That is,

$$\mathrm{Corr}(\varepsilon_t,\ \varepsilon_{t-h}\,|X) \neq 0\ . \qquad (2.3.2)$$

The conditioning on *X* in Equation 2.10 refers to the fact that it is the correlation in the errors of the model that contain X_t that is relevant to discussions of serial correlation. It may be the case that including an additional independent variable in a data model might result in errors that are not serially correlated. The model we ran for the public responsiveness example had the same form as Equation 2.3.1. Including a lagged-dependent variable, as we discussed, might not just avoid a violation of the zero conditional mean assumption; it might also ensure errors that are not serially correlated. We will discuss other forms of serial correlation and methods to address them in Chapter 3 and models that include a lagged-dependent variable in Chapter 4.

2.4 Stationary Stochastic Process

As is common with cross-sectional variables, time series variables contain a stochastic component. For example, the data-generating process for Y_t may be a function of a constant, zero or more independent variables, and a random error:

$$y_t = \alpha_0 + \beta_1 x_t \ldots + \varepsilon_t.$$

The stochastic error component means that our particular observed values of Y_t are only one possible realization of this time series process. The particular values we observe will depend on one particular realization (draw) of the error term for each time point (as well as the observed values of x_t): $\{\varepsilon_1, \varepsilon_t, \ldots, \varepsilon_T\}$. As this component is stochastic, another draw could have been observed. Each random draw of the set of values $\{\varepsilon_1, \varepsilon_t, \ldots, \varepsilon_T\}$ results in a different realization of Y_t. Of course, we can only ever witness one realization of Y_t. For each point in time, we only ever have one observation. This is distinct from cross-sectional data, as it is (usually) easy to conceptualize drawing additional observations for a single point in time from the population of cases.

In the case of cross-sectional data, we use regression analysis to predict $E(Y)$, based on the average value of Y across the cases in our sample (this is usually done conditioning on independent variables: $E(Y|X)$). The equivalent in time series analysis would be predicting $E(y_t)$ based on the average value of y_t for all cases at time t.

If we could, we would take a random sample of realizations of Y_t and estimate $E(y_t)$ for a particular time point t from the sample of values for y_t at that time point. But we only have one value of y_t at time point t. Therefore, we average the values across all time points (t = 1, …, T) of a single realization of Y_t to estimate $E(y_t)$.

This presents a problem. For this to be valid, we must assume that the expected value of Y_t across time converges (as the number of time points increases) on the expected value of y_t for any single time point t. This then implies that the expected value across time of a single realization of Y_t converges on the expected value of y_t for all time points t.

$$\text{As } T \rightarrow \infty\text{: } E(Y_t) \rightarrow E(y_t) \text{ for all } t,\ t = 1,\ \ldots,\ T \qquad (2.4.2)$$

Note: The left expectation is the mean of the values of a single realization of Y_t across all time points, while the right expectation is the mean of all possible realizations of y_t at any specific t. This implies that $E(y_t)$, and therefore $E(Y_t)$, must be constant over time. This is also true for estimating the variance and autocovariances of the y_ts from the variance and autocovariances of our single realization of Y_t.

This brings us to the idea of stationarity. A stochastic time series process Y_t is stationary if the distributions and joint distributions of the y_ts from t - 1 to $t = h$ are the same as they are from $t = 1 + k$ to $t = h + k$ for every and any $k \geq 1$.

$$\{y_{t=1}, y_{t=2}, y_{t=3}, \ldots, y_{t=h}\} \qquad \{y_{t=1+k}, y_{t=2+k}, y_{t=3+k}, \ldots, y_{t=h+k}\}$$

Identically Distributed

Thus, stationarity implies that the y_ts are identically distributed over time. Put another way, all marginal and joint distributions of the process are invariant to time. Therefore, we can estimate the parameters of the distributions of the y_ts from the distribution of the single realization of Y_t.

Covariance-Stationary Process

Usually, for the type of time series analysis covered in this book we shall assume a weak form of stationarity. This is called *covariance*stationarity. A stochastic process is covariance stationary if $E(y_t)$ is constant, Var(y_t) is constant, and for $k \geq 1$, Cov($y_t = k$, $y_t = h+k$) depends only on h and not on k (Greene, 2003, p. 612). The assumption regarding the covariances of y_t means that the covariance between y_t and y_{t-1} may differ from (1) y_t and y_{t-2} but not from (2) y_{t-1} and y_{t-2}. The difference is that (1) is the covariance between y_t and its second lag, which can differ from the covariance between y_t and its first lag. However, (2) is the covariance between y_t and its first lag at a different point in time. It is assumed that this makes no difference.

$$\{y_{t=1}, y_{t=2}, y_{t=3}, \dots, y_{t=h}\} \qquad \{y_{t=1+k}, y_{t=2+k}, y_{t=3+k}, \dots, y_{t=h+k}\}$$

Identical Expected
Value, Variance, and
Covariances

If Y_t is a covariance-stationary process, the expected value (mean), variance, and covariances of the y_ts can be approximated or estimated from a single realization of Y_t. For the mean,[1]

$$\bar{y}_t = \frac{\sum_{t=1}^{T} y_t}{T}. \qquad (2.4.3)$$

For the variance,

$$\widehat{\mathrm{Var}}\left(y_t\right) = \frac{\sum_{t=1}^{T}\left(y_t - \bar{y}_t\right)^2}{T-1}. \qquad (2.4.4)$$

For the *h*-order (auto)covariance,

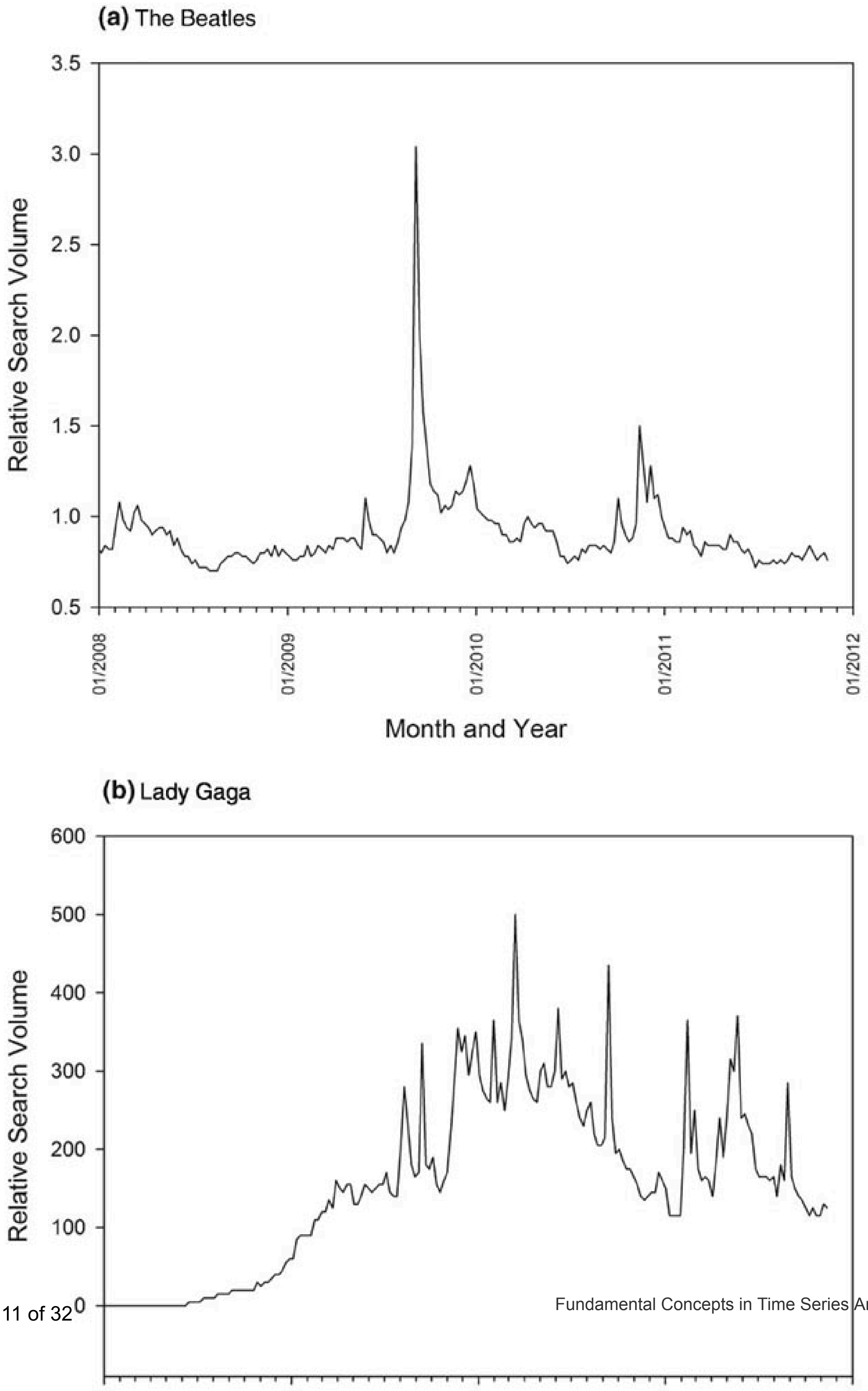
$$\widehat{\mathrm{Cov}}\left(y_t, y_{t-h}\right) = \frac{\sum_{t=1}^{T}\left(y_t - \bar{y}_t\right)\left(y_{t-h} - \bar{y}_t\right)}{T-1}. \qquad (2.4.5)$$

However, estimating the *h*-order covariance requires lag *h* of y_t. This will result in the loss of *h* data points, as described in Chapter 1—each time you lag an observed variable, you lose one data point. So unless we are in a position to collect additional data, for the *h* time points before our first observation, the estimate will actually be based on the following:

$$\widehat{\mathrm{Cov}}\left(y_t, y_{t-h}\right) = \frac{\sum_{t=1+h}^{T}\left(y_t - \bar{y}_t\right)\left(y_{t-h} - \bar{y}_t\right)}{T-h-1}. \qquad (2.4.6)$$

Instead of the covariances, we often examine the autocorrelations. These are just the covariances divided by the time series variance. We can get an idea of what type of time series will and will not be covariance stationary by considering the weekly volume of Google searches for "The Beatles" and comparing this with the weekly volume of Google searches for "Lady Gaga." Weekly data for these Google searches from the beginning of 2008 until the end of 2011 are displayed in Figure 2.3. These data are the weekly volume of Google searches for “The Beatles” and “Lady Gaga” relative to the total number of searches on Google. Each series is scaled to the mean value for the 2004–2011 period: A value of 1 indicates that the volume is equal to this mean; a value of 2 reflects a volume of twice this mean.

Figure 2.3 Weekly Volume of Google searches for “The Beatles” and “Lady Gaga”


(a) The Beatles
Relative Search Volume
3.5
3.0
2.5
2.0
1.5
1.0
0.5
01/2008
01/2009
01/2010
01/2011
01/2012
Month and Year
(b) Lady Gaga
Relative Search Volume
600
500
400
300
200
100
0

If we like, we can split the 2008–2011 period into two 2-year periods: 2008–2009 and 2010–2011. If the relative search volume is covariance stationary, then the mean, variance, and covariances (or autocorrelations) for the search volume time series should be the same in the first 2-year period as in the second 2-year period. Importantly, the assumption is that this is true for the data-generating process. However, we would also expect this to be approximately true in the observed data.

Looking at the plotted data, we would not expect the Lady Gaga search volume to be covariance stationary. The average volume is much greater in the second period relative to the first. For the Beatles search, it seems possible that the mean volumes in the first and second periods are the same. However, there is some volatility in the first period of the Beatles search volume, which is not seen in the second period—this might mean that the variances differ. The covariance stationarity assumption is about the mean, variance, and autocovariances/autocorrelations of the data-generating process, but we can look at estimates of these from our data to check the plausibility of the stationarity assumption.

For the Beatles search, the estimated mean volume is 0.93 for the 2008–2009 period and 0.88 for the 2010–2011 period. These two means are not exactly the same, but given that they are estimates, it is possible that they both come from the same data-generating process with the same mean. For the Lady Gaga search, the estimated mean volume is 94 for the 2008–2009 period and 220 for the 2010–2011 period. Clearly, the mean volume for the Lady Gaga search is much higher in the second of these two periods.

For the Beatles search volume, the estimated variance is 0.08 for the 2008–2009 period and 0.02 for the 2010–2011 period. For the Lady Gaga search volume, the estimated variance is 9,683 for the 2008–2009 period and 6,379 for the 2010–2011 period. The proportionate difference is actually bigger for the Beatles search volume. This is primarily due to the spike in searches in the first 2 weeks of September 2009, because digitally remastered versions of all the Beatles studio albums were released then.

For the Beatles search volume, the estimated first-order autocorrelation is 0.74 for the 2008–2009 period and 0.78 for the 2010–2011 period. For the Lady Gaga search volume, the estimated first-order autocorrelation is 0.98 for the 2008–2009 period and 0.73 for the 2010–2011 period. We could also estimate higher-order autocorrelations.

These are only estimates, but it would appear that the covariance stationarity assumption is badly violated for the Lady Gaga search volume time series. This assumption seems to be much more realistic for the Beatles search volume time series, although when we arbitrarily split the 2008–2011 data into two equal periods, there was greater variance in the first due to a single event that produced a large spike in searches.

When we talk below about trending, periodicity, and structural breaks, we will find that the assumption of covariance stationarity, necessary to estimate models of time series data, only needs to hold once we have controlled for the independent variables in the model—it is covariance stationarity conditioning on x_t that is relevant. Put another way, if the data-generating process is a covariance-stationary time series with a spike in search volume in the first half of September, our assumption of covariance stationarity is met as long as we control for that spike in our model. Failing to account for the spike may lead to a violation of covariance stationarity (Hendry, 2003, pp. 99–100).

It is important to note that if controls are necessary to meet the condition of covariance stationarity, we are restricted in our interpretation of the estimated model results for those controls. A term controlling for a nonstationary element should not be given any substantive interpretation beyond that of a control. Intervention analysis does soften this rule substantially—more on this in Chapters 3 and 5. If there are independent variables in our model to which we do wish to give substantive meaning and we do not think they are necessary to control for nonstationary elements, we would first need to confirm that the assumption of covariance stationarity is met without including these variables.

This is important, as it is not at all unusual to find that social science data, without any controls, violate the covariance stationarity assumption (Pickup, 2009). We have already seen this with the data used in the public responsiveness example in Chapter 1 (Figure 1.1), where the violation is due to trending—more on this to come. Any analysis of such data would need to take this problem into account.

2.5 Exogeneity

The usual exogeneity assumption in a regression model of cross-sectional data is the zero conditional mean assumption. This is described as follows:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

$$E(\varepsilon_i \mid \boldsymbol{X}) = 0. \qquad (2.5.1)$$

This implies that the expected value of the error term is not a function of the explanatory variable(s) (Greene, 2002, p. 14). For models of time series data, it is also necessary to make an exogeneity assumption of this sort:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t,$$

$$E(\varepsilon_t \mid \boldsymbol{X}) = 0. \qquad (2.5.2)$$

The bold *X* implies that the expected value of the error term in any given period is not a function of the explanatory variables in any time period—that is, at any lead or lag. For a single independent variable x_t, this can be written as

$$E(\varepsilon_t \mid x_{t+h}) = 0, \ \forall h. \qquad (2.5.3)$$

(Note: $\forall$ = "for all.") If this condition is met, x_t is said to be strictly exogenous for the estimation of β_1. What does it mean for the expected value of ε_t to be independent from an explanatory variable in the same or another time period? An obvious violation of independence is if ε_t and x_t are correlated (have covariance): $COV(\varepsilon_t, x_t) \neq 0$. The contemporaneous covariance between ε_t and x_t is

$$\mathrm{Cov}(\varepsilon_t, x_t) = E((\varepsilon_t - \mu_\varepsilon)(x_t - \mu_X)). \qquad (2.5.4)$$

The covariance between εt and Xt from another time period is

$$\mathrm{Cov}(\varepsilon_t, x_s) = E((\varepsilon_t - \mu_\varepsilon)(x_s - \mu_X));\ s = t + h,\ h \neq 0. \qquad (2.5.5)$$

For example, h = -1:

$$\mathrm{Cov}(\varepsilon_t, x_{t-1}) = E((\varepsilon_t - \mu_\varepsilon)(x_{t-1} - \mu_X)).$$

The presence of such covariance would violate the assumption of strict exogeneity. The result is a biased model estimate. A weaker assumption than strict exogeneity is that xt is contemporaneously exogenous:

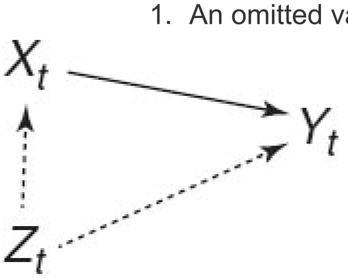
$$E(\varepsilon_t \mid x_t) = 0. \qquad (2.5.6)$$

The expected value of εt is independent of xt but is permitted to have some sort of dependence on lags or leads of xt. Contemporaneous exogeneity will be sufficient for asymptotic unbiasedness in large T and requires the additional assumption of *weak dependence* (Wooldridge, 2006, pp. 382–384).[2] This will be discussed after we review some common violations of the strict-exogeneity assumption.

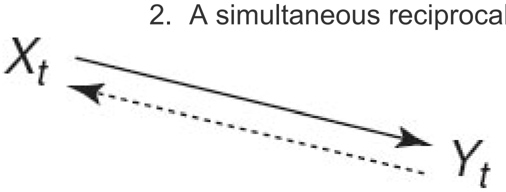
The violation of exogeneity is called endogeneity. This is a complex concept, but it is vital to understand it if we are interested in testing causal theories. But what could cause a violation of the zero conditional mean assumption? Let us say that we are interested in testing the causal effect of Xt on Yt. Our data model is as follows:

For this discussion, we will limit ourselves to time series data. For other types of data, such as those described in Chapter 1, there are other possible data-generating processes and other endogeneity concerns. The following two data-generating processes will result in two forms of endogeneity. Each will complicate a test of the causal effect of Xt on Yt.

1. An omitted variable causally prior to Xt and Yt



2. A simultaneous reciprocal relationship



A simultaneous reciprocal relationship is very difficult to deal with. It suggests that Xt and Yt concurrently have causal effects on each other, not because of their relationship with any other variable but simply because they each cause each other at the same point in time. This is a very tricky problem and must be approached with advanced, multivariate time series techniques, such as vector autoregression. This is a topic for an advanced text on multivariate time series analysis (see Brandt & Williams, 2007). For our purposes, we will assume that the *contemporaneous* direction of causality runs in one direction and we know that direction. This leaves us with the omitted-variable problem.

Before discussing the different forms of the omitted-variable problem, let us examine it in general. Suppose the data-generating process is given as

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 z_t + \varepsilon_t,$$

$$E(\varepsilon_t \mid \boldsymbol{X}) = 0. \qquad (2.5.7)$$

but we estimate the data model as

$$y_t = \tilde{\beta}_0 + \tilde{\beta}_1 x_t + \mu_t \qquad (2.5.8)$$

and therefore, $\mu_t = \beta_2 z_t + \varepsilon_t$. The ordinary least squares (OLS) estimator of $\tilde{\beta}_1$ is

$$\tilde{\beta}_1 = \frac{\sum_{t=1}^{T}\left(x_t - \bar{x}\right) y_t}{\sum_{t=1}^{T}\left(x_t - \bar{x}\right)^2}. \qquad (2.5.9)$$

A demonstration that $\tilde{\beta}_1$ is an unbiased estimator of β_1, would be that $E(\tilde{\beta}_1 \mid \boldsymbol{X}) = \beta_1$, where $\boldsymbol{X}$ denotes $\{x_1, x_2, \ldots, x_T\}$. In other words, the estimator is, on average, the true value.

If we take the expected value of Equation 2.5.9, conditioning on $\boldsymbol{X}$, we can derive the following (Wooldridge, 2006):

$$E\left(\tilde{\beta}_1 \mid \boldsymbol{X}\right) = \beta_1 + \left(\frac{1}{\sum_{t=1}^{T}\left(x_t - \bar{x}\right)^2}\right) \sum_{t=1}^{T}\left(x_t - \bar{x}\right) E\left(\mu_t \mid \boldsymbol{X}\right). \qquad (2.5.10)$$

From $\mu_t = \beta_2 z_t + \varepsilon_t$, we can work out the following:

$$E(\mu_t \mid \boldsymbol{X}) = E(\beta_2 z_t + \varepsilon_t \mid \boldsymbol{X}),$$

$$= \beta_2 z_t. \qquad (2.5.11)$$

Substituting Equation 2.5.11 into Equation 2.5.10,

$$E\left(\tilde{\beta}_1 \mid \boldsymbol{X}\right) = \beta_1 + \left(\frac{1}{\sum_{t=1}^{T}\left(x_t - \bar{x}\right)^2}\right)\sum_{t=1}^{T}\left(x_t - \bar{x}\right)\beta_2 z_t,$$

$$= \beta_1 + \beta_2 \frac{\sum_{t=1}^{T}\left(x_t - \bar{x}\right) z_t}{\sum_{t=1}^{T}\left(x_t - \bar{x}\right)^2}. \qquad (2.5.12)$$

This reveals that $\tilde{\beta}_1$ may be a biased estimator of β_1, and the magnitude of the bias is

$$\beta_2 \frac{\sum_{t=1}^{T}\left(x_t - \bar{x}\right) z_t}{\sum_{t=1}^{T}\left(x_t - \bar{x}\right)^2}. \qquad (2.5.13)$$

The $\frac{\sum_{t=1}^{T}\left(x_t - \bar{x}\right) z_t}{\sum_{t=1}^{T}\left(x_t - \bar{x}\right)^2}$ term is the OLS estimator of the slope coefficient from the regression of z_t on x_t. This means that even if $\beta_1 = 0$, $E(\tilde{\beta}_1 \mid \boldsymbol{X})$ will not be so, except in two instances. If x_t and z_t are uncorrelated in the sample, then $\frac{\sum_{t=1}^{T}\left(x_t - \bar{x}\right) z_t}{\sum_{t=1}^{T}\left(x_t - \bar{x}\right)^2} = 0$, and the magnitude of the bias is 0. Alternatively, if $\beta_2 = 0$, the magnitude of the bias is 0. This amounts to saying that z_t is not actually in the data-generating process and so is not omitted from the data model.[3]

These consequences of omitting variables from the data model are easily extended to models with additional independent variables. The direction of bias is less clear, but the lesson is the same. If an independent variable in the data-generating process is omitted from the data variable and that omitted variable is correlated with an independent variable in the data model, the OLS estimator for the slope coefficient on the variable in the data model will be biased.

An omitted variable Z_t can be of different types, resulting from different data-generating processes:

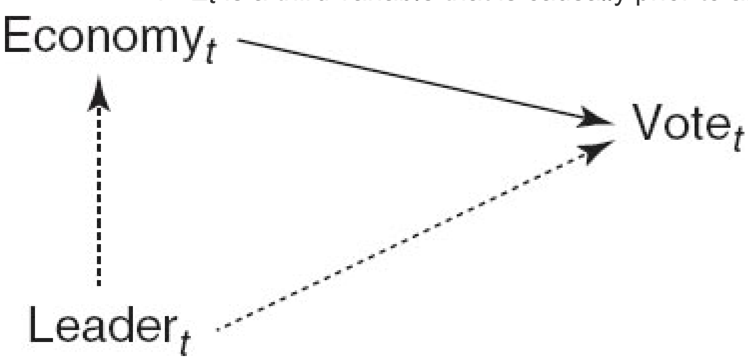
1. Z_t is a third variable that is causally prior to and correlated with current values of X_t and Y_t.
2. Z_t is a past value of a third variable (Z_{t-1}) that is causally prior to and correlated with current values of X_t and Y_t.
3. Z_t is a past value of Y_t that is causally prior to and correlated with current values of X_t and Y_t

The form of endogeneity from the first data-generating process is familiar to those who have studied the analysis of cross-sectional data. The forms of endogeneity due to the second and third data-generating processes are more specific to time series analysis, although those who have studied the use of instrumental variables in cross-sectional data may be familiar with these endogeneity problems (Greene, 2002, chap. 5). Let us use an example to discuss these endogeneity problems in a more concrete manner.

Let us say that Y_t is an aggregate measure of government vote intention, like that for the German government in Figure 2.1. Let us say that X_t is an aggregate measure of subjective evaluations of how the economy has performed. Our data model is the following:

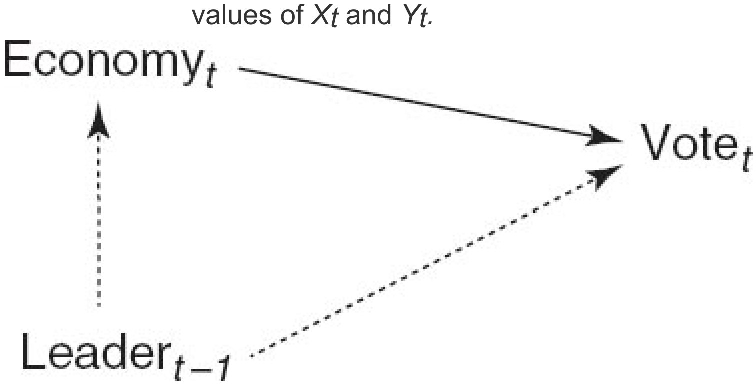
We can now review the different types of data-generating processes that will lead to an omitted-variable bias:

1. Z_t is a third variable that is causally prior to and correlated with current values of X_t and Y_t.



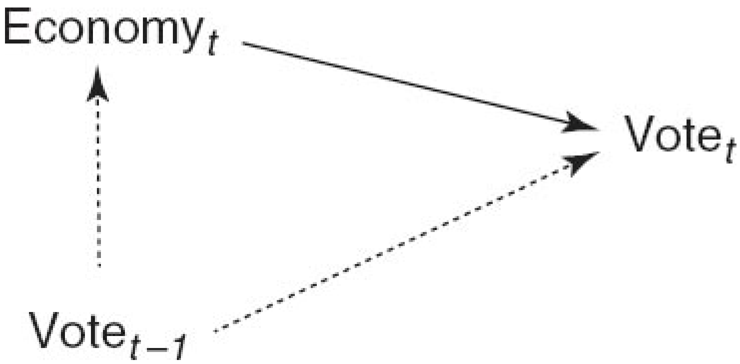
For example, both current government vote intention and current economic evaluations are a function of current leadership evaluations. Leadership effects on vote intention are well documented in a number of countries (Bittner, 2011; Johnston, 2002; Stewart & Clarke, 1992). Leadership effects on economic evaluations, such as those of prime ministerial and presidential approval, have also been demonstrated (Evans & Andersen, 2006; Evans & Pickup, 2010; Pickup & Evans, 2013). Concurrently, these will result in an apparent causal relationship between current economic evaluations and current government vote intention.

2. Z_t is a past value of a third variable (Z_{t-1}) that is causally prior to and correlated with current values of X_t and Y_t.



For example, both current government vote intention and current economic evaluations are a function of past leadership evaluations, resulting in an apparent relationship between current economic evaluations and current government approval. Work demonstrating the contemporaneous effects of leadership effects on economic valuations and vote intention is somewhat ambiguous as to whether it is contemporaneous leadership evaluations or earlier leadership evaluations that are relevant (Evans & Andersen, 2006).

3. Z_t is a past value of Y_t that is causally prior to and correlated with current values of X_t and Y_t.



For example, both current government vote intention and current economic evaluations are a function of past government vote intention, resulting in an apparent causal relationship between current economic evaluations and current government approval. Recent work on economic evaluations makes the case that past vote intention influences current economic evaluations (Evans & Pickup, 2010; Ladner & Wlezien, 2007).

In practice, we can resolve these endogeneity problems by including the omitted variable. Of course, if the omitted variable is the past value of some variable in the context of cross-sectional analysis, there is not much we can do about it unless we have an appropriate instrumental variable to be used in an instrumental variable regression (see Angrist & Pischke, 2009). This demonstrates a key benefit of the analysis of longitudinal data, such as in time series analysis. We are much more likely to be in a position to include variables that would result in bias if omitted. Longitudinal data can be a powerful tool for addressing such endogeneity problems, which can go a long way in helping us test causal hypotheses.

It should be noted that the second endogeneity problem is a violation of strict exogeneity (Equation 2.5.2) but not contemporaneous exogeneity (Equation 2.5.6). It should also be noted that the third type of data-generating process will always lead to a violation of strict exogeneity, even if we include the omitted variable—the past value of Y_t. To see this second point, consider the following.

Our data-generating process is as follows:

$$\text{Vote}_t = \alpha_0 + \beta_1 \text{Econ}_t + \beta_2 \text{Vote}_{t-1} + \varepsilon_t,$$

$$\varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2).$$

Our data model, estimated by OLS, is the following linear model (i.e., Vote$_{t-1}$ is omitted):

$$\text{Vote}_t = \tilde{\alpha}_0 + \tilde{\beta}_1 \text{Econ}_t + \tilde{\mu}_t.$$

Note: $\tilde{\mu}_t$ denotes the OLS-estimated residuals from the model that excludes Vote$_{t-1}$; and $\tilde{\beta}$ denotes the OLS-estimated coefficient that results from this exclusion. To see that this violates contemporaneous exogeneity, and $\tilde{\beta}$ may be a biased estimate of β_1, we can rewrite the data-generating process as

$$\text{Vote}_t = \alpha_0 + \beta_1 \text{Econ}_t + \mu_t,$$

$$\mu_t = \beta_2 \text{Vote}_{t-1} + \varepsilon_t.$$

The data-generating process for μ_t is a function of Vote_{t-1}. Therefore, if Econ_t is also a function of Vote_{t-1} in the data-generating process, then

$$E(\mu_t \mid \text{Econ}_t) = E(\beta_2 \text{Vote}_{t-1} + \varepsilon_t \mid \text{Econ}_t) = \beta_2 \tilde{\delta}_1,$$

where $\tilde{\delta}_1$ is the slope coefficient from the regression of Vote_{t-1} on Econ_t.

Therefore, we do not have contemporaneous exogeneity: $E(\mu_t \mid \text{Econ}_t) \neq 0$; and our estimate of β_1 will be biased . This may be resolved by explicitly modeling Vote_{t-1} as in the following data model, again estimated by OLS:

$$\text{Vote}_t = \hat{\alpha}_0 + \hat{\beta}_1 \text{Econ}_t + \hat{\beta}_2 \text{Vote}_{t-1} + \hat{\varepsilon}_t .$$

The residual will no longer contain Vote_{t-1}, so we have contemporaneous exogeneity:

$$E(\varepsilon_t \mid \text{Econ}_t) = 0.$$

However, we will not have strict exogeneity. This is because Vote_t is, by the definition of the data-generating process, a function of ε_t, and vice versa. Therefore, ε_t is a function of a lead or lag of the independent variable Vote_{t-1}. Specifically, ε_t is a function of the first lead of Vote_{t-1}:

$$E(\varepsilon_t \mid \text{Vote}_{t-1+1}) = E(\varepsilon_t \mid \text{Vote}_t) \neq 0.$$

This is a problem, but in Chapter 3, we will see that if an assumption called weak dependence can be met and we have contemporaneous exogeneity, even if we do not have strict exogeneity, we can get asymptotically unbiased estimates of the parameters in our model. Before discussing weak dependence, let us make a brief foray into a more complicated discussion of exogeneity.

We have been defining exogeneity in terms of expectations and covariances, and will continue to do so. However, Engle, Hendry, and Richard (1983) have shown that this can lead to an ambiguity. Because of the ambiguity that can occur when defining exogeneity in terms of expectations, Engle et al. (1983) define what they call weak exogeneity as the necessary exogeneity condition for an unbiased estimation of the parameters in a model. They also define two more restrictive conditions called strong exogeneity and super exogeneity, for the purposes of forecasting and intervention analysis, respectively. Weak exogeneity resolves the ambiguity problem but very much increases the necessary complexity of the discussion. Even the relatively advanced econometrics text by Greene (2002) provides a brief definition before referring the interested reader to what is described as challenging reading (p. 381).[4] An incomplete description is provided here.

For our purposes, we consider the model of y_t as a function of x_t. We begin by denoting the joint density for

y_t and x_t as $f(y_t, x_t)$; and the distribution of y_t conditional on x_t and a set of parameters **ß** as $f(y_t \mid \beta x_t)$. The conditional distribution $f(y_t \mid \beta x_t)$ is our model. Also, we denote the marginal distribution of x_t as $f(x_t \mid \boldsymbol{\theta})$, where **θ** is the set of parameters of the marginal distribution. The marginal distribution $f(x_t \mid \boldsymbol{\theta})$ is the data-generating process for x_t. Weak exogeneity of x_t for the estimation of **ß** in our model is established if we can write the joint distribution for y_t and x_t as

$$f(y_t, x_t) = f(y_t \mid \boldsymbol{\beta} x_t) f(x_t \mid \boldsymbol{\theta}) \qquad (2.5.14)$$

such that no element of **ß** is functionally related *to* $f(x_t \mid \boldsymbol{\theta})$. This means that the parameters of $f(x_t \mid \boldsymbol{\theta})$ can be safely ignored in the estimation of the parameters of $f(y_t \mid \beta x_t)$. Note that the exogeneity of x_t is established with respect to a specific set of parameters of interest in a specific model.

2.6 Weak Dependence

A stationary time series is weakly dependent if Y_t and Y_{t+h} are "almost independent" as h increases. If for a covariance stationary process $\mathrm{Corr}(y_t, y_{t+h}) \to 0$ as $h \to \infty$, we say that this covariance stationary process is weakly dependent.

Say we have a time series with the following data-generating process:

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t, \text{ with } \varepsilon_t \sim \mathrm{NID}\left(0, \sigma_\varepsilon^2\right). \qquad (2.6.1)$$

This is the autoregressive process of order 1 (AR(1)) described above. For this process to be weakly dependent, it must be the case that $|\alpha_1| < 1$. This is because for an AR(1) process, $\mathrm{Corr}(y_t, y_{t+h}) = \alpha_1^h$, which becomes small as h increases if $|\alpha_1| < 1$.[5]

This suggests one example of a time series process that is not weakly dependent. In Equation 2.6.1, set α_1 is equal to 1:

$$y_t = y_{t-1} + \varepsilon_t. \qquad (2.6.2)$$

For such a process, α_1 is no longer interpretable as the correlation between itself and the first lag of itself, and $\mathrm{Corr}(y_t, y_{t+h}) \nrightarrow 0 \text{ as } h \to \infty$ as $h \to \infty$. This particular process is called a unit root; we will revisit it at the end of this chapter and in Chapter 6.

When it is met, the assumption of weak dependence, combined with the assumption of stationarity, allows us to weaken the strict-exogeneity assumption to contemporaneous exogeneity for large-T analysis. Under these conditions, the sample variances and covariances converge on the population variances and covariances as T goes to infinity. In this case the β_1 from the regression of y_t on x_t,

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t,$$

can be estimated by

$$\hat{\beta}_1 = \frac{\widehat{Cov}(Y_t, X_t)}{\text{Var}(X_t)}. \tag{2.6.3}$$

This estimate is asymptotically unbiased and consistent as $T \rightarrow \infty$.

2.7 Trending, Periodicity, and Structural Breaks

We saw in Chapter 1 that a trending time series is one in which the mean increases or decreases over time. Recall that for a time series to be covariance stationary, its mean must be constant over time, as do its variance and (auto)covariances. Therefore, a trending time series cannot be stationary, since the mean is changing over time.

As in the public responsiveness example from Chapter 1, social science time series often have trends, and it is important to spot this form of violation of the covariance stationarity assumption. If two series are trending together, we will probably estimate a strong correlation between the two, but it is highly likely that it is a spurious result produced by the fact that both variables are a function of time. Often, both will be trending because of other, unobserved, factors.

The easiest way to spot a trending time series, in the first instance, is to plot and examine the data. When considering whether a time series trends or not, it is important to keep in mind that there are actually a number of dynamic processes that are called trending. The fact that a time series is trending is often not difficult to spot. It is a little harder to determine the type of trend. Let us consider the simplest, a deterministic, linear trend. (We will discuss other types of trends and how to distinguish between them in Chapter 6.) An example of a data-generating process with a deterministic, linear trend is the following:

$$y_t = \beta_0 + \beta_2 t + \varepsilon_t. \tag{2.7.1}$$

In this equation, t is a variable that counts up from the beginning of the time series: $t = \{1, 2, 3, \ldots, T\}$. The interpretation of such a process is that y_t increases or decreases, on average, by the magnitude of β_2 in each time period.[6] The average growth or decline of y_t may not remain constant in the long run, but if it does so for the period of observation, then this is an acceptable way to express the data-generating process. Otherwise, a more complex function of time may be used to express a nonlinear trend. A quadratic trend is one option. Splines are a more advanced option (Durlauf & Blume, 2010; Keele, 2008).

A deterministic linear trending series can be weakly dependent—the correlation between y_t and y_{t+h} approaches 0 as h increases. If a series is weakly dependent and is stationary about its trend, we call it a trend-stationary process. When we state that the series is stationary about its trend, we mean that the series is stationary once we have partialled out the trend. If the time series is trend stationary, we can control for the trend to avoid a spurious result. As stated previously, the assumption of covariance stationarity, necessary to estimate models of time series data, only needs to hold once we have controlled for the independent variables in the model. It is covariance stationarity conditioning on X_t that is the relevant assumption. Let us return to the public responsiveness example from Chapter 1.

$$R_t = \beta_0 + \beta_1 P_t + \beta_2 W_t + \varepsilon_t. \qquad (2.7.2)$$

Recall that R_t is the public's relative preference for policy spending in a given year—that is, the difference between the public's preferred level of policy spending and the level that it actually gets. P_t is the actual level of policy spending in a year. W_t represents other, exogenous effects on the public's relative preferences. This is an example of a static model, which we shall cover in Chapter 3.

In the Canadian model for social welfare spending developed by Soroka and Wlezien (2010), W_t *is a* counter variable: {1, 2, 3, …}. It is included to capture the upward trend in spending demands. The counter variable is just a trend variable, *t.* The results of regressing the public's relative spending preference on the level of government spending and the counter variable are presented in Table 2.1.

Table 2.1 Canadian Public Responsiveness Model, Social Welfare Spending

Preference	*Coefficient*	*Standard Error*	*t Statistic*	*P Value*
Program spending	–0.61	0.14	–4.46	0.001
Counter	3.95	0.50	7.89	<0.001
Constant	117.51	22.27	5.28	<0.001

NOTE: $R^2 = 0.87$, $T = 16$; T = number of time points.

The results indicate a statistically significant negative relationship between government spending and the public's relative spending preference. For each billion-dollar increase in government spending, the public's relative preference for spending decreases by 0.61 percentage points. As government spending goes up, the public's preference for greater spending decreases. This is the result predicted by the thermostatic model. What if Soroka and Wlezien had not included the trend variable? We already noted in Chapter 1 that both the public's relative preference for social welfare spending and the actual social welfare spending are trending upward over time. This means that there is a danger of estimating a positive relationship between the two variables simply because they both trend upward and not because there is a positive causal relationship.

This is indeed what we found when we estimated such a model without a trend in Chapter 1. We found a positive relationship between government social welfare spending and the public's relative preference. Controlling for the trending in the data has clear consequences for our analysis and the conclusions we reach about the relationship between government spending and the public's relative preference for spending. If we were choosing between these two models, we would choose the model with the trend. We can see from the statistical significance of the trend term in Table 2.1 that the public's relative preference does contain a trend. The estimates from the model that fails to control for this trend violate the assumptions of covariance stationarity. As we will discuss in Chapter 6, a deterministic, linear trend can also be removed by first differencing the data. However, this has consequences that generally make it an unattractive option.

Less common than trending but still common is the problem of data containing periodicity. This is a time series process with an equilibrium that cycles up and down. If it cycles up and down in time with the seasons, it is often called seasonality. Examples include quarterly data on retail sales, which tend to jump up in the fourth quarter, and crime rates, which tend to be higher in warmer months. Seasonality can be dealt with by adding a set of seasonal or monthly dummies. We can also seasonally difference the data. For example, if we have monthly data and we wish to control for periodicity at the monthly level, 12th differencing the data (subtracting the 12th lag of each variable from itself) might eliminate the periodicity resulting in stationary data. Substantively speaking, this transforms the data into the change since the same month in the previous year. This is discussed further in Chapter 6.

Trending and seasonality are instances where the mean of the time series is changing with time and is therefore not stationary. The mean could also change in a less systematic way. It could simply shift to a new level at a particular time. This shift is called a structural break or equilibrium shift. If this equilibrium shift is explained by an exogenous variable in our model, then all is well. Otherwise, an unaccounted-for equilibrium shift will result in biased estimations of our data model parameters.

In the German government approval/vote intention example in Chapter 1, we saw an example of a structural break in the variance of the time series. As noted, structural breaks can also occur in the covariance between time series. This means that the relationship between two time series variables might change at a specific point in time. Such a possibility is both something that needs to be taken into account and possibly of substantive interest. We shall discuss this further in the context of intervention analysis in Chapter 5. In Chapter 3, we will discuss other ways in which we can test and control for a potential deterministic trend, seasonality, or structural breaks.

2.8 Instability and Integration

There are reasons other than trending, seasonality, and structural breaks that will cause a time series to be nonstationary. The most important of these is instability. Stability relates to the concepts of convergence and equilibrium. A time series is stable if it has an equilibrium to which it converges in the long run.

Consider the, now familiar, AR(1) process:

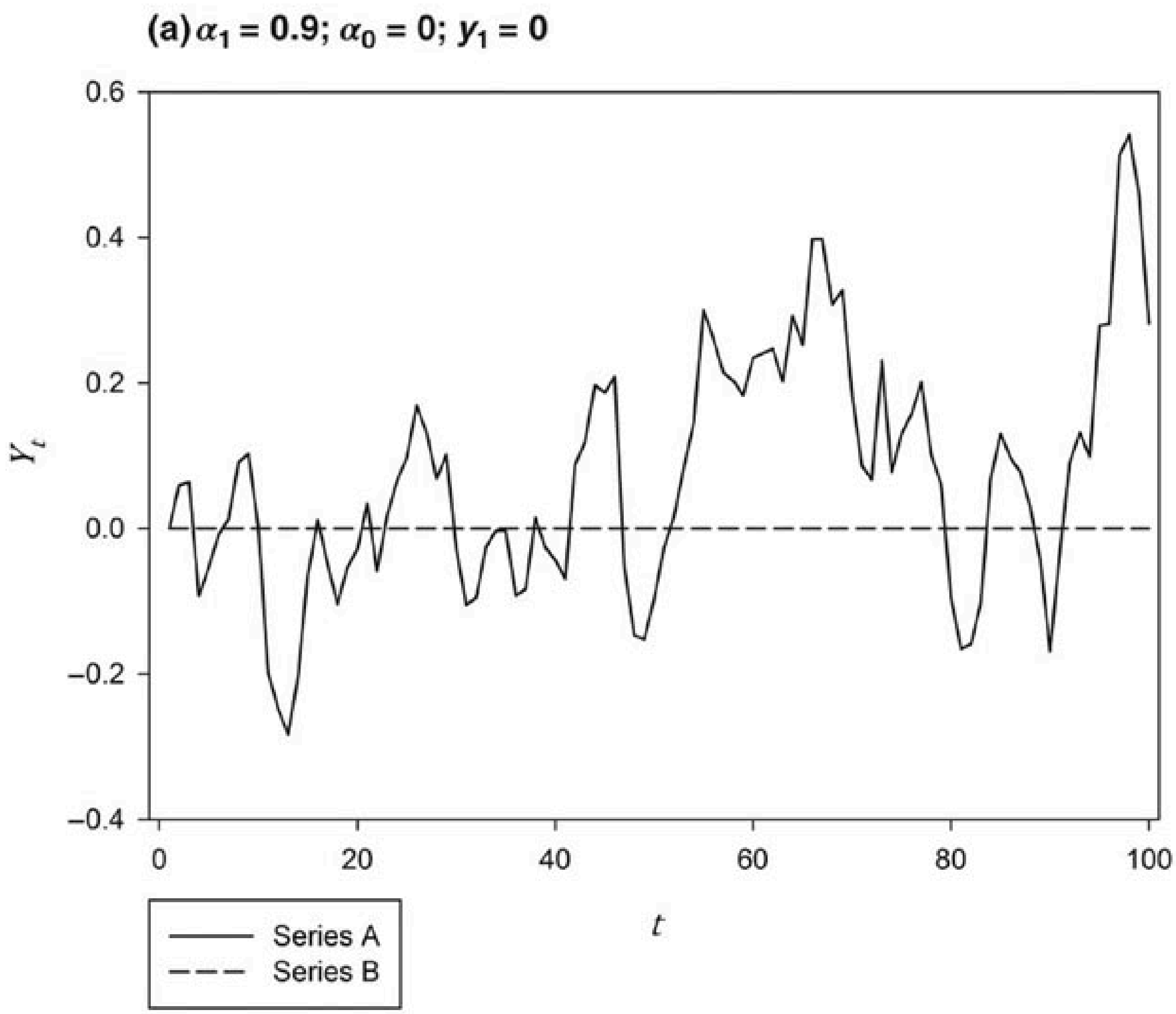
$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t. \qquad (2.8.1)$$

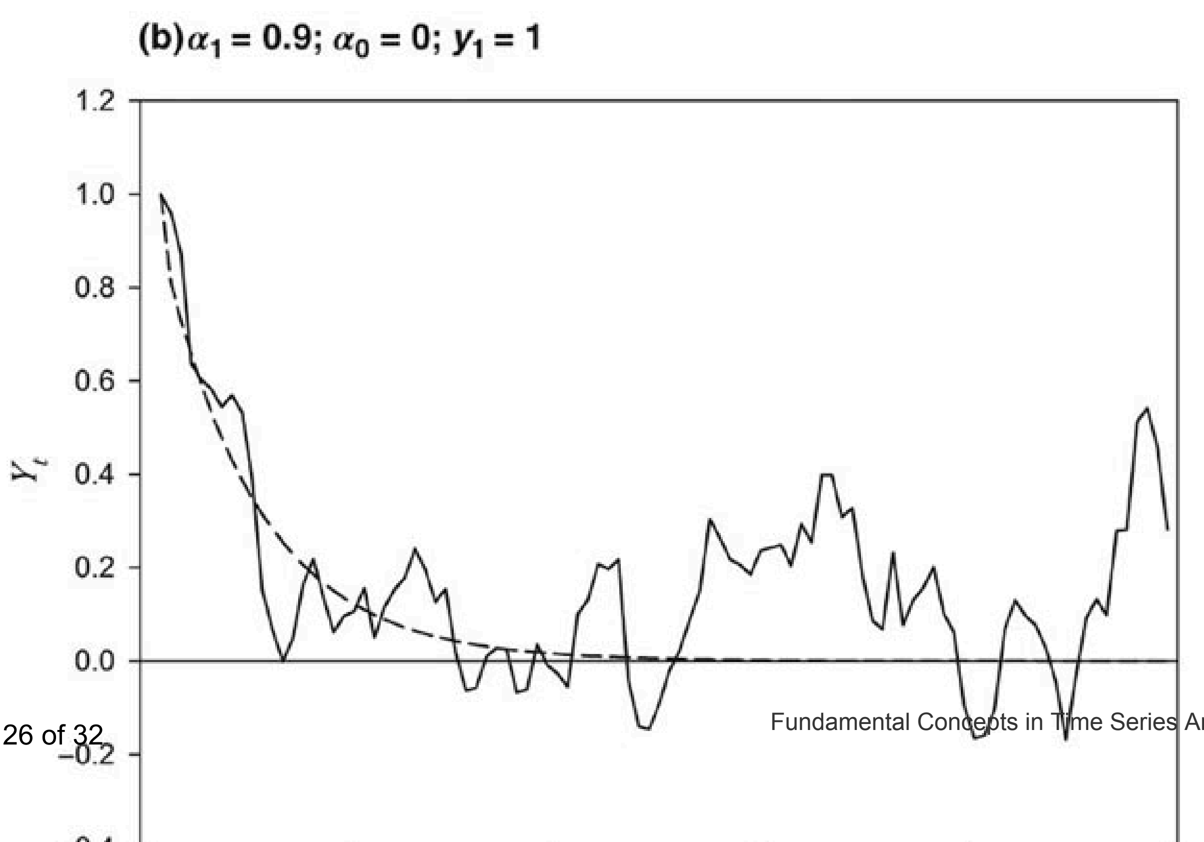
We noted that this process is weakly dependent if $|\alpha_1| < 1$. This is also a requirement for this process to be stable. As stability is a necessary condition for stationarity, the AR(1) process cannot be stationary unless $|\alpha| < 1$.

To see why $|\alpha_1| < 1$ is a necessary condition for stability, let us generate one realization of the stochastic $\{\varepsilon_t\}$ sequence and plot one realization of the data-generating process, described by Equation 2.8.1, for different values of α_1. For simplicity, let $\alpha_1 = 0$ and $\varepsilon_t \sim N(0, 1)$. Now let us consider the cases where α_1 takes on the values 0.9, 0.5, -0.5, 1.2, and -1.2. We are interested in determining in which of these cases the time series process is stable and therefore meets a necessary condition for stationarity. We do this by examining Figures 2.4a to f. In each plot, Series A is our randomly generated time series. Series B is the nonstochastic part of A: $y_t = \alpha_1 y_{t-1}$. Series A moves stochastically around the nonstochastic part—that is, Series B. The nonstochastic series gives us an impression of the value to which the time series will converge (or not) in the long run $(t \to \infty)$, when the stochastic component becomes arbitrarily small $(\varepsilon_t \to 0)$.

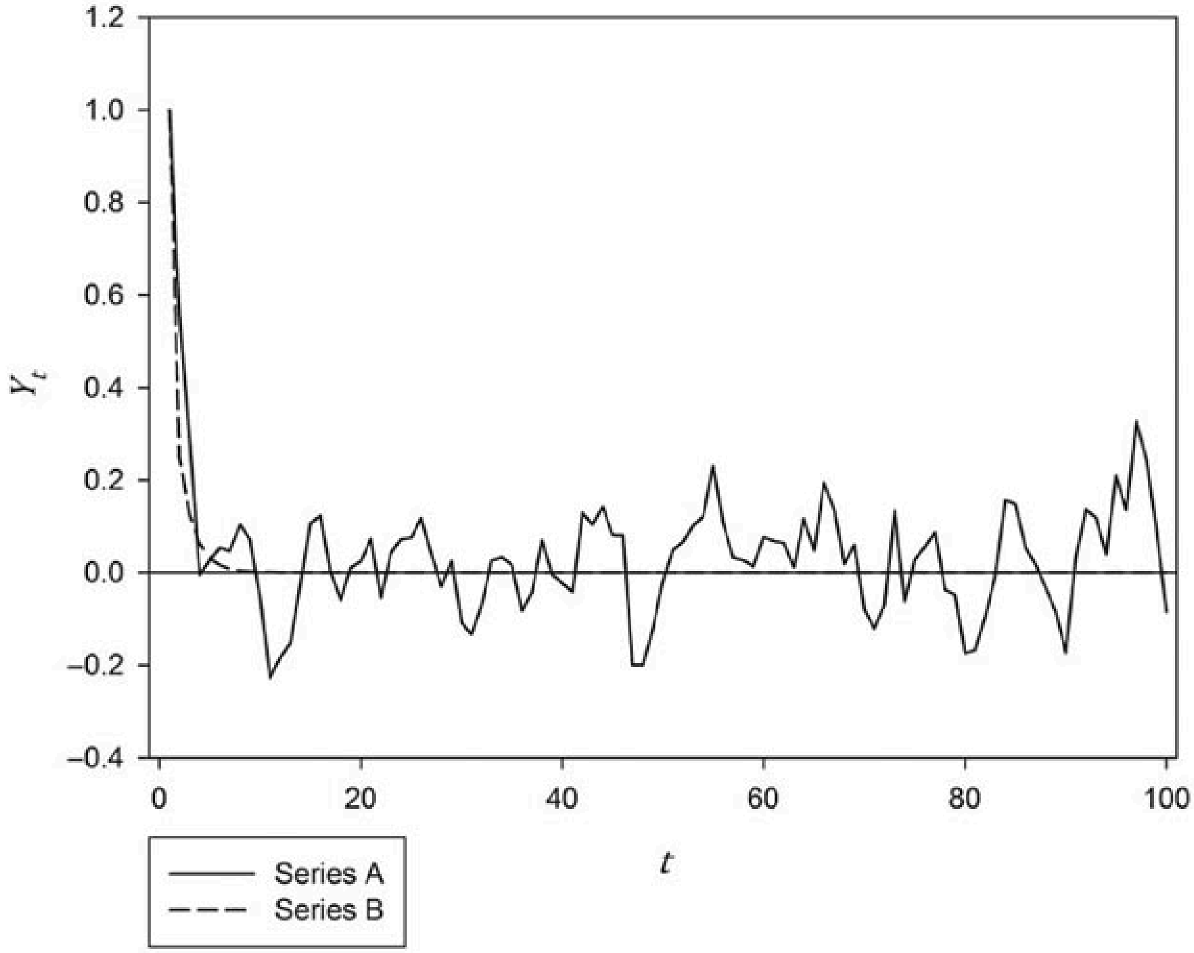
Figure 2.4a is the plot of $y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t$, where $\alpha_0 = 0$ and $\alpha_1 = 0.9$. We have also set the initial value for y_t at $y_{t=1} = 0$. In this plot, you can see from the nonstochastic part of the series that the time series process converges to (remains in) an equilibrium of 0. The equilibrium value also happens to be the initial value, so the series is not so much converging on its equilibrium as it is remaining at its equilibrium. An example of this type of convergent sequence might be monthly approval ratings for the U.S. president (although α_0 would probably not equal 0 and all values would be positive).

Figure 2.4 Stable and Unstable Time Series Processes

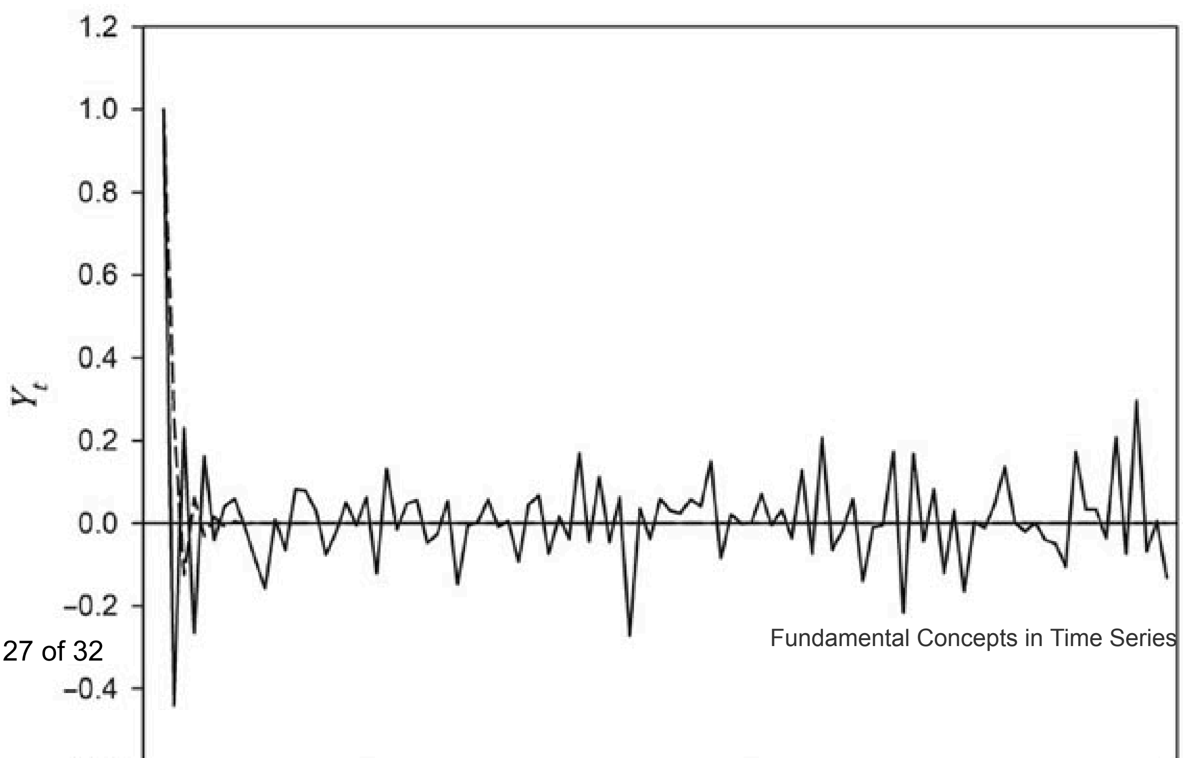

(a) $\alpha_1 = 0.9$; $\alpha_0 = 0$; $y_1 = 0$
0.6
0.4
0.2
0.0
−0.2
−0.4
Y_t
0
20
40
60
80
100
t
Series A
Series B


(b) $\alpha_1 = 0.9$; $\alpha_0 = 0$; $y_1 = 1$
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
Y_t

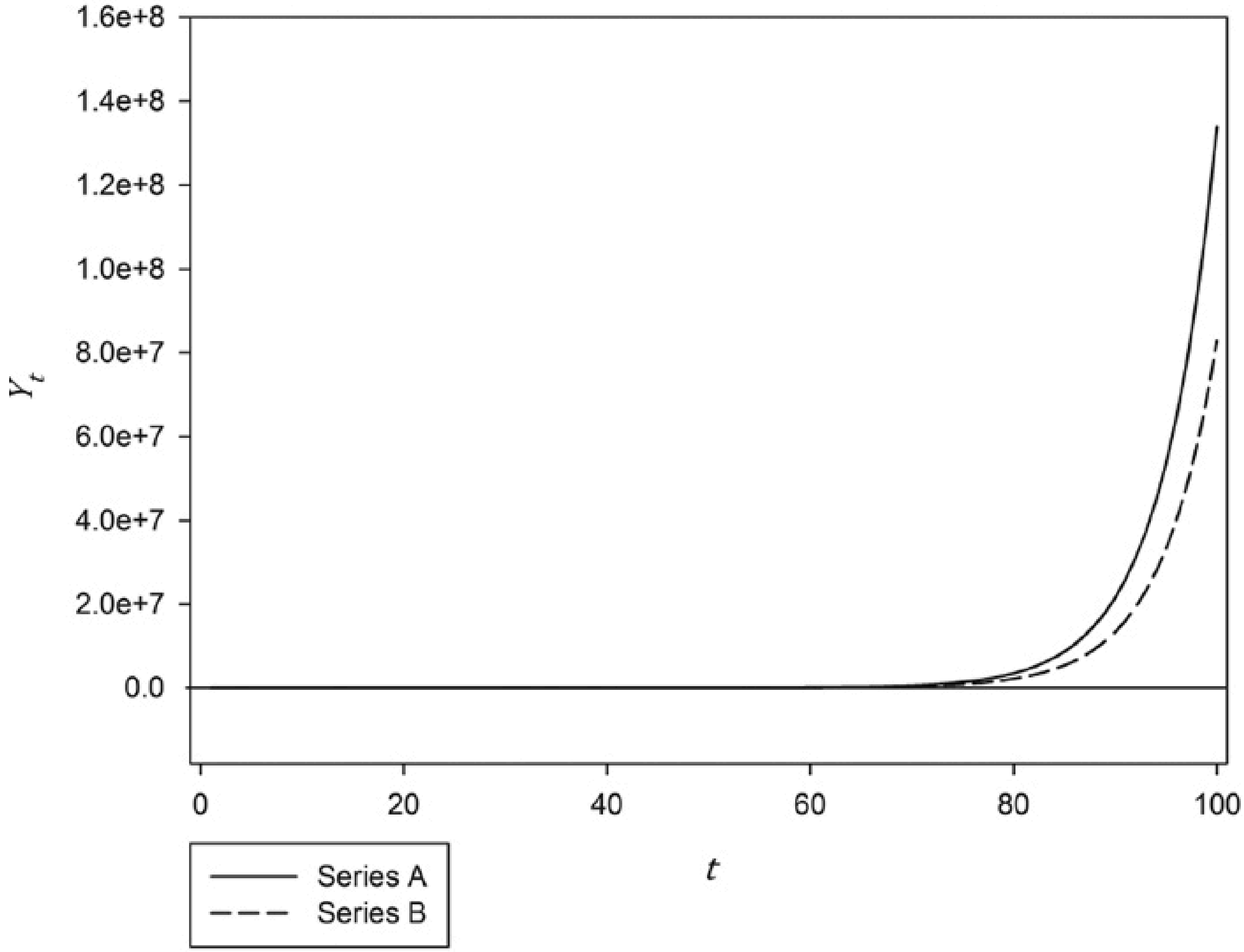
(c) $\alpha_1 = 0.5; \alpha_0 = 0; y_1 = 1$



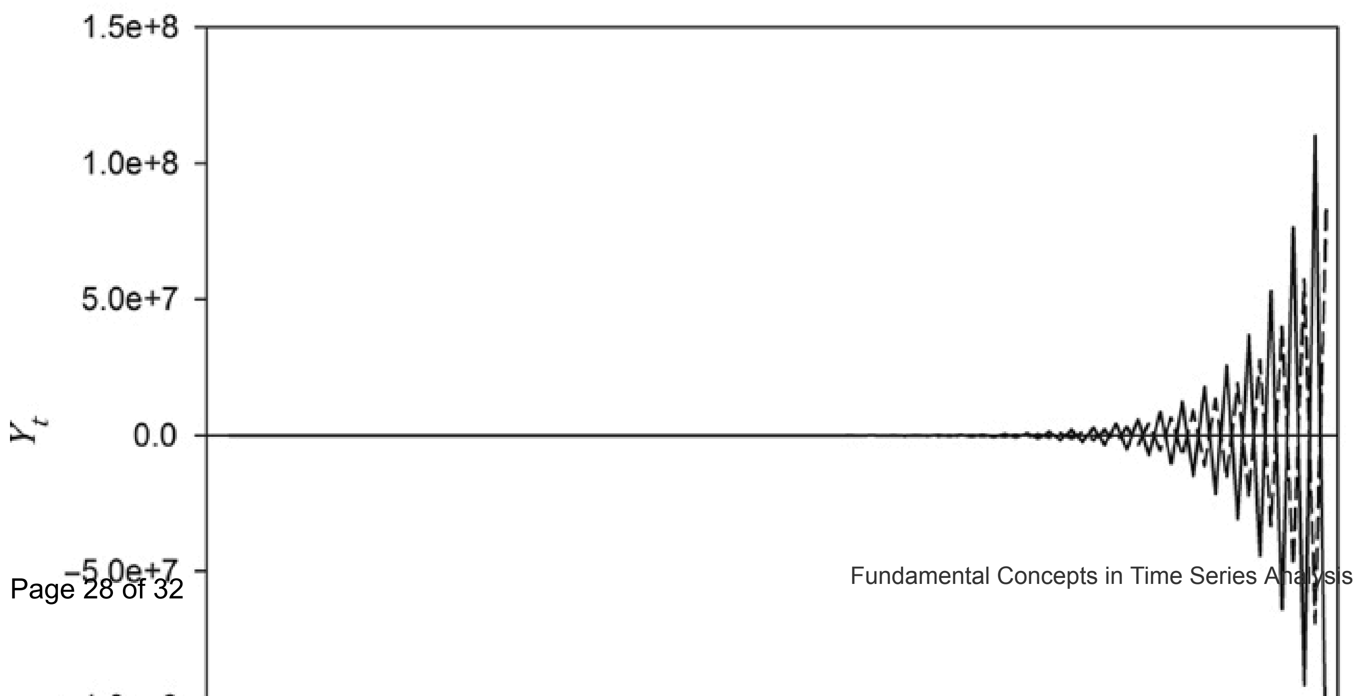
(d) $\alpha_1 = -0.5; \alpha_0 = 0; y_1 = 1$



(e) $\alpha_1 = 1.2$; $\alpha_0 = 0$; $y_1 = 1$



(f) $\alpha_1 = -1.2$; $\alpha_0 = 0$; $y_1 = 1$



We can get a clearer picture of convergence by choosing an initial value that is out of equilibrium, for example, $y_{t=1} = 1$. This is plotted in Figure 2.4b. The monthly approval of the U.S. president could also possibly be an example of this type of sequence. Presidents often start their term with a honeymoon period in which their support is atypically high—that is, a president's approval often starts out of equilibrium.

The comparison of Figures 2.4a and b raises an interesting point. The time series processes plotted in these two figures are the same, except that we began the second one out of equilibrium. Both of these processes are stable in that they converge to an equilibrium in the long run. However, the series that started out of equilibrium is not stationary. As is clear from the plot, it trends downward until it reaches its equilibrium. Having done so, the process continues as a stationary process.[7] If we observed this process as it equilibrated, it would not be stationary. If we observed this process at a later point in time, long after its initial out-of-equilibrium start, it would be stationary. We will discuss the necessary conditions for stationarity further in Chapters 5 and 6. For now it is important to note that stability is a necessary but not sufficient condition for stationarity.

Now consider Figures 2.4c and d, in which α_1 is 0.5 and -0.5, respectively. Again, we start the time series processes out of equilibrium: $y_{t=1} = 1$. Both of these series are stable in that they converge to their equilibrium in the long run, but the second series is a stable, oscillating process. It converges on its equilibrium through a decaying, oscillating path.

Each of the series with $|\alpha_1| < 1$ converged to an equilibrium. The equilibrium of such series can be calculated as

$$E\left(y_t\right) = \frac{\alpha_0}{\left(1-\alpha_1\right)}. \qquad (2.8.2)$$

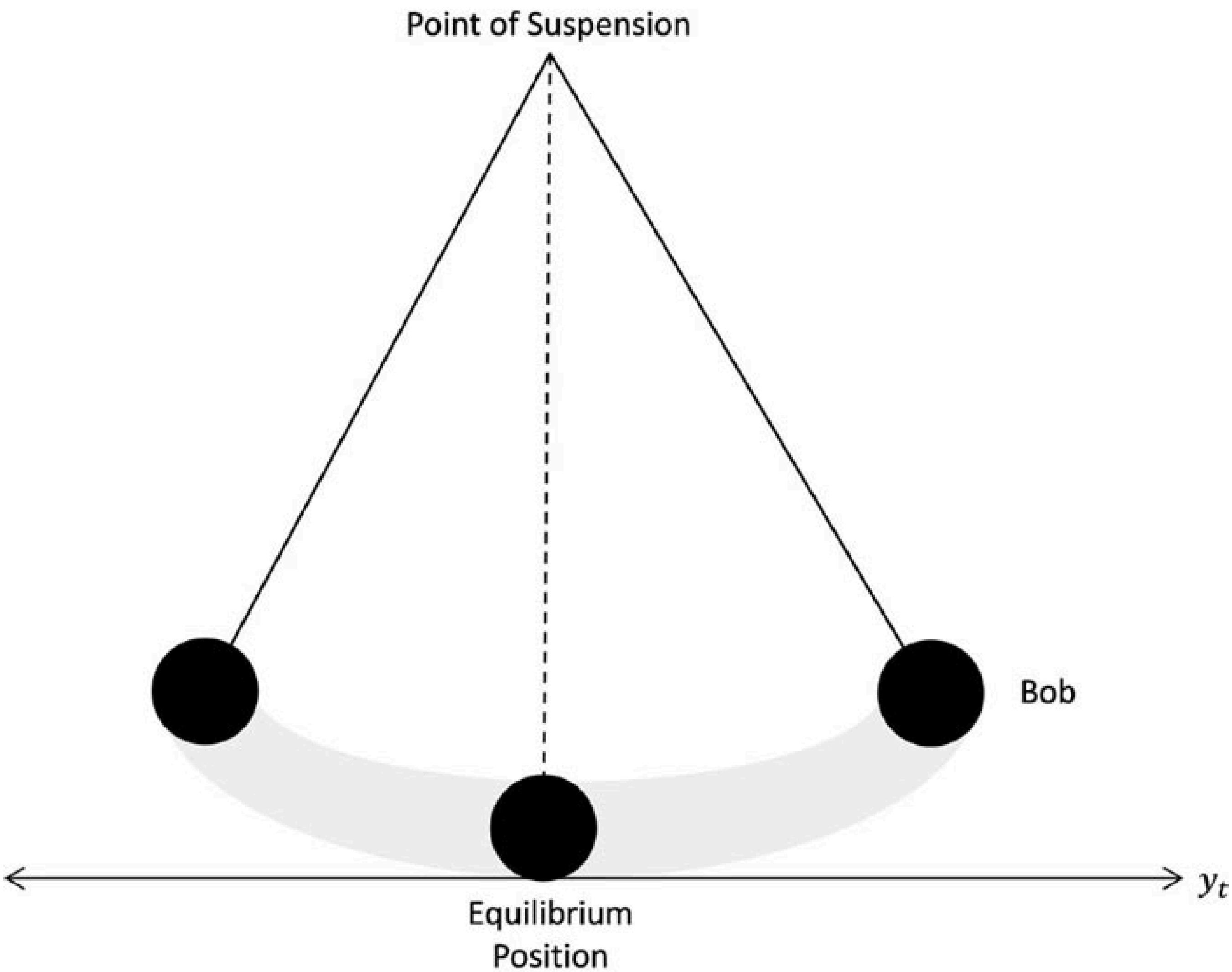
For example, when $\alpha_0 = 0$ and $\alpha_1 = 0.9$:

$$E\left(y_t\right) = \frac{\alpha_0}{\left(1-\alpha_1\right)} = \frac{0}{\left(1-0.9\right)} = 0.$$

The convergence to equilibrium is different for each of the processes considered so far. The time series process with $\alpha_1 = 0.5$ converges quicker than for $\alpha_1 = 0.9$, and $\alpha_1 = -0.5$ oscillates around its equilibrium before settling down. An example of a convergent series with $0 < \alpha_1 < 1$ is changes in the yearly amount of alcohol consumed in Denmark (liters per capita): $y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t$, where y_t is the yearly change in the log of the amount of alcohol consumed, $\hat{\alpha}_1 = 0.806$, $SE = 0.15$. (Bentzen & Smith, 2004).[8]

Convergent series with $\alpha_1 < 0$ are far less common, particularly in the social sciences. However, everyone is familiar with a pendulum (Figure 2.5). The horizontal distance that the bob of the pendulum travels is a classic example of a stable system that oscillates into its equilibrium after it has been started out of equilibrium. If disturbed, the pendulum will swing left and right until gravity returns it to its original position. Gravity moves the pendulum back to equilibrium through an oscillation.

Figure 2.5 Pendulum



Convergent sequences reflect stable systems. Let us now look at systems that are not stable.

Specifically, let us consider the same process: $y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t$, where $y_0 = 1$ and $\alpha_0 = 0$ but $|\alpha_1| > 1$. Figures 2.4e and f correspond to $\alpha_1 = 1.2$ and -1.2, respectively. The time series process in Figure 2.4e is an explosive process. The time series process in Figure 2.4f is an oscillatory explosive process. As these processes do not return to any equilibrium, they are not stable, and therefore they are not stationary.

When discussing weak dependence, we considered the possibility that $\alpha_1 = 1$. This produces a time series process with very special properties. As with the case where $|\alpha_1| > 1$, the time series does not converge to an equilibrium and is not stationary. However, the process is not explosive. This is an example of a time series process called a unit root process.

A unit root process is an example of a category of time series processes we call integrated. An integrated time series process is one that is not stationary but can be made stationary by differencing one or more times. A unit root process can be made stationary by differencing once as follows:

$$y_t = \alpha_0 + y_{t-1} + \varepsilon_t. \quad (2.8.3)$$

Subtract y_{t-1} (the lag *of* y_t) from both sides:

$$y_t - y_{t-1} = \alpha_0 + y_{t-1} - y_{t-1} + \varepsilon_t,$$

$$\Delta y_t = \alpha_0 + \varepsilon_t. \quad (2.8.4)$$

As $\varepsilon_t \sim \text{NID}(0,\sigma^2_\varepsilon)$, Δy_t is a constant plus a white noise term. This process is stable and meets the conditions of covariance stationarity. The substantive interpretation of this process is that it has infinite memory. Consider the case when $\alpha_0 = 0$. When something intervenes to change y_t, the process, in the absence of any further intervention, remains at its new value. In other words, the effect of the intervention remains forever. We will consider this and other integrated time series processes in detail in Chapter 6.

Summary

In this chapter, you have been introduced to fundamental concepts in time series data and analysis. These concepts are the basis of the analytical techniques discussed in the chapters to follow. In the next chapter, you will see how each of these fundamental time series concepts are applied as we examine how time series data can be modelled using two of the most basic time series models: static models and finite distributed lag models.

[1]We could also denote this as $\overline{Y}_t$ but we are assuming that this is the same as $\overline{y}_t$ for all *t*.

[2] This parallels the idea in cross-sectional data analysis that the exogeneity assumption can be weakened when *N* is large to an assumption of no correlation between the error term and the independent variables. This is a weaker assumption than that of the expected value of the error term not being a function of the explanatory variables.

[3] Without going into the mathematics, omitting a variable from the data model may also have consequences for the estimated variance of the OLS estimator of β_1. If we use the usual estimate of the variance of an OLS estimator and compare the variance of $\tilde{\beta}_1$ from our misspecified data model (Equation 2.5.8) with the variance of $\tilde{\beta}_1$ from the correctly specified data model,

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \varepsilon.$$

we will find that the estimated variance of $\tilde{\beta}_1$ (conditioning on x_1 and x_2) is less than that of $\hat{\beta}_1$. This is true unless X_1 and X_2 are uncorrelated, in which case the variances are the same. The estimated variance of $\tilde{\beta}_1$ is too small. The conditioning on x_1 and x_2 means that this result may not hold in practice because of the potentially reduced estimate of the error variance due to the inclusion of x_2. As the sample size grows, the variance of each estimator shrinks to 0, making the difference in variance less important. However, remember that the misspecified model is biased unless $\beta_2 = 0$ or $\mathrm{Corr}(X_1, X_2) = 0$.

[4] In addition to Engle et al. (1983), these advance readings include Zellner (1979), Sims (1977), and Granger (1969).

[5] In fact, Equation 2.6.1 is not an autoregressive process unless $|\alpha_1| < 1$.

[6] Deviations from this average increase/decrease are modeled as random.

[7] Such a process might be called asymptotically stationary.

[8] The Bentzen and Smith model also includes the price of alcohol.

http://dx.doi.org/10.4135/9781483390857.n2