Chapter 2: Univariate Time Series Models

January 16, 2022

Contents

List of Figures

Introduction to Part I

In Part I, we introduce the key features of data that drive the approaches to time series analysis covered in the remainder of the book. How can we characterize the dynamics of our data? Chapter 2 presents the archetypal patterns of persistence that can be observed in a single time series. The chief message is that univariate description reveals a great deal that informs any modeling endeavor. We introduce the distinction between stationary and non-stationary time series and explain its importance. Modeling choices in applied analyses are determined by the conclusions we draw about the univariate features of our data: Whether successive observations are dependent, whether the series can be classified as stationary or non-stationary, and how the persistance in data can be characterized.

Next, Chapter 3 considers the "unit root question." We describe formal tests to arbitrate between stationary and non-stationary processes and offer a strategy for doing so in applied work. A hypothesis about a long-run relationship is at the heart of almost every time series analysis. Are two or more variables related over time and, if so, how are they related? For two variables to be related, there must be an equilibrium at the heart of the relationship. However, before the relationship can be identified, the analyst needs to know what type of equilibrium they are looking for.

A stationary series has a predictable long-run mean and variance. A hypothesis test about a specific relationship with a stationary regressand is a test for the existence of a conditional stationary equilibrium between the variables against the null that the regressand has an unconditional stationary equilibrium. A non-stationary series does not have a predictable long-run mean and variance on its own. Such a series can only have an equilibrium with respect to another non-stationary series. The tools required to test for the existence of a conditional stationary equilibrium are different from the tools used to test for a co-integrating equilibrium, so the analyst must classify the series as stationary or non-stationary in advance to apply conventional hypothesis testing procedures. There is vast literature on unit root and stationarity tests which speaks to the significance of the unit root question and the problems that it creates. Chapter 3 introduces a handful of the most prominent tests and discusses the significance of classification as a prerequisite to applied analysis.

Chapter 4 tackles the exogeneity question. It has become standard dictum in applied statistics to note that correlation does not imply causation. Yet, regression models presume a causal relationship; the regressors are part of the conditional model for the regressand. This can be a difficult assumption to validate with cross-sectional data and the issue of exogeneity becomes more consequential in the time series context. Entire schools of thought and approaches to data analysis stem from the assumptions the analyst makes about exogeneity. If the regressand can be presumed exogenous to the regressors, the approaches in Part II of the text are appropriate. If an exogenous relationship cannot be assumed, or if theory dictates that the relationship is better characterized as an endogenous system, the approaches presented in Part III of the book should be used. Chapter 4 will explain the intuition underlying different assumptions about exogeneity and endogeneity, define these concepts, detail how they relate to the types of information that we can learn from any statistical analysis, and explain how to assess whether or not weak exogeneity is a reasonable assumption.

1 Introduction

This book is a practical guide to time series analysis using ordinary least squares (OLS) regression. The statistical foundations of OLS regression are built on independent random variables but time series observations are seldom, if ever, independent. Observations ordered sequentially in time exhibit time dependence. Observations close together in time tend to be more similar, while those more distant from each other are less similar. Time dependence can take many forms, some of which present significant problems for regression. Time series can exhibit trends, structural breaks, and other features that must be addressed for statistical inference to be meaningful. These different features of the data require different strategies for estimation and inference. As such, the first stage of any time series analysis is to identify the underlying dynamics of each individual series.

Each time series you work with can be thought of as one *realization* of a *stochastic process* representing the evolution of a set of random variables indexed over time. Each observation of your time series $\{y_t\}_{t=1}^{T} = \{y_1, y_2, ..., y_T\}$ is produced by an underlying ranodm variable $\{Y_t\}_{t=-\infty}^{\infty} = \{\ldots Y_1, Y_2, \ldots, Y_T\}$; the value y_1 is one potential outcome of the random variable Y_1, the value of y_2 is one potential outcome of the random variable Y_2, and so on. The realization you work with is the sequence of outcomes produced by the underlying stochastic process.

Consider the presidential approval time series we introduced in Chapter 1. Presidential approval at each point in time is a random variable. At any point in time it is possible that no one approves of the job the president is doing. It is also possible that everyone approves. Neither of these outcomes is likely. It is more likely that the proportion of Americans who approve of the president lies somewhere between zero and one, and even more likely that this value falls between 0.40 and 0.60. The particular constellation of economic outcomes, extrordinary events (e.g., terrorist attacks and pandemics), partisan politics, and other factors determine the proportion of adults that approve of the president at any point in time, y_t. The same is true at every other point in time. The possible outcomes always range from zero to one and some values are always more likely to be observed than others.

These random variables (Y_1, Y_2, etc.) at successive time points are not independent of one another; approval this month tends to be very similar to approval in the previous month. The model that describes the evolution of a particular sequence of random variables can take many forms. These forms are the different types of stochastic processes. You will use the the realizations of the stochastic process that produced the time series you observed (your time series variables), to describe the dynamic properties of your individual time series. This determine the type of regression models you will estimate and how you will draw inference from those models.

This chapter introduces the different types of stochastic processes you are likely to encounter and the vocabulary you need to describe them. There are two broad classes: stationary processes and non-stationary processes. There are different varieties within each class. White noise, autoregressive (AR), moving average (MA), and hybrid autoregressive-moving average (ARMA) processes are the most common stationary processes. Non-stationary processes include deterministic trends and stochastic trends. A deterministic trend time series has a predictable long-run pattern. The patterns we observe in time series data produced by stochastic trends depend on whether a process is a pure stochastic trend (a random walk) or whether there are other deterministic features of the underlying process (random walk with with drift or random walk with trend and drift).

The concepts discussed in this chapter represent the raw materials you will use to build your time series models. Classifying your series as stationary or non-stationary will determine which regression models are appropriate for your data and what types of relationships are possible. The exposition in this chapter will introduce a lot of new terms unique to time series analysis but if you develop an intuition for the underlying concepts and the patterns they represent, the nomenclature will be less challenging.

2 Plotting Your Time Series

If identifying the dynamics of one's time series is the first stage of a time series analysis, plotting these data is one's first task. As the name suggests, a time series plot is a plot of one or more time series variables over time. The x-axis displays the time dimension of the variable. The y-axis is the metric for the variable. Whether you have realized it or not, you see time series plots all the time. Every time you see the daily values of a stock price or stock index on the news, every time you see a graph of presidential approval ratings in the run up to an election, and every time you see a heartbeat monitor in network television hospital drama; you are looking at a time series plot. This is the most common way to depict time series data because the sequential depiction of a series contains so much useful information.

Some of the information in a time series plot is descriptive. The x-axis of the plot highlights two critical features of the data. The sampling window is denoted by the beginning and end of the time series, the first and last values on the x-axis. The sampling interval is the periodicity of the axis. For most data, the index on the x-axis will be given in units of time: seconds, hours, days, months, or years. In some cases, the x-axis only denotes an order of the observations. This may reveal that the data do not have a meaningful sampling interval, as with some of the simulated data we depict below. The minimum and maximum values of the realization will also be easy to identify in the plot. Where these values appear can signal important features of the data.

Interpreting a time series plot can be thought of as a rudimentary form of time series analysis. While one will not be able to draw any formal conclusions about a series, one can often rule out certain conclusions about the data. If the observations are increasing over time, the data exhibit a positive trend. If the observations are decreasing over time, they are exhibiting negative trend. If one observes either pattern, one can reasonably assume the data cannot be classified as mean stationary. Seasonality will appear in the data as regular patterns. If there appear to be periods of higher and lower variability in the data over time, it suggests heteroskedasticity. If the data are slowly moving up and down, rather than crossing a mean in rapid succession, one should expect that there are dynamics in the data that need to be modeled. One cannot know the exact properties of these dynamics from the plot alone but it will often be clear, almost immediately, whether the data are dynamic. The problem with this rudimentary analysis is that some patterns are consistent with multiple processes, so one cannot rely on observation alone. There are formal tests. Like many test in statistics, these formal tests serve to verify what you see in simple plots. We highlight many of these tests, along with the patterns they reveal, below and in Chapter 3.

3 Stationary Time Series Processes

Stationary stochastic processes represent a particularly important class of time series. Time series that are stationary can be analyzed in the traditional regression framework. What makes a time series stationary? A stochastic processes is said to be *covariance stationary* if its mean and variance are independent of time and if its covariances are finite and depend only on the number of periods separating observations. This implies these features of the data do not depend on the time period in which the data are observed.[1,2] Formally, a process y_t is covariance stationary if for all time periods, t, and all lags, s and j:

$$E(y_t) = E(y_{t-s}) = \mu_y \tag{1}$$
$$V(y_t) = V(y_{t-s}) = \sigma_y^2 = \gamma_0 \tag{2}$$
$$cov(y_t, y_{t-s}) = cov(y_{t-j}, y_{t-j-s}) = \gamma_s. \tag{3}$$

Intuitive, right? Perhaps not. Let's take a moment to digest what is going on in these equations. Equation 1 is the equation for the mean, μ_y. Like any random variable, the expected value ($E(\ )$) of the variable, absent any additional information, is the mean. Equation 1 says that the expected value of any observation at a particular point in time t is the same as the expected value of any observation observed at a previous period $t - s$; the mean μ_y does not change over time. In other words, the series gravitates to the same mean value over the full sampling window.

Equation 2 is the equation for the variance, σ_y^2 or γ_0. Like the equation for the mean, we can see that the variance of a variable at t ($V(y_t)$) should be the same as it was s periods before t ($V(y_{t-s})$). The variance is the same for all time periods, σ_y^2. Said differently, the series is homoskedastic over time. There is one (*homo*) dispersion or spread of the variable (*skedasticity*) over the full sampling window.

The final equation, Equation 3, is the equation for the covariance. Equation 3 says that the covariance between observations observed at t and $t - s$, two observations that are s periods apart, is the same as the covariance between any two other observations at $t - j$ and $t - j - s$, two observations that are also s periods apart. As long as the distance between the observations is the same, the covariance coefficient γ_s that describes that relationship should be the same; the covariance is the same over the full sampling window.

[1] A strictly stationary process is one in which the observations are jointly identically distributed. Strictly stationary time series are covariance stationary but the reverse need not be true.

[2] Why must these conditions hold? First, if we are to describe the stochastic process that generated the data based on the observed values of the data, the expected value of y_1 (and all other realizations) must be approximated by the mean of all realizations of the underlying stochastic process that generated the data. This condition will only be met if the mean is constant over time and we observe the process over a sufficiently long period of time. Second, the variance of y_1 (and all other realizations) must likewise be approximated by the variance of the realizations of the process in our sample. For this condition to hold the variance must be constant over time and we must observe the series for a sufficiently long time period. Finally, the correlation between y_1 and y_2, y_2 and y_3 and so on must be approximated by the lag one autocorrelation in the sample. The lag 1 autocorrelation and all other autocorrelations must depend only on the time period separating the observations and not on the time period itself for this conditions to be met. If all of these conditions are met, the mean, variance, and covariances of the process do not depend on the time period in which the data are observed. These assumptions are the equivalent of assuming the process is covariance stationary.

So, to review, a *covariance stationary* variable meets all three of the criteria outlined in Equations 1, 2, and 3. The mean, variance, and covariance do not change over time. How the variable behaves and how observations are related to one another doe not depend on whether we are looking at observations at the beginning of the time window, the end, or anywhere in between.

With these criteria in hand, we can define another that can be used to describe stationary stochastic variables, autocorrelation. An autocorrelation coefficient, or an autocorrelation, describes the dependence between observations of the same time series observed at different points in time. Consider two observations from the same time series, one observed at t and another observed at $t - s$. The autocorrelation coefficient (ρ_s) is the covariance between y_t and y_{t-s} divided by the variance of y_t:

$$\rho_s = \gamma_s / \gamma_0. \tag{4}$$

There are two points about Equation 4 that warrant reflection before moving forward. First, if you have taken a course that includes regression, you should remember that a covariance between two variables divided by the variance of one of those variables is the formula for a bivariate slope coefficient, $\hat{\beta}_x = cov(y_i, x_i)/var(x_i)$. An autocorrelation coefficient is like a slope coefficient but the variables involved are two realizations of the same time series, one observed at t and one observed at $t - s$. Imagine the following exercise. You have a data set that contains one variable, a time series you are interested in modeling. You create another variable that is the lag of that variable. These variables are almost exactly the same but the lag has one fewer observation. If the original time series has 100 observations, the lag of the time series only has 99. If you regress the observed time series on the lag of the time series, the slope coefficient is the autocorrelation coefficient, ρ_1. If you created different variables, the second or third lags of your time series, you could calculate different autocorrelation coefficients, ρ_2 and ρ_3. So while ρ_s, γ_s, and γ_0 may not be the greek letters you are used to seeing, we are actually treading in familiar territory.

Second, the formula presented in Equation 4 is comprised of components defined in Equations 2 and 3. Thus, the properties described above also apply to the autocorrelation coefficient. If the variance and covariances of a time series are independent of time, the autocorrelations are also independent of time.

An *autocorrelation function* (ACF) is a graphical summary of autocorrelation coefficients calculated over multiple time periods. Autocorrelation coefficients (ρ_s) are calculated for s lags and then the coefficients are plotted against s, with the magnitude of the autocorrelation coefficients depicted along the y-axis and the lags depicted on the x-axis.[3] This depiction of the autocorrelations gives a compact visual summary of the dependence observed in the data. Autocorrelations for stationary series vary between negative one and positive one. The larger the magnitude of autocorrelation, the more similar the variable observed at t and $t - s$. As the value of s increases along the x-axis of the plot, the magnitudes of the coefficients will decline because the variables used to calculate the autocorrelations are less and less similar. Negative autocorrelations indicate that successive values have opposite signs. The magnitudes will gradually decline but the coefficients will appear to oscillate from positive to negative.

[3] It is good practice is to specify a s to encompass all lag lengths at which we might expect correlations. For monthly data, we would want at least 12 and preferably 24. For quarterly data, at least four lags and preferably eight lags or more. This would enable us to observe repeated patterns in autocorrelations due to the periodicity of the data that we will discuss below.

We can also calculate the *partial autocorrelations*, ϕ_s, between observations separated by s lags. This gives us the correlation between y_t and y_{t-s} after controlling for the intervening lags. For example, the lag three partial autocorrelation is given by

$$\phi_3 = \frac{cov(y_t, y_{t-3}|y_{t-2}, y_{t-1})}{\sqrt{var(y_t|y_{t-1}, y_{t-2})var(y_{t-3}|y_{t-1}, y_{t-2})}} \tag{5}$$

If you can see the correspondence between an autocorrelation coefficient and a bivariate slope coefficient, you should also recognize the correspondence between a partial autocorrelation coefficient and a partial slope coefficient estimated in a multiple regression model. While the correspondence is not exact, the analogy provides insight into what a partial autocorrelation coefficient reflects. The variances and covariances are calculated conditional on, or net, the other dependencies observed in the data. Like the autocorrelation coefficient, the properties summarized in Equations 2 and 3 also apply to the partial autocorrelation coefficients; the partial autocorrelations will also be independent of time if the series is stationary and the values will range between negative one and positive one. Also like the the autocorrelation coefficients, we can plot s partial autocorrelation coefficients against s to produce a summary of the dependence in the data. A plot of partial autocorrelations is called a *partial autocorrelation function* (PACF).

As we will see later, different types of stationary processes have characteristic patterns of theoretical autocorrelations and partial autocorrelations. The empirical ACF, in conjunction with the PACF, provides clues to help us identify the underlying stochastic process that generated the data. For a detailed treatment of the standard patterns observed in the ACF and PACF see Enders (2015); especially Table 2.1, which summarizes this information for each of the common stationary processes we discuss below.

Stationary time series processes are often referred to as integrated of order zero or $I(0)$ processes. This notation indicates that a process need not be differenced in order to be stationary. A number of processes are covariance stationary, including white noise processes, autoregressive (denoted AR(p)) processes, moving average (denoted MA(q)) processes, and hybrid processes that are produced by combining these different elements (denoted ARMA(p,q) or ARIMA($p, 0, q$)). Whenever it avoids confusion, we will use the simpler notation. We discuss these different stationary processes in turn.

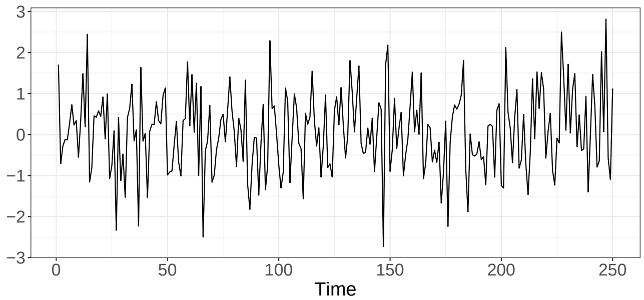
3.1 White Noise Process

The white noise process is the simplest stationary time series process. Its defining characteristic is the lack of any predictable pattern over time. Formally a time series process $\{\varepsilon_t\}$ is *white noise* if for all observations t, and all lags s and j:

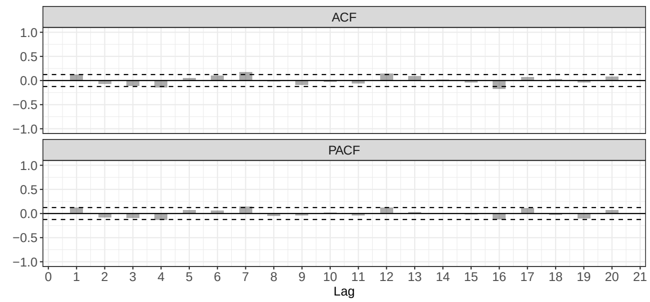
$$E(\varepsilon_t) = E(\varepsilon_{t-s}) = 0 \tag{6}$$
$$\gamma_0 = E(\varepsilon_t^2) = E(\varepsilon_{t-s}^2) = \sigma_\varepsilon^2 \tag{7}$$
$$\gamma_s = E(\varepsilon_t \varepsilon_{t-s}) = E(\varepsilon_{t-j}, \varepsilon_{t-j-s}) = 0. \tag{8}$$

These equations say that the sequence of observations ε_t has a mean of zero, a constant variance, and that all the covariances of the series are equal to zero; white noise is a sequence of random

Figure 1: White Noise: $y_t = \varepsilon_t$



(a) Time Series Plot



(b) Autocorrelation and Partial Autocorrelation Function

shocks.[4] In what follows, $\{\varepsilon_t\}$ will always denote a white noise process. The process y_t is thus a white noise process if:

$$y_t = \varepsilon_t. \tag{9}$$

A white noise process is the simplest stationary time series process. If you compare the formal definition outlined in Equations 6, 7, and 8 to the criteria for a stationary process outlined in Equations 1, 2, and 3 you can see that all the criteria are met. The mean of a white noise series is constant over the sampling window (0), the variance is homoskedastic over time (σ^2_ε), and the covariances for all the time points are also constant over time because they are all zero. Because a white noise process is stationary and is not characterized by any AR or MA dependence, these processes are often denoted ARIMA$(0, 0, 0)$ processes.

Figure 1a shows a time series plot of a (simulated) white noise process. The series is moving randomly about a mean of zero. Figure 1b shows the ACF and PACF of this series. The dashed lines represent 95% confidence bounds for each autocorrelation and partial autocorrelation. Because the series is random, there are no patterns in the ACF or PACF. Consistent with random chance, the lag 16 correlations are just beyond the confidence interval, but all remaining correlations are not significantly different from zero.

If white noise processes characterized the time series data with which we work, there would be no need for an entire text on time series analysis. But white noise processes seldom, if ever, describe the time series social scientists work with. Instead social science time series tend to be persistent, to have relatively similar values at consecutive points in time.

3.2 Autoregressive Process

It is difficult to imagine a social science time series that is purely random. Most of the time series data we encounter exhibit dependence, or inertia. Observations close in time tend to be similar. Consider the changes in the annual budget of a federal agency. Research on budget dynamics

[4] See, for example, Hamilton (1994) or Enders (2015) for a detailed treatment of different forms of white noise processes and for proof that these processes have a mean of zero, constant variance, and covariances equal to zero for all lags.

demonstrates that most budgets are built by starting with the previous year's figures and adjusting from there. Or consider the Index of Consumer Sentiment (ICS) time series. The value of the ICS at any given time point is a function of current (and past) economic conditions but public perceptions of the economy are also determined, in large part, by previous perceptions of the economy; people have memories and are slow to change their evaluations. As a result, the ICS at time t likely depends on evaluations at $t-1$. In general, when dealing with people, behavior and beliefs tend to persist.

The persistence observed in a time series often reflects an underlying autoregressive process. An *autoregressive process* is one where the current value of a variable y_t is a function of its own past values. The equation for a simple autoregressive process that includes one lag is written:

$$y_t = c + \phi y_{t-1} + \varepsilon_t \tag{10}$$

where c is a constant and ϕ is a parameter giving the weight attached to the previous observation, y_{t-1}.[5] Equation 10 is a stochastic difference equation for an AR(1) processes, an autoregressive process of order 1, because only one lag of y_t is included in the model. The constant (c) is not essential; we could still describe the process summarized in Equation 10 as an autoregression if it did not include a constant. Including this term makes the expression more general - the constant can only be omitted if $c = 0$ - and the expression highlights the affinity between an autoregression and the regression models that you encountered in the past. As we mentioned above, if you were to regress y_t on y_{t-1}, you would have estimated an autoregression and the coefficient ϕ would be the autoregression coefficient.

You should note that the final term in Equation 10 is the white noise error term (ε_t). In this simple formulation, the white noise error term is a proxy for all the factors, random shocks, that influence the trajectory of the series that are not accounted for by y_{t-1}. Consider our budget example. If a budget follows an AR(1) process, the previous year's budget would determine the current budget, updated by the novel information contained in the current error term. You can probably formulate theories and conjectures about other variables that could influence the level of the budget (tax revenues, political priorities, etc.) but these variables are not included in a univariate model of the process, where the primary goal is to characterize the underlying dynamics of the series.

We can use the intuition from the simple model presented in Equation 10 to produce a more general representation of autoregressive processes. If y_t is a stationary process, then the autoregressive parameter ϕ not only describes the relationship between y_t and y_{t-1}, it also describes the relationship between y_{t-1} and y_{t-2}. Both observations are one unit apart. By definition, the covariances and the variance of the series are constant. If we were to write out the equation for y_{t-1}, we would get the following expression.

$$y_{t-1} = c + \phi y_{t-2} + \varepsilon_{t-1} \tag{11}$$

With an equation for y_{t-1}, we can substitute the righthand side of Equation 11 into Equation 10. We could repeat this process for y_{t-2}, y_{t-3}, y_{t-4}, and so on. This recursive substitution allows us

[5]The mean of the process is given by $c/(1-\phi)$, the variance is given by $\sigma_\varepsilon^2/(1-\phi^2)$, and the covariance is given by $\sigma_\varepsilon^2\phi^s/(1-\phi^2)$.

to rewrite the AR(1) process in Equation 10 in the more general form

$$y_t = c\sum_{i=0}^{t-1}\phi^i + \phi^t y_0 + \sum_{i=0}^{t-1}\phi^i \varepsilon_{t-i} \tag{12}$$

Equation 12 highlights the constraints that must be placed on ϕ for an autoregression to be meaningful. The ϕ^i parameters give the weights attached to the previous values of the error term and must have an absolute value less than 1.0 in order for the series to be stationary. The outcome, in this formulation, is expressed as a progressively discounted summation of everything that happened in the past, where the discounting is exponential; the effects of shocks matter less and less over time. For values of $\phi = 1.0$, the shocks cumulate rather than decay. Such a process is said to be a unit root; we will discuss unit root processes more below. If $\phi > 1.0$, the impact of past shocks will become progressively larger over time. These processes are called explosive and they are very rare in applied work.[6]

The AR(1) process generalizes to allow p lagged values of the process to determine the current value. An AR(p) process is given by:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t. \tag{13}$$

We can calculate the conditions that ensure the AR(p) process is stationary, but will not do so here. The conditions required for stationarity are well established. For an AR(2) process, for example, it must be the case that $\phi_1 + \phi_2 < 1.0$, $|\phi_2| < 1.0$ and $\phi_2 - \phi_1 < 1.0$. Similar rules hold for an AR(3) process. The key is that the sum of the AR coefficients must be less than 1, otherwise the previous random shocks cumulate and the series is a unit root.[7]

In contrast to the random patterns in the plot of a white noise process, the plot of an AR(1) process is relatively smooth. How smooth will depend on ϕ. Figure 2a shows the behavior of an AR(1) process with $\phi = 0.90$. Smaller values of ϕ would produce a series that appears less smooth. A negative value would produce an oscillating time series.

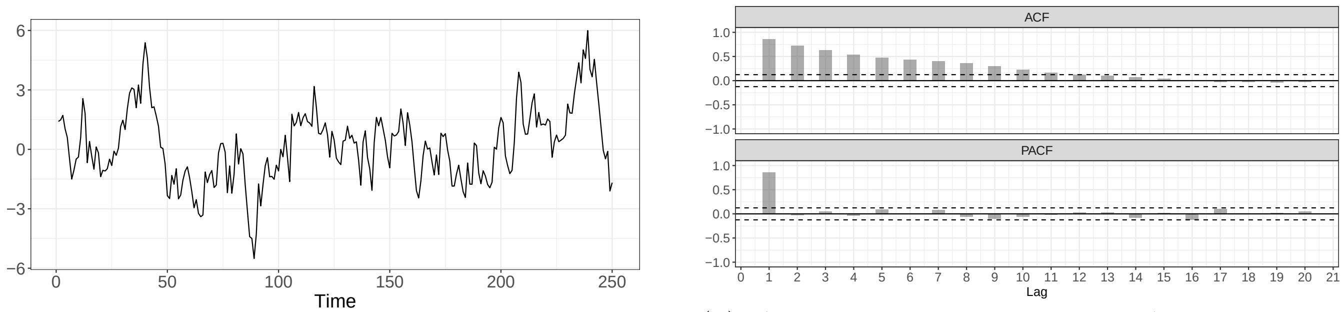
The characteristic patterns in the ACF and PACF of an AR(1) process are illustrated in Figure 2b. The ACF shows a spike at lag 1 that is equal to ϕ and the autocorrelation at each successive lag decays approximately exponentially (for negative ϕ the autocorrelations will oscillate between negative and positive values).[8] The autocorrelations for higher order AR(p) processes will also decay. In contrast, the PACF shows only a single spike (given by ϕ) at lag 1. All other partial correlations have an expected value of 0 because having controlled for the lag one autocorrelation, y_t is uncorrelated with all higher lags. The PACF of an AR(2) process would show 2 spikes and an AR(p) process p spikes. For either smaller values of ϕ or higher order AR(p) processes, the patterns

[6]For values of ϕ less than but close to 1.0, the series may be said to be near-integrated. Such a process, while asymptotically stationary, mimics integrated series in that their variance depends on time in finite samples. See De Boef and Granato (1997); Baillie (1996).

[7]The conditions that ensure the AR(p) process is stationary are determined by the roots of the characteristic polynomial, which are themselves determined through differential calculus. Interested readers should see the appendix to Box-Steffensmeier et al. (2014) and Chapter 1 in Enders (2015), both of which detail methods for solving stochastic linear difference equations.

[8]ACF(s)=ϕ^2.

Figure 2: AR(1): $y_t = \phi y_{t-1} + \varepsilon_t$, $\phi = 0.90$



(a) Time Series Plot

(b) Autocorrelation and Partial Autocorrelation Function

are less discernible and *patterns are key.* It may be that all or most of the autocorrelations for an AR(p) process are inside the standard error bounds; but if they show a pattern of decay, an autoregressive process likely describes the dynamic behavior of the time series.

3.3 Moving Average Process

A moving average process captures a different type of dependence. The defining feature of a white noise process is its lack of dependence, each realization of the process is simply a random shock determined by the error term, ε_t. A *moving average process* relates the current value of a series, y_t, to q past values of the error term. In each period the process adjusts in response to current and previous disturbances. Returning to our budget example, an unexpected shortfall (or windfall) in revenues that is captured in ε_t might lead policy makers to adjust the current year's budget in order to re-establish some acceptable deficit level. Because policy makers' efforts are often imperfect, further adjustments may need to be made in the subsequent year. If a one year adjustment is sufficient to re-equilibrate the budget, the process would be a moving average (MA) process of order one or, more simply, an MA(1) process:

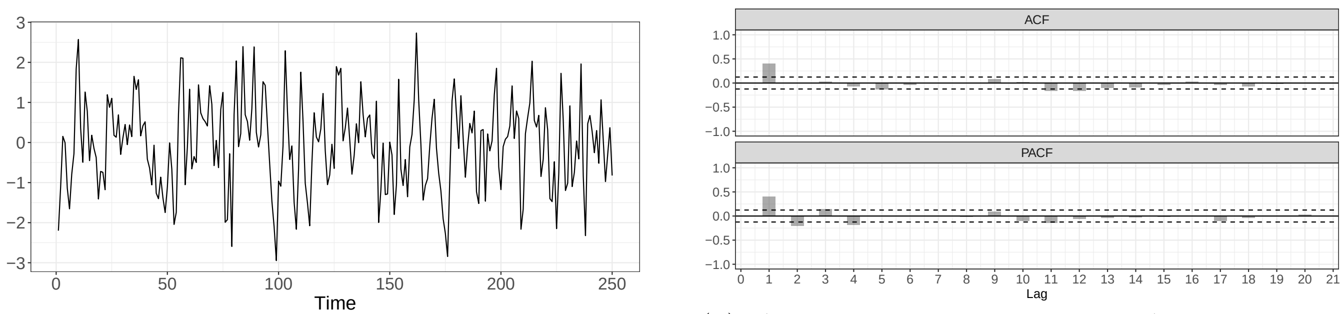
$$y_t = c + \theta\varepsilon_{t-1} + \varepsilon_t \tag{14}$$

where c is a constant, θ is a parameter denoting the weight attached to the previous error, and ε_t is a white noise error term.[9] Compare the equation for the MA(1) process to Equation 12, which presents the AR(1) process in terms of past shocks. You can see that – while the effects of a shock to an AR(1) process decay slowly, at an exponential rate determined by the magnitude of ϕ_1 – the effects of a shock to a MA(1) process are short-lived, lasting only one period.

Figures 3a and 3b illustrate the prototypical behavior of an MA(1) process with $\theta = 0.60$. The evolution of the time series is neither random, as with a white noise process, nor smooth, as the AR(1) process. The ACF looks exactly as we would expect: because the typical effect of a shock to an MA(1) process lasts exactly one period, there is exactly one spike in the ACF. The value of the first autocorrelation is given by $\theta/(1+\theta^2)$. The PACF is also typical of an MA(1) process, showing

[9]The mean of a moving average process is given by c, the variance is given by $(\theta^2 + 1)\sigma^2_\varepsilon$, and the covariance is given by $\theta\sigma^2_\varepsilon$ for $s = 1$, else $\gamma_s = 0$.

Figure 3: MA(1): $y_t = \theta_1 \varepsilon_{t-1} + \varepsilon_t$, $\theta = 0.60$



(a) Time Series Plot

(b) Autocorrelation and Partial Autocorrelation Function

evidence of decay. For a positive θ, the decay in the PACF will oscillate; for a negative θ, the decay will be direct, negative, and gradual.

Note that the ACF for an AR(1) process and the PACF for a MA(1) process mirror each other. That's not a coincidence. The MA(1) process can be rewritten as an infinite AR process after recursive substitution for ε_{t-1}. The MA(1) process is said to be invertible if this recursive substitution results in a stationary infinite autoregressive process. Invertibility in an MA process is the counterpart to stationarity in an AR process. The MA(1) process is invertible if $|\theta| < 1.0$.

We can generalize the MA(1) process to allow q additional past values of the random error to influence y_t:

$$y_t = c + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t. \tag{15}$$

The ACF of an MA(q) process will contain spikes at the first q lags. The PACF will exhibit decay. An MA(2) process is invertible if $|\theta_2| < 1$, $\theta_1 + \theta_2 < 1.0$ and $\theta_2 - \theta_1 < 1.0$. Like the stationarity conditions for an autoregressive process, differential calculus is used to to determine whether an MA(q) process is invertible but, again, this is not necessary in most applications, so we will not reproduce it here.

3.4 Autoregressive Moving Average Process

A process may also contain both autoregressive and moving average components. This would be true for budgets as we have described them: AR components to capture the inertia in the budgeting process and MA components to capture the adjustments associated with unanticipated revenue shortfalls and windfalls. We write the general ARMA(p, q) model as:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + ... \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{16}$$

where p gives the highest number of lags of y_t to enter the model and q gives the highest number of lagged values of the error, ε_t, in the model. The values of p and q tend to be small, as the patterns of autocorrelations and partial autocorrelations can typically be captured with relatively parsimonious models. The ACF and PACF of an ARMA(p, q) process will each show evidence of decay. An ARMA(1,1) process is stationary and invertible if $|\phi_1| < 1.0$ and $|\theta_1| < 1.0$.

Many time series texts give detailed information on how to identify ARMA(p, q) processes and estimate ARMA models. The procedure, pioneered by Box and Jenkins (1976), begins by examining the ACF and PACF of the time series under analysis in order to identify candidate models consistent with the observed correlation structure in the data. The candidate models are then estimated and a series of diagnostic tests is performed to ensure that the residuals exhibit no systematic correlations. Models that meet this criterion are satisfactory representations of the underlying DGP. Information criteria are typically used to select a final, parsimonious model that does not include unnecessary lags of y_t or ε_t. McCleary, McDowall, and Bartos (2017) and McDowall, McCleary, and Bartos (2019) provide in depth treatments on the identification and estimation of ARMA models.

We choose not to elaborate on the identification and estimation of ARMA(p, q) models. The Box-Jenkins methodology represents a different approach to time series analysis than the methods outlined in this text. ARIMA(p, q) models were initially developed for forecasting; analysts use sample data to build a dynamic model of the underlying process in hopes of predicting future values of the series. While the models discussed in this book can be used for forecasting, in theory, that is not our primary focus and, as we will discuss in subsequent chapters, conditioning forecasts on independent regressors requires exogeneity assumptions that are not typically plausible in most social science settings. Rather than trying to capture the dynamic properties of time series using AR and MA components, we will discuss how you can build dynamic regression models of dynamic relationships using lags of the regressands y_t and the regressors x_t for hypothesis testing; we will not discuss the estimation of MA models using lags of ε_t in this text.

You may be wondering why we would introduce MA(q) and ARMA(p, q) processes if we are not going to tell you how to estimate ARMA(p, q) models. As we mentioned at the beginning of the chapter, our goal here is to introduce you to the different types of dynamic processes that you can encounter in applied time series analysis, this includes MA(q) processes. As it happens, these dynamics can be captured using lags of y_t and x_t. The problem, of course, is that using lags of the regressand and / or the regressors to capture MA(q) dynamics will bias your estimates (Cook and Webb 2021). Not to worry, in the chapters where we discuss the specification and estimation of dynamic regression models, we will tell you how you can test for this bias and how you can use alternative econometric methods to draw reliable statistical inferences in the presence of these dynamics. For now, it is only necessary for you to know what types of stochastic dynamics are possible. Which leads us to one final digression before we discuss non-stationary processes.

3.5 Seasonality

Occasionally there will be regular patterns in the time dependence exhibited by a time series that repeat over some number of time periods, s. This type of behavior is referred to as *seasonality* and is often associated with behavior that varies predictably over the calendar year. Monthly unemployment rates, for example tend to be lower during holiday months when employers hire additional help in anticipation of greater demand, while they tend to be higher at the start of the summer, when students enter the work force. Similarly, the number of hearings held in Congress drops predictably during periods of recess over the months comprising a Congress. And the level of interest in politics measured monthly rises predictably with the election cycle.

If $s = 12$ and we have monthly data, then the value 12 months ago might be related to the current value. For example, suicide rates this December are likely to be similar to the previous December. The same holds for any other month of the year.[10] These patterns can be captured with simple extensions to ARMA(p, q) models by adding a lag equal to the length of the season. Thus monthly data featuring the US Congress, which is reconstituted every two years, might have a seasonal autoregressive lag at 24 months, in addition to any non-seasonal autoregression:

$$y_t = c + \phi_1 y_{t-1} + \phi_{24} y_{t-24} + \varepsilon_t. \tag{17}$$

This model is denoted as an ARMA$(1, 0)(1_{24}, 0)$ model, where there is one non-seasonal lag of y_t and one seasonal lag, at 24 periods. Note that the ones in parentheses tell us that both the nonseasonal and seasonal autoregression is of order $p = 1$. The zeroes in parentheses tell us that there are no moving average processes in the data, either non-seasonally or seasonally. The nonseasonal descriptors are listed first and followed by the seasonal features of the model. The subscript denotes at what lag the seasonality operates. Even if you will not be estimating ARMA or seasonal ARMA models, the ARMA notation provides a succinct way to describe seasonal processes.

It is also possible to have other combinations of seasonal and non-seasonal autoregressive or moving average terms. An ARMA$(0, 2)(1_{12}, 0)$ model, for example, has no non-seasonal autoregression, two nonseasonal lags of the error term ($q = 2$) and one seasonal lag of y at 12 months ($p = 1$). The seasonality occurs with a lag of 12 periods and there is no seasonal moving average component to the model.[11]

4 Non-stationary Processes

Not all stochastic time series are stationary. The defining feature of all stationary processes is that they equilibrate, they return to their means in the long run; non-stationary time series do not. A covariance stationary process has a constant mean, constant variance, and constant covariances. A *non-stationary* stochastic process is a series that has a mean, a variance, and / or a covariance that change(s) over time. In the social sciences, non-stationarity is typically introduced by trends. The defining feature of a trend is that it has a permanent effect on the series (Enders 2004, 158). Trends may be either deterministic, where a fixed amount is added to the value of y_t in each period, or stochastic, where a random shock permanently alters the path of the time series.

4.1 Deterministic Trends

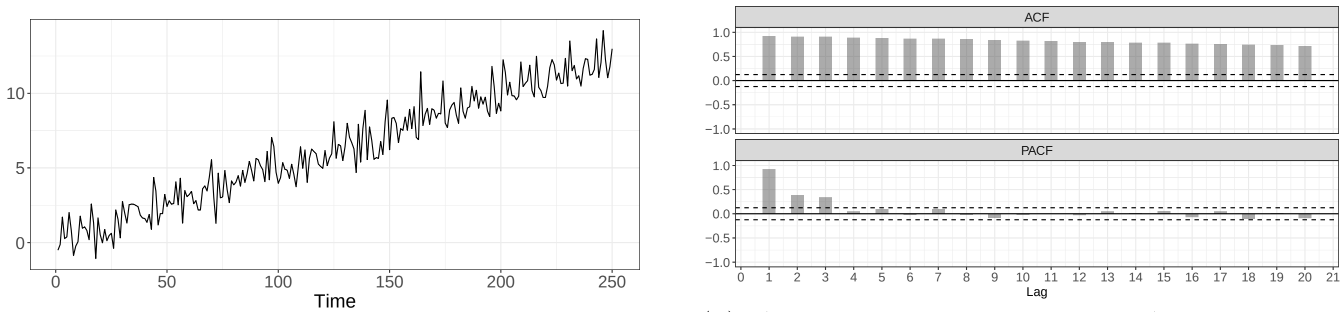
A *deterministic trend* is a variable that changes as a function of time. The most common type of deterministic trend in applied work is linear. Formally, a variable y_t follows a linear deterministic trend if

$$y_t = c + \delta t + \varepsilon_t \tag{18}$$

[10] The behavior of many economic time series is tied to the calendar. In fact, the Bureau of Labor Studies and Bureau of Economic Statistics, for example, distribute seasonally adjusted versions of much of their data.

[11] The ACF and PACF of seasonal processes parallel their nonseasonal counterparts with evidence of seasonality appearing in each at s lags in accord with the nature of the seasonal effect, whether AR or MA.

Figure 4: Linear Deterministic Trend: $y_t = \delta t + \varepsilon_t$, $c = 0$, $\delta = 0.05$



(a) Time Series Plot

(b) Autocorrelation and Partial Autocorrelation Function

where c is a constant, t is a time counter, ε_t is a white noise error term, and δ is a slope parameter that defines the relationship between t and y_t. As before, Equation 18 could be simplified if $c = 0$ but allowing $c \neq 0$ makes the expression more generalizable. It is also possible for there to be other AR and MA dynamics in a series that contains a deterministic trend.

The defining feature of Equation 18, what make the series a *deterministic trend*, is δ; δ is the amount that y_t changes in each period. If δ is positive, y_t increases over time. If δ is negative, y_t decreases over time. When we say that y_t is a "linear function" of time, we mean that Equation 18 is literally the equation for a line. Instead of $y_i = mx_i + b$ (where m is the slope and b is the y-intercept) or $y_t = \alpha + \beta x_t + \varepsilon_t$ (where β is the slope and α is the y-intercept), Equation y_t has a slope δ and a y-intercept c that summarize the linear relationship between the variable t (time) and the outcome y_t. The outcome y_t is increasing by a factor of δ each period and δ is constant over the sampling window, so we say that y_t is a *deterministic trend.*

A deterministic trend is non-stationary because it violates the first criterion for a covariance stationary series, the mean is not constant over time. Figure 4a shows a deterministic trend, with $\delta = 0.05$, $c = 0$, and $\sigma^2_\varepsilon = 1$. It is easy to see that the mean of the series changes over time. Imagine the following exercise. If you calculated the mean for observations 1 through 150 and then calculated the mean for observations 100 through 250, would the means of the two subsamples be the same? Obviously not, $\mu_{1:150} \neq \mu_{100:250}$.

The regular increase in the value of the time series has a dramatic effects on the ACF and PACF, depicted in Figure 4b. Although the error is a white noise process, there is substantial autocorrelation in y_t that persists over all 20 lags and the PACF shows a high level of correlation in the first lag. This is a pattern you will see throughout this section. The non-stationary elements of the time series dominate all the subordinate dynamics that could be observed. As a consequence, we always see the same slow rate of decay in the ACF and a high ($\phi_1 \approx 1$) level of correlation in the PACF. Here, we also see some decay in the PACF. Even after controlling for the correlation at intervening lags, substantial correlation exists for several lags.

An otherwise stationary process that contains a deterministic trend, like the simple one summarized in Equation 18 and depicted in Figure 4a, can be made stationary by filtering out the trend. This is why we call this kind of series *trend stationary.* This filtering process, called *de-trending*, is accomplished by estimating Equation 18, regressing y_t on the time variable or time counter in your data set. The residuals from this auxiliary regression are the stochastic elements of the series less the deterministic trend. One can regress y_t on t and work with the de-trended series but it is more common to see people include t as a variable in the final regression model. When you include a trend term as a regressor, you are "controlling for" the trend in the series, so there is no need to go through this two-step procedure of de-trending and building a model of the de-trended data. Nonlinear deterministic trends – t^2, t^3, etc. - can also filtered out of a series in an analogous fashion, but these kinds of non-linear trends are much less common.

4.2 Stochastic Trends and Unit Roots

A stochastic trend is quite different than a deterministic trend. Where a process containing a deterministic trend adds a fixed value, determined by t, to a time series in each period, a process containing a stochastic trend adds a random value, determined by the error term, in each period. In other words, a stochastic trend is defined by the permanence of random shocks to the process while a deterministic trend permanently incorporates time. Processes that contain stochastic trends are referred to as *unit root* processes.[12] We begin by discussing a process containing only a stochastic trend, the random walk process, then we introduce the random walk with drift and the random walk with drift and trend.

The Random Walk. The simplest unit root process is a random walk. A random walk looks like an autoregressive process where $\phi = 1.0$.

$$y_t = y_{t-1} + \varepsilon_t. \tag{19}$$

The current value of y_t is equal to its previous value plus a random shock. This process is called a "random walk" because each successive observation – each *step* the process takes – is a random draw from the distribution for ε_t. While the expected value of $\varepsilon_t = 0$, the nature of randomness means that there will be periods of successive positive (negative) shocks to the process that cause it to wander unpredictably upward (downward) for periods of time.

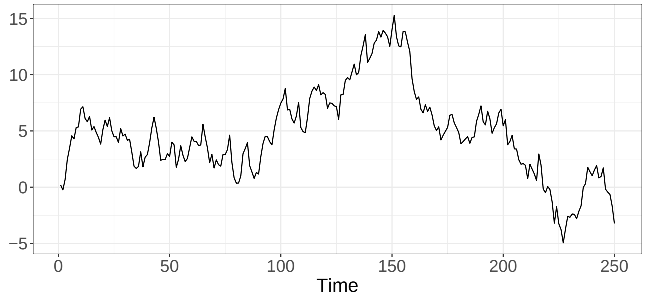
Some algebra will provide further insight into the unpredictable nature of random walk processes. If $y_t = y_{t-1} + \varepsilon_t$, then $y_{t-1} = y_{t-2} + \varepsilon_{t-1}$, $y_{t-2} = y_{t-3} + \varepsilon_{t-2}$, and so on. If we recursively substitute these successive equations into one another, we can write the random walk as the sum of past shocks:

$$y_t = y_0 + \sum_{i=1}^{t-1} \varepsilon_t. \tag{20}$$

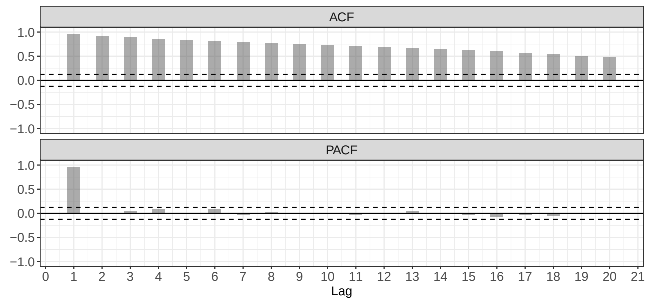
where y_0 is the first observation of the series. This highlights that each random shock is permanently incorporated into the value of y_t, creating a series that wanders unpredictably based on the values of ε_t.

[12] Random walk, unit root, and stochastic trend can be used interchangably. A random walk plus drift can be called a unit root with drift. The unit root (random walk) elements of the stochastic processes described in this section are what impart the stochastic trend.

Figure 5: Random Walk: $y_t = y_{t-1} + \varepsilon_t$



(a) Time Series Plot



(b) Autocorrelation and Partial Autocorrelation Function

As the expected value of $\varepsilon_t = 0$, the expected value of a random walk process is the first value. Thus, a random walk is mean stationary. However, the variance of a random walk process increases over time.

$$\begin{aligned} \gamma_0 &= E(y_t^2) = E(y_o + \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \varepsilon_{t-3} + ... + \varepsilon_1)^2 & (21) \\ &= E(y_o)^2 + E(\varepsilon_t^2) + E(\varepsilon_{t-1}^2) + E(\varepsilon_{t-2}^2) + E(\varepsilon_{t-3}^2) + ... + E(\varepsilon_1^2) & (22) \end{aligned}$$

Because y_0 is a constant, $E(y_o)^2 = 0$. The remaining terms all equal the variance of ε_t, so we can simplify the previous expression to

$$\gamma_0 = t\sigma^2_{\varepsilon_t} \tag{23}$$

This makes it clear that the variance of a random walk increases over time. For similar reasons, the covariances of a random walk change over time: $\gamma_s = (t - s)\sigma^2_\varepsilon$. As a result, the autocorrelations for unit root processes are also a function of time.

Figure 5a shows a plot of a simulated ranodm walk. This illustrates the tendency for unit root processes to wander. The expected value of this random walk is $y_0 = 0$ for all time points but the random walk will go for long stretches without ever crossing zero. The ACF and PACF for a unit root are similar to to what you would observe with an AR(1) process, there is decay in the ACF and a single spike in the PACF. However, the decay in the ACF is very slow. [13] The affinity between a stationary AR(1) process and a random walk process makes classifying a series as stationary or non-stationary difficult. We will discuss this issue further in Chapter 3 when we introduce unit root tests.

Unit roots hold a special place in the analysis of financial time series. In his classic manuscript, *A Random Walk Down Wall Street*, eminent economist Burton Malkiel explains why stock prices are so difficult to predict (Malkiel 1999). Because there are so many people participating in stock markets, buying and selling stocks billions of times a day, all the information relevant to future values of stocks is reflected in the prices at the end of the day. This is the efficient market hypothesis

[13] Recall the formula for the autocorrelation: $\rho_s = \gamma_s/\gamma_0$. Thus, $\rho_s = (t-s)\sigma^2_\varepsilon / t\sigma^2_\varepsilon$ or $(t-s)/t$. For small values of s, this ratio will be close to one, but as s grows, there will be slow decay in the ACF. This makes it hard to differentiate the decay in an AR(1) process with ϕ approaching 1.0 from the slight decay in a random walk process.

(EMH). While the EMH has its supporters and detractors, the theory does a good job of explaining the wandering nature of stock prices. The price of a stock when the market opens is the value that it was when the markets closed on the previous day ($p_t = p_{t-1}$), plus or minus any revisions that occur based on new information received over night. If you picked up this book hoping to make a fortune on the stock market, you may be disappointed but you might be able to learn something about modeling stock returns, changes in stock prices over time.

The random walk can be removed by *first-differencing*, subtracting the previous value y_{t-1} from the current value y_t. In this case the difference transformation produces a stationary (white noise) time series:

$$\Delta y_t = y_t - y_{t-1} = \varepsilon_t \tag{24}$$

where Δy_t is shorthand for the one period change in y_t. Because the random walk is stationary after first-differencing, it is said to be *difference-stationary*.

Equation 24 reveals an important property of random walk processes. If ε_t is a white noise process, the expected change in y_t from period to period is zero. If you difference a stock price series (the series in levels), you get the stock returns series (the differenced series). The expected value of returns series tend to be zero. If the average return for buying and holding a stock were above zero, investors could make money buying a stock and selling it for a higher price each day. If such a stock existed, the opportunity wouldn't exist for long because investors would swoop in to make their fortunes. A non-zero mean after differencing would imply $c \neq 0$ in Equations 19 and 24. This would not only suggest a different kind of financial theory, it would also generate a very different kind of stochastic behavior.

The Random Walk Plus Drift. If we add a constant to the unit root process the resulting process is a *random walk plus drift*.

$$y_t = c + y_{t-1} + \varepsilon_t. \tag{25}$$

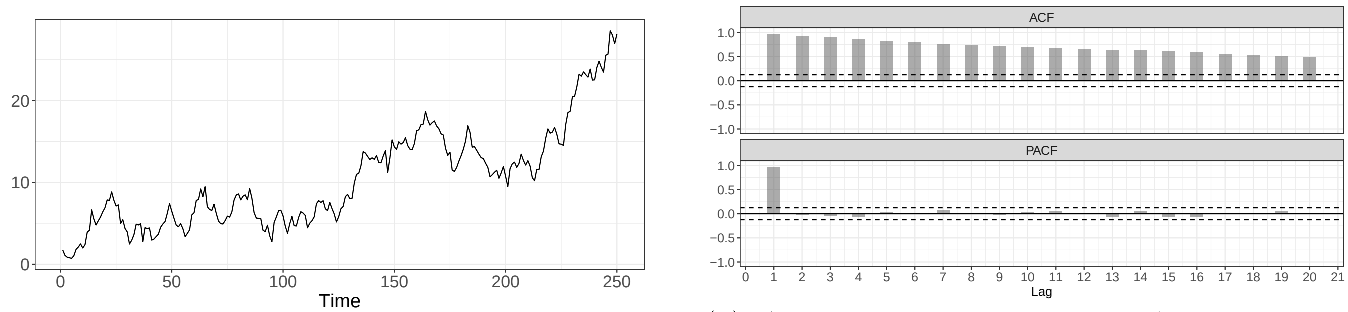
Though it is not readily obvious from Equation 25, this process contains both a deterministic trend and a stochastic trend. The deterministic trend is revealed when you recursively substitute for y_{t-1}.

$$y_t = y_0 + ct + \sum_{i=1}^{t-1} \varepsilon_t. \tag{26}$$

With each realization of y_t, a fixed amount given by c is added to the process. This imparts a deterministic trend. When the deterministic trend is caused by c instead of δt, we refer to this trending behavior as *drift*. The series is called a random walk plus drift because the errors also accumulate, imparting a stochastic trend similar to the pure random walk. The random walk plus drift is non-stationary in mean ($ct + y_0$), variance ($\sigma^2_{\varepsilon_t}$), and covariance ($\gamma_s = (t-s)\sigma^2_{\varepsilon}$).

Figure 6a shows a random walk with positive drift. The constant ($c = 0.05$) is equivalent to the trend coefficient ($\delta = 0.05$) used to generate the series in Figure 4, but the trajectories of the series are different. You can see the tendency for a random walk with positive drift to increase by c in each time period. A random walk with a negative drift term would show a similar tendency to decrease over time. However, because each random shock is permanently incorporated into the value of the process, sometimes the increase will be greater than c and sometimes less; only on

Figure 6: Random Walk Plus Drift: $y_t = c + y_{t-1} + \varepsilon_t$, $c = .05$



(a) Time Series Plot

(b) Autocorrelation and Partial Autocorrelation Function

average will the process increase by c. This behavior produces a more “fluid” trend than the trend stationary series presented in Figure 4. However, Figure 6b shows that the ACF and PACF of the random walk plus drift exhibit the same patterns as a pure random walk.

The first-difference of a random walk plus drift results in a stationary time series with a mean equal to c.

$$\Delta y_t = y_t - y_{t-1} = c + \varepsilon_t. \tag{27}$$

Thus, like the random walk, the random walk with drift is difference stationary.

Random Walk with Drift and Trend A unit root process can also contain drift, a deterministic time trend, and a stochastic trend:

$$y_t = c + \delta t + y_{t-1} + \varepsilon_t. \tag{28}$$

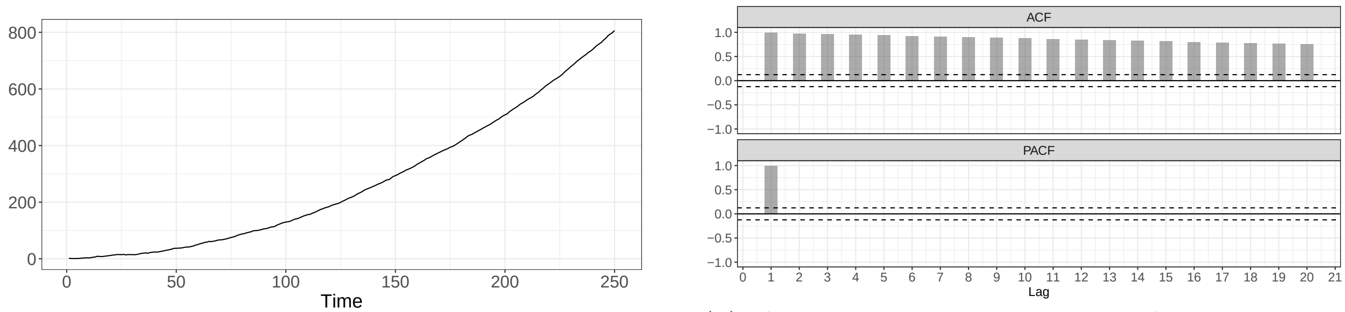
This kind of unit root process is called a *random walk with drift and trend.* If we recursively substitute for y_{t-1} again, we get a sense of the behavior of this type of process.

$$y_t = c(t) + \delta \frac{t(t+1)}{2} + y_0 + \sum_{i=1}^{t} \varepsilon_i \tag{29}$$

Because the process contains a stochastic trend, the errors accumulate over time and in each period a constant c is added to the value of y_t. In addition, the quantity $(\delta \frac{t(t+1)}{2})$ is added; how much is added depends on the value of t. Ignoring the other elements of Equation 29, the quantities added for the first five periods are $\delta(1)$, $\delta(1.5)$, $\delta(2)$, $\delta(2.5)$, and $\delta(3)$. The fact that this term is getting larger in each period means the process is changing at an increasing rate: increasing or decreasing quadratically depending on the sign of δ. In each period the magnitude of the deterministic change is accentuated or attenuated depending on the value of the error term ε_t.

A simulated random walk with drift ($c = 0.025$) and trend ($\delta = 0.025$) is presented in Figure 7a. The difference beween a quadratic trend and a linear trend is clear. Figures 4 and 6 show a trajectory that could be traced with a straight line. The trajectory of the series presented in Figure 7a is more curved; the random walk with drift and trend is increasing at a faster rate than either

Figure 7: Random Walk Plus Drift and Trend: $y_t = c + y_{t-1} + \delta t + \varepsilon_t$, $c = .025$, $\delta = .025$



(a) Time Series Plot

(b) Autocorrelation and Partial Autocorrelation Function

of these series. Despite the differences in trajectory, the ACF and PACF for the random walk with trend and drift, shown in Figure 7b, follow the same patterns as the other unit root processes: slow decay in the ACF and a single significant spike in the PACF.

A random walk with trend and drift must be differenced and detrended to be made stationary. If you subtract y_{t-1} from both sides to difference a random walk with trend and drift, you have removed the unit root but you are left with a time series that is still increasing over time

$$\Delta y_t = c + \delta t + \varepsilon_t, \tag{30}$$

so you would still need to regress Δy_t on t to remove the effect of time.

Random walks, random walks with drift, and random walks with drift and trend are all said to be integrated of order one, denoted $I(1)$ or $ARIMA(0, 1, 0)$; they are all unit roots. The one refers to the number of times the series must be differenced to be made stationary. The term *integrated* refers to the fact that the current value of the process is the sum of all previous shocks; another ε_t is integrated into y_t in each period.

5 Why are these Distinctions Important?

Stationary versus non-stationary, trend stationary versus difference stationary, random walk versus random walk with drift: why do these distinctions matter? Why must we have names for all these different stochastic processes and why is it important to know the differences? Determining whether a process is stationary or contains a unit root and identifying the deterministic features of your data are essential tasks in applied time series analysis for a number of reasons.

First and foremost, classifying your data as stationary or non-stationary is important for establishing the veracity of results from traditional regression models. Regression analyses of unit root processes are prone to the *spurious regression problem.* Yule (1926) and Granger and Newbold (1974) showed that a regression involving independent (not related) unit root processes produce evidence of a statistically significant relationship at high rates. This problem worsens as the sample size increases. The spurious regression problem results from the tendency of unit root series to wander. Two unit root processes will tend to wander together, or apart, and thus appear to be positively or negatively correlated. To be sure, unit root processes can be related. When two unit root processes are related in the long run, they are *co-integrated*. We will discuss this topic and the error correction models used to examine these relationships in Chapter 6.

In some cases, it will be clear that your data are non-stationary. If you plot a series and it increases over time, as in Figures 4 and 6, you will not find yourself agonizing over the possibility that the series has a predictable long-run mean. Still, the accurate classification of non-stationary dynamics can have important consequences for model building. You need to be able to differentiate a trend stationary series from a random walk with drift in order to identify the transformation required to render the series stationary. Subtracting a deterministic time trend from a difference stationary process will fail to produce a stationary time series, as the stochastic trend has not been eliminated. First-differencing a deterministic trend induces a noninvertible moving average, producing a different type of non-stationarity. A random walk with drift and trend must be first-differenced and detrended. It is not enough to identify a trend in your data, you need to know what type of trend you are working with. A variety of statistical tests are available to help us infer the nature of any trends in the data. We discuss a testing strategy in Chapter 3.

We will also see that the variety of non-stationarity has important consequences for hypothesis testing. Many standard results in probability theory - like the Central Limit Theorem and the Law of Large Numbers - do not apply to unit root processes. Hypothesis tests involving time series with stochastic trends require nonstandard distribution theory. When we begin working with these nonstandard distributions later in the text, you will see that the variety of non-stationarity determines the shape of the underlying distribution.

Finally, the properties of your data can help you discern the plausibility of your theory. Scholars debate the nature and prevalence of trends in social science data. On one hand, some are skeptical that any effects are truly permanent. This makes them skeptical that either deterministic or stochastic trends really exist. On the other hand, many time series appear to trend over time. These patterns have lead some to develop economic and social theories in terms of trends. At one time, the predominant view held by economists was that growth tended to be deterministic. GDP, for example, was expected to grow by a fixed amount in each period, caused by the regular march of technological advancement. Spurred, in part, by the unit root tests conducted by Nelson and Plosser (1982), most researchers now believe that many economic time series are better characterized as unit root processes with drift, increasing *on average*, rather than by a fixed amount, over time.

Theoretical arguments for stochastic trends are rare in the social sciences. Consider a notable exception. Williams and McGinnis (1988) offered a rational expectations theory to explain the arms race between the United States and U.S.S.R. during the Cold War. They contend that "political actors [on both sides] made efficient use of their existing capabilities for information processing" (973). Intelligence gathering informed expectations about rival defense spending. Along with a domestic appetite for security, these expectations fueled public support for the arms race. From year to year, information was permanently incorporated into the current level of defense spending. Only surprise behavior by the rivals (random shocks) could explain changes in spending over time. Despite the plausibility of this argument, there is no consensus on its validity; some even suggest that the autocorrelation in defense spending is indicative of a stationary process. As we will see in

Chapter 3, you can use hypothesis tests to discriminate between unit root and (trend) stationary processes to test the consistency of these arguments with your data.

6 Fractional Integration

Stationary ARIMA$(p, 0, q)$ processes and non-stationary ARIMA$(p, 1, q)$ processes impose a dichotomy that may be too restrictive. Granger and Joyeux (1980) and Hosking (1981) proposed fractional integration as an alternative stochastic process that relaxes the assumption that the level of integration is a whole number and allows it to hold any real value. A general autoregressive fractionally integrated moving average or ARFIMA(p, d, q) process is given by:

$$\phi(L)(1 - L)^d Y_t = \theta(L)\varepsilon_t \tag{31}$$

where L is the lag operator such that $L^b Y_t = Y_{t-b}$, d is the fractional differencing parameter, ϕ are stationary autoregressive parameters, θ are stationary moving average parameters, and ε_t is a well-behaved (white noise) error term. A fractionally integrated series is variance and covariance stationary when $0 < d < .5$ but not when $0.5 \leq d < 1$. The process is said to exhibit intermediate memory (anti-persistence), or long-range negative dependence, for $d \in (-0.5, 0)$.

A fractionally integrated $I(d)$ process looks and behaves differently than processes that are $I(0)$ or $I(1)$. Recall that shocks to a stationary ARMA process have short-term effects on the series. In contrast, shocks to a non-stationary $I(1)$ process cumulate over time. Fractionally integrated processes are called *long memoried* because the effects of shocks die down eventually, but slowly. They show statistically significant correlations over long lags (Baillie 1996) and are technically mean stationary. The autocorrelation function of a fractionally integrated process will decline at a hyperbolic rate – slower than the exponential rate of an autoregressive process, but faster than the perfect memory of an integrated process.

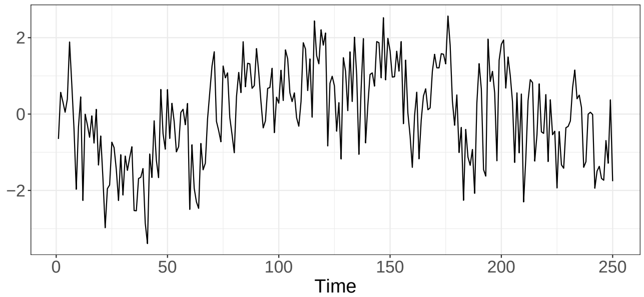
Figure 8 presents a time series plot, an ACF, and a PACF for an ARFIMA$(0, d, 0)$ process with $d = 0.40$. The plot itself is not particularly telling, but the ACF shows the very slow decay characteristic of an ARFIMA process. Even at 20 lags the autocorrelations are significantly different from zero. The ACF exhibits wave like patterns; this is the hyperbolic decay that is indicative of a fractionally integrated process.

Like unit root processes, fractionally integrated processes are prone to the spurious regression problem (Lebo, Walker, and Clarke 2000; Tsay and Chung 2000). An analyst conducting a regression involving two independent fractionally integrated series is too likely to wrongly conclude a relationship exists and this problem increases as d approaches 1. Fractionally integrated processes may be related to each other in level-form, in which case the series are *fractionally cointegrated*.

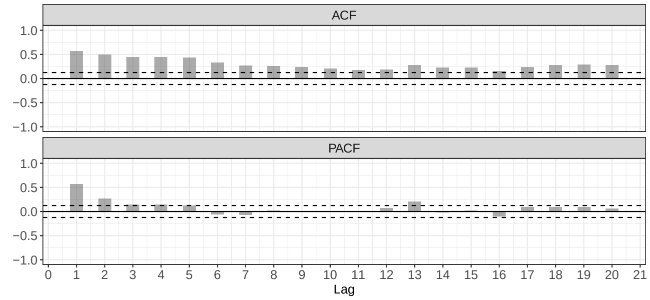
Fractionally integrated data can also be transformed. Just as first-differencing can transform an integrated series into a stationary series, fractional differencing transforms a fractionally integrated series into a stationary series (Lebo, Walker, and Clarke 2000). The fractional difference is given by:

$$(1 - L)^d \tag{32}$$

Figure 8: ARFIMA$(0, d, 0)$: $d = .40$



(a) Time Series Plot



(b) Autocorrelation and Partial Autocorrelation Function

where d is from an estimated ARFIMA(p, d, q) model. To fractionally difference a series, you apply a Hosking's (1981) filter:

$$(1 - L)^d = \sum_{s}^{\infty} \frac{\Gamma(s - d)L^s}{\Gamma(s + 1)\Gamma(-d)} \tag{33}$$

where Γ is the Gamma function and s is the lag length. Applying the correct ARFIMA filter to a time series will create a new time series of residuals that should be white noise.[14] The filtered series is the fractionally differenced version of the variable.

How common is fractional integration? Granger (1980) offers a powerful argument for FI in aggregate time series processes. He shows that time series constructed as aggregations of individual units will evolve naturally as $I(d)$ processes if some individual units follow purely autoregressive patterns while others follow unit root processes (see also Zaffaroni (2004).) For example, Achen (1992) argues that individual-level party identification is unlikely to be homogenous in the population. Building on this idea, Box-Steffensmeier and Smith (1996) argue that the permanence of party identification for individuals can exist anywhere along a continuum from "complete or perfect permanence to no persistence," and hypothesize that macropartisanship is fractionally integrated. They find that macropartisanship is best characterized as an $ARFIMA(0, 0.839, 0)$ process. Analysts have found many other political processes to be fractionally integrated including macro-ideology (Box-Steffensmeier and DeBoef 2001), presidential approval (Lebo, Walker, and Clarke 2000), yearly Supreme Court justice behavior (Lanier 2011), and party approval ratings (Byers, Davidson, and Peel 2000).

[14]This process of "pre-whitening" is a main feature of the Box-Jenkins (1976) approach to time series that relies on the diagnostic tools we outline in the next chapter. It's important to note that filtering data to create a white noise time series does not mean that all meaningful information has been lost. Rather, the process is meant to extract the way a time series depends on its own past history in order to be left with just that portion of the series that might be explained by independent variables in a multivariate model.

7 Structural Breaks

In the chapters that follow we will introduce a number of methods that can be used to model dynamic processes. All of these models assume that the underlying data generating process is constant within the sampling window. A *structural break* occurs when there is a change in the underlying data generating process.

Structural breaks can be temporary or permanent. Temporary breaks are short term changes in a process. An unexpected event might cause an unexpected increase or decrease in the series but eventually the series returns to its pre-event equilibrium. The plot of Presidential Approval during the George W Bush administration in Chapter 1 shows a temporary structural break. The first 8 months of his presidency were relatively normal but the events of September 11, 2001 cause his approval rating to skyrocket beyond 90%. His approval rating settled back to a normal range by the end of his first term but, for a short period, George W Bush enjoyed the highest approval rating of any president in the history of the United States.

Permanent structural breaks occur when changes in the underlying process are not temporary. A policy change or the introduction of a new technology may permanently alter the underlying dynamics of a process. Brandt and Sandler (2010) use time series of terrorist attacks to examine how terrorist tactics and targeting have changed over time. One of the significant changes they identified in terrorist tactics occurred in 1973, when the United States and other developed countries began introducing metal detectors in major airport terminals. These security provisions made it more difficult for passangers to smuggle guns and other weapons onto planes, permanently altering the tactics of would be hijackers.

Structural breaks can complicate time series analyses. If the structural change in the process is relatively small compared to the other variation in the series, the effects are minor. In these cases, breaks or events will appear as noise in the data. If the change is large compared to the other variation in the data, the problem is more severe. Major shifts in the underlying data generating process can hamper the tools we use to classify time series. The ACFs and PACFs for series with structural breaks will be distorted, making it difficult to identify the dynamic features of the series. Many of the unit root tests that we outline in the next chapter will be even more prone to misleading results. A failure to control for major events or policy shifts can also produce omitted variable bias problems in your final analyses.

In some cases structural breaks are easy to identify, in some cases they are not. You will often be able to see structural breaks when you plot your time series. The 30% rise in Presidential Approval following the 9-11 terrorist attacks stands out. When you plot a differenced series, you may be able to spot major events. Structural breaks are among the reasons that you should begin every time series analysis by plotting your data. In some cases, you may need to test for structural breaks. Maddala and Kim (1998) provide a survey of the standard methods for identifying and evaluating structural breaks when the locations of breaks are known and unknown. When the location of the break is known, you can apply an F-test like the one described by Chow (1960) to determine whether accounting for the break will be important for the analysis. If you are not sure about the location of the break or whether a break exists, you can apply recursive tests that allow for multiple breakpoints like those developed by Bai (1994, 1997) and others.

You will need to incorporate information about structural breaks at various points in your analysis. Unit root and stationarity tests can be sensitive to the presence of structural breaks, so you will need to modify the testing procedures to accomodate these features of the data. If major events or policy shifts explain a significant amount of variation in your outcome variable, you will need to control for these effects in your analysis. In each case, the approach is the same: you create (a) deterministic variable(s) to "control for" of "partial out" the effects of the breaks. These variables are deterministic because, like a time counter that you would include in an analysis to control for trend, the variables you create to account for breaks do not vary in a random fashion. We will describe how these variables can be created and incorporated into your analysis throughout the text. For now, it is sufficient for you to know that structural breaks can play an important part in the analysis of stochastic variables and that we create deterministic variables to account for these effects.[15]

8 The Takeaway

There are two broad types of stochastic processes you will encounter: stationary processes and non-stationary processes. There are three key pieces of information to remember from this chapter. First, a stationary process is characterized by inertia (unless the process is purely random) and mean reversion. As such, successive values of stationary ARMA(p, q) processes are predictable based on their past behavior. Second, a non-stationary time series may be deterministically trending as a function of t – evolving with the addition (or subtraction) of a fixed amount in each successive period – or stochastically trending or *both*. Stochastically trending processes follow a random walk through time where each step is determined by a random shock (possibly plus drift) and is hence unpredictable. The expected value of the process is the current value (plus any drift), shocks are persistent, and the level of the process wanders without any tendency toward a long run mean. A time series can contain a random walk with drift and trend, although we seldom see this behavior in the social sciences. Such a process will appear to contain a quadratic trend. Finally, a fractionally integrated process falls somewhere between, with the effects of shocks decaying more slowly than an ARIMA$(p, 0, q)$ process.

Each of these types of dynamic processes is best thought of as a prototype. Each is a theoretical characterization of a time series and each is described as distinct. However, multiple characterizations may fit the data equally well and these different types of processes may combine in a single series, e.g., an ARMA process with a deterministic trend and seasonality. *The key issue for analysts who wish to test the relationship between time series is whether the data are stationary or unit root processes.* Coupled with assumptions about the exogeneity and endogeneity of the time series under analysis, the answer to this question determines the appropriate modeling strategy and the distribution of test statistics used to test hypotheses.

[15]There is another approach that treats break points as the focus of the time series analysis. In an interrupted time series analysis, the analyst not only wants to determine if a break exists but also how the intervention changes the dynamic process and how that change transpires (McDowall, McCleary, and Bartos 2019). Extensive treatments of this approach to time series analysis can be found elsewere (see Enders 2015; McCleary, McDowall, and Bartos 2017).

In the next chapter we describe a series of tests designed to help us determine these key features of the data generating process underlying our time series and offer a strategy for inference. Chapter 4 provides tools to determine whether single-equation or multiple-equation analysis is appropriate. In Chapter 5 we discuss how to model relationships between time series that we are confident are ARMA stationary processes (possibly after detrending) given weak exogeneity. In Chapter 6 we consider models that are appropriate when we are confident our time series are unit root processes, again assuming weak exogeneity. However, it is important to recognize at the outset that there may be considerable uncertainty as to the true nature of the DGP for any given time series, even after careful analysis of the individual time series. In Chapter 7 we offer a modeling strategy and hypothesis testing framework for analyzing the relationship between time series given this uncertainty. In Chapters 8 and 9, we cover models for time series that violate the weak exogeneity assumption. Chapter 8 considers the case of stationary time series and Chapter 9 deals with unit root processes.

References

Achen, Christopher. 1992. "Social Psychology, Demographic Variables, and Linear Regression: Breaking the Iron Triangle in Voting Research." *Political Behavior* 14: 195–211.

Bai, Jushan. 1994. "Least squares estimation of a shift in linear processes." *Journal of Time Series Analysis* 15(5): 453–472.

Bai, Jushan. 1997. "Estimation of a change point in multiple regression models." *Review of Economics and Statistics* 79(4): 551–563.

Baillie, Richard T. 1996. "Long Memory Processes and Fractional Integration in Econometrics." *Journal of econometrics* 73(1): 5–59.

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

Box-Steffensmeier, Janet, and Suzanna DeBoef. 2001. "Macropartisanship and Macroideology in the Sophisticated Electorate." *Journal of Politics* 63: 232–248.

Box-Steffensmeier, Janet M., and Renée M. Smith. 1996. "The Dynamics of Aggregate Partisanship." *American Political Science Review* 90(September): 567–580.

Box-Steffensmeier, Janet M, John R Freeman, Matthew P Hitt, and Jon CW Pevehouse. 2014. *Time Series Analysis for the Social Sciences.* Cambridge University Press.

Brandt, Patrick T, and Todd Sandler. 2010. "What do transnational terrorists target? Has it changed? Are we safer?" *Journal of Conflict Resolution* 54(2): 214–236.

Byers, David, James Davidson, and David Peel. 2000. "The Dynamics of Aggregate Political Popularity: Evidence from Eight Countries." *Electoral Studies* 19(1): 49–62.

Chow, Gregory C. 1960. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions." *Econometrica: Journal of the Econometric Society* pp. 591–605.

Cook, Scott J, and Clayton Webb. 2021. "Lagged Outcomes, Lagged Predictors, and Lagged Errors: A Clarification on Common Factors." *Political Analysis* DOI: https://doi.org/10.1017/pan.2020.53: 1–9.

De Boef, Suzanna, and Jim Granato. 1997. "Near-integrated Data and the Analysis of Political Relationship." *American Journal of Political Science* 41(April): 619–640.

Enders, Walter. 2004. *Applied Econometric Time Series.* 2nd ed. New York: Wiley.

Enders, Walter. 2015. *Applied Econometric Time Series.* 4th ed. New York: Wiley and Sons.

Granger, Clive WJ. 1980. "Long Memory Relationships and the Aggregation of Dynamic Models." *Journal of Econometrics* 14(2): 227–238.

Granger, Clive W.J., and Paul Newbold. 1974. "Spurious Regressions inEconometrics." *Journal of Econometrics* 26: 1045–1066.

Granger, Clive WJ, and Roselyne Joyeux. 1980. "An Introduction to Long-Memory Time Series Models and Fractional Differencing." *Journal of Time Series Analysis* 1(1): 15–29.

Hamilton, James D. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hosking, Jonathan RM. 1981. "Fractional Differencing." *Biometrika* 68(1): 165–176.

Lanier, Drew N. 2011. "Acclimation Effects and the Chief Justice: The Influence of Tenure and Role on the Decisional Behavior of the Court's Leader, 1888-2007." *American Politics Research* 39(4): 682–723.

Lebo, Matthew J., Robert W. Walker, and Harold D. Clarke. 2000. "You Must Remember This: Dealing with Long ¡emory in Political Analyses." *Electoral Studies* 19(1): 31–48.

Maddala, Gangadharrao S, and In-Moo Kim. 1998. "Unit roots, cointegration, and structural change.".

Malkiel, Burton Gordon. 1999. *A random walk down Wall Street: including a life-cycle guide to personal investing*. WW Norton & Company.

McCleary, Richard, David McDowall, and Bradley Bartos. 2017. *Design and Analysis of Time Series Experiments*. New York, NY: Oxford University Press.

McDowall, David, Richard McCleary, and Bradley Bartos. 2019. *Interrupted Time Series Analysis*. New York, NY: Oxford University Press.

Nelson, Charles R., and Charles I. Plosser. 1982. "Trends and Random Walks in Macroeconmic Time Series : Some Evidence and Implications." *Journal of Monetary Economics* 10(2): 139–162.

Tsay, Wen-Jay, and Ching-Fan Chung. 2000. "The Spurious Regression of Fractional Integrated Processes." *Journal of Econometrics* 96: 155–182.

Williams, John T, and Michael D McGinnis. 1988. "Sophisticated Reaction in the US-Soviet Arms Race: Evidence of Rational Expectations." *American Journal of Political Science* pp. 968–995.

Yule, G Udny. 1926. "Why Do We Sometimes Get Nonsense-Correlations Between Time-Series?–A Study in Sampling and the Nature of Time-Series." *Journal of the Royal Statistical Society* 89(1): 1–63.

Zaffaroni, Paolo. 2004. "Contemporaneous Aggregation of Linear Dynamic Models in Large Economies." *Journal of Econometrics* 120(1): 75–102.