1

Modeling Social Dynamics

Many social processes are best understood in terms of changes over time. Current social conditions are outgrowths of people's memories of past conditions and, sometimes, also of people's expectations about future conditions. Social processes are rarely, if ever, at rest. Rather they are moving in time, trending in a regular way or exhibiting patterns that recur in time. Understanding the causal relationships between social variables across time and the forces that drive these variables through time is essential to explaining and forecasting social processes. We call these relationships and forces social dynamics. Time series methods are a powerful means by which to analyze these dynamics. In his essay on the history of time series, Tsay (2002, 121) emphasizes that studies of dynamic structures and dynamic relationships between variables have played a vital role in the development of the field, especially through applications in business and economics.

Many important social problems are conceived in terms of dynamics. Indeed, a broad array of questions across the social sciences cannot be addressed without utilizing data comprised of measurements of variables over time. The substantive bite of many empirical findings from the social world comes not from knowing the static level of an important variable at one frozen moment, but rather from understanding the dynamics of that variable. For instance: Is not is the phenomenon of interest rising? Falling? Repeating a pattern seasonally? Drifting randomly? Reverting to a stable value that we can forecast? And how do the dynamics in the phenomenon relate and react to the dynamics in other phenomena? Substantively meaningful answers to questions of this sort require that we conceive our data as a dynamic process and utilize techniques such as those we present in this book.

Moving beyond the abstract, what sorts of phenomena can be conceptualized and studied under the framework we promote here? Consider economic

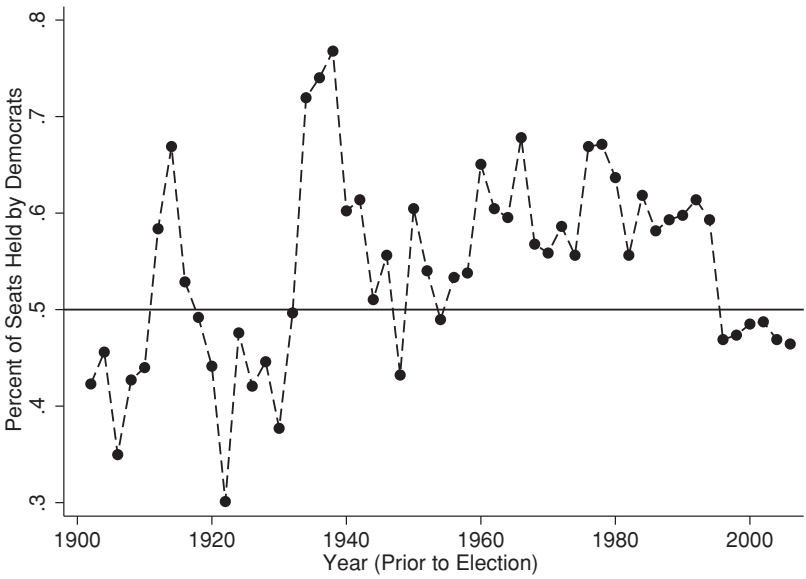

FIGURE 1.1. Democratic Party Seat Share in the U.S. House of Representatives, 1990–2004. Compiled by the authors from data supplied by the Clerk of the House of Representatives.

performance indicators such as unemployment, inflation, or gross domestic product. As both independent and dependent variables, they are integral to empirical analysis in many disciplines, including economics and political science. Measurements of aggregate fertility, obesity, smoking prevalence, and mortality over time are of keen interest in population studies and public health. In the field of education, fundamental data such as rates of literacy, graduation, and dropouts can all be well conceptualized as dynamic processes. A sophisticated understanding of the complex dynamic processes undergirding the rates of various crimes, incarceration, and recidivism is nothing short of foundational to the field of criminology. Adolescent fertility and drug use rates, along with rates of suicide and interpersonal violence across time, are all dynamic processes of great interest in sociology. These are, of course, but a few examples. The point is that, across the social sciences, some of our most interesting and fundamental empirical questions can be best addressed by properly understanding that our data are composed of dynamic processes and then modeling them as such.

What do data composed of dynamic processes look like in practice? We now turn to several illustrations of interesting time series data from a variety of fields. To begin, an important idea in political science is that of representation in democracies. Among other things, this concept implies a recurring alternation of power between contending groups. Consider, for example, Figure 1.1, which shows the balance of power in the U.S. House of Representatives over the

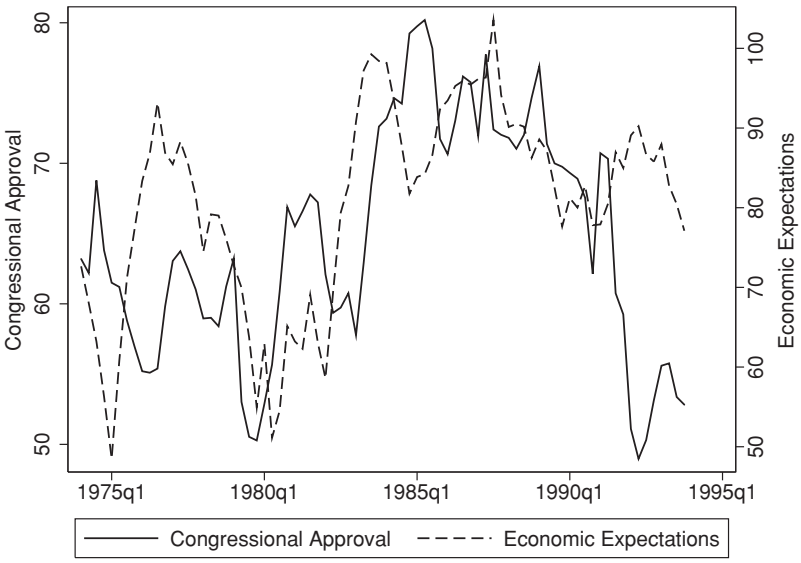

FIGURE 1.2. Congressional Approval and Economic Expectations in the U.S., 1974q1–1993q4. Compiled by the authors from replication data on website of Christina Wolbrecht (Durr, Gilmour, and Wolbrecht, 1997).

previous century.[1] The data are proportions of the seats held by the Democratic Party in the House ordered over time. Do the workings of our institutions produce regular shifts in the balance of power between the Republicans and Democrats? Is there a pattern of alternating control over time? The figure suggests that there might be, but the timing between changes in partisan control is not clear. There was a prolonged period of Democratic control in the middle of the century, but then power shifted to the Republicans. So conceivably there are recurring alterations of power between competing groups, but only over long periods of time. Time series methods help us characterize this process of representation and predict its behavior. We discuss methods for modeling and forecasting from a univariate time series like that in Figure 1.1 in Chapters 2, 5, and 7.

Patterns of political accountability, a concept closely related to representation, provide shorter term insights into this subject. Figure 1.2 shows time series for the proportion of citizens who approve of Congress in each quarter from the beginning of 1974 to the end of 1993. It also depicts the proportion of respondents who have positive expectations about the economy in the same time period. The two series appear to move together, suggesting that the public dispenses approval on the basis of how it evaluates economic policy. If so, the

[1] We denote annual time series by the first and last year of the series, for instance, 1985–1993. We denote quarterly data by the year and quarter, e.g., 1995q1 (see Figure 1.2 above). Monthly time series are denoted by the year and month of the observation, for example, 1997:5–2006:9. Data and STATA replication code for all figures and tables in this book are available at www.politicalscience.osu.edu/faculty/jbox/tsass.

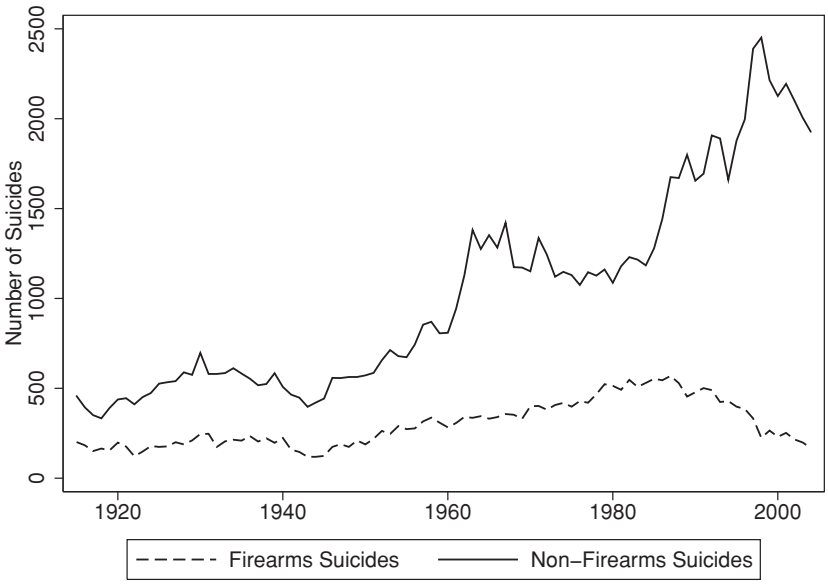

FIGURE 1.3. Suicide Deaths in Australia, 1914–2004. Compiled by the authors from replication data on website of Andrew Leigh (Neill and Leigh, 2008).

data indicates that Congress is held accountable for its policies. But what exactly is the nature of this relationship? Are the evaluations of the economy temporally prior to changes in representation? Is the effect immediate or lagged? Time series methods, such as those we discuss in Chapters 3, 4, and 6, enable us to describe exactly how these variables trend together.

Let us turn to a different field: since the pioneering work of one of the discipline's founding thinkers, Emile Durkheim, sociologists have studied the question of suicide. In very broad strokes, some sociologists study questions such as the following: What are the different types of suicidal behavior? What, if anything, can (or should) governments do to discourage people from committing suicide? Figure 1.3 shows deaths from firearms and non-firearms suicides in Australia from 1915–2004, first collected and analyzed by Neill and Leigh (2010). Non-firearms suicides seem to spike in the 1960s and again in the 1990s, whereas firearms suicides seem to decline in the mid-1990s. After several mass killings, the Australian government implemented a massive firearms buyback program in 1997 to reduce the number of guns in circulation in Australia. Did this program reduce the amount of suicides by firearms in Australia? Using time series intervention models (discussed in Chapter 2), and endogenous structural break models (discussed in Chapter 7), analysts can explore this and other similar questions.[2]

[2] For more on the sociology of suicide, Wray, Colen, and Pescosolido (2011) provide an excellent overview covering research predating Durkheim to modern times. Interested readers are

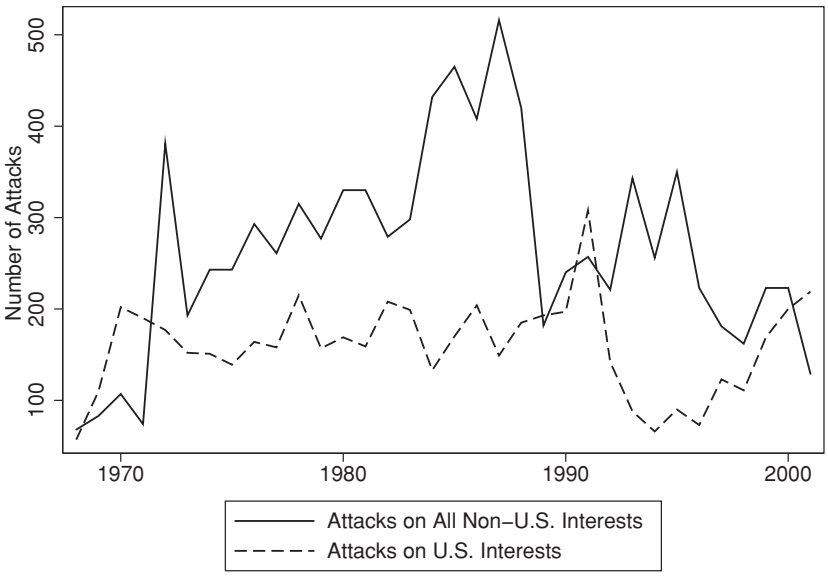

FIGURE 1.4. Terrorist Events and Attacks on the U.S. and other interests, 1968–2003. Reprinted with permission of Elsevier under STM guidelines (Sandler and Enders, 2004).

A different social problem stemming from the field of international relations is terrorism. Figure 1.4 charts the number of terrorist attacks on the United States and on other countries between 1968 and 2003.[3] The two series tend to move in opposite directions between the mid-1980s and mid-1990s. But, in the 1970s and early 1980s as well as in the late 1990s to 2003, the two series often move in the same direction. In addition, the surge in attacks on U.S. interests in 1991 (308) appears to be a precursor to the surge in attacks on other countries in 1993 (343) and 1995 (340). In contrast, the surge in attacks on other countries in 1999 (226) and 2000 (226) seems to foreshadow the surge in attacks on the United States in 2001 (219).[4] Are these series, in fact, related? If so what is the nature of this relationship? More importantly, if we can model this relationship could we use it to forecast and even prevent terrorist attacks? We discuss forecasting as used in the field of time series in Chapter 7.[5]

also referred to O'Brien and Stockard (2006), who utilize a technique (SUR) for time series data we cover in the online appendix to this book (available at www.politicalscience.osu.edu/faculty/jbox/tsass).

[3] Data taken from Sandler and Enders (2004).

[4] Time series data on terrorism now are published regularly by government agencies such as the State Department. See, for example, Sabasteanski (2005). An example of a website on terrorist activity is that of START, www.start.umd.edu/data/gtd/.

[5] For a review of the literature on forecasting in international relations see Brandt, Freeman, and Schrodt (2011).

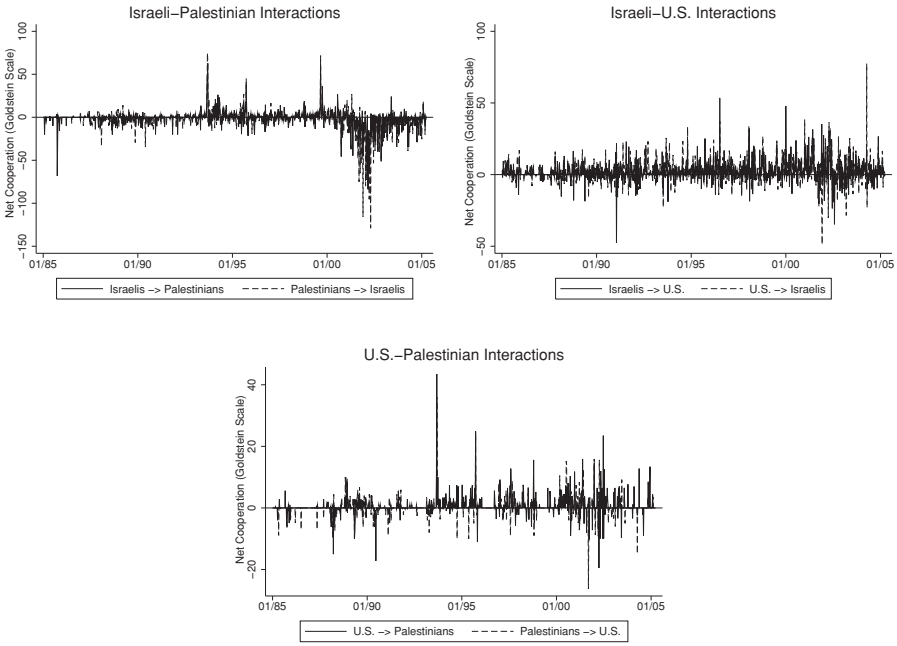

FIGURE 1.5. Israeli-Palestinian-American Relations. Compiled by authors from event data produced by Kansas Event Data System.

Finally, the top-left panel in Figure 1.5 is a representation of relations between the Israelis and Palestinians, which obviously have enormous implications for international politics. The data are based on events data coded to capture the day-to-day interactions of actors in the Levant. This panel shows that the amount of cooperation and conflict in the directed behavior of the two toward each other ordered in time. Events data such as these are based on content analysis and, in recent years, on automated text parsing. Events are extracted from newspaper text and then coded into particular kinds of directed behavior. Each such behavior is given a numerical value indicating the level of cooperation or hostility between the parties.[6]

The density of these data is much greater than that on which Figures 1.1–1.4 are based. This makes it difficult to discern trends and relationships. But, clearly there were short periods of cooperation between the Palestinians and Israelis in and around 1994, 1995, and 1999, as well as periods of marked conflict in 1985 and especially in 2001–2002. These two time series suggest that a complex social dynamic exists between the Israelis and Palestinians. For instance, at times they reciprocate each other's behavior, and at other times they do not: neither a stable relationship nor one that connotes a common

[6] For an overview of this kind of data see Monroe and Schrodt (2008). See also Schrodt (1994) and Merritt, Muncaster, and Zinnes (1993).

trend in behavior appears. Does this mean there is no basis for the evolution of cooperation between the Israelis and Palestinians? Despite the lack of a consistent pattern, like many international actors – ranging from the great powers, to international organizations, to peace organizations – we would like to be able to forecast Israeli-Palestinian relations and perhaps even design foreign policy interventions to help resolve areas of conflict. On the basis of the pattern in the top panel of Figure 1.5, what is the best forecast of relations between these two parties? With respect to an intervention, many observers think that the United States has a role to play. The middle and bottom panels in Figure 1.5 show how the United States acts toward and responds to the Israelis and Palestinians. How are the six time series in the figure related? Do the patterns in Figure 1.5 reveal the potential for mediation by the United States?

Studies of social dynamics such as those illustrated in Figures 1.1–1.5 involve a "process orientation" or an appreciation of how social variables both are composed of trends and cycles and are governed by relationships that evolve over time. Reconsider Figures 1.1 and 1.2. Someone who studied the period between the Great Depression and Reagan presidency might have warned of the capture of American institutions by the Democratic Party (Sprague 1981). But, in fact, the series shows that the Democrats eventually lost control of the House, so such a warning would have been premature.[7] Regarding Figure 1.2, suppose that one researcher studied the question of political accountability in the period between 1975 and 1979 while another analyzed the 1982–1984 period. The two scholars would reach very different conclusions. The former would infer there was no relationship between the two series, whereas the latter would conclude there was a positive relation between them. In all likelihood neither researcher would detect the possibility that the two series move together. Progress in the study of political accountability would suffer as a result.

A process orientation also is required for forecasting. Consider the Israeli-Palestinian conflict. As we will learn, forecasts based on simple extrapolations of time series such as those in the top-left panel of Figure 1.5 are unlikely to be accurate. Nor do such extrapolations provide meaningful ways to evaluate the possible effects of third party intervention. Even rigorous forecasts are often criticized for what is called the off-on-time problem. That is, these forecasts can only tell policy makers that a certain decision is likely to be made, *not when this decision will be made*.[8]

To make sound causal inferences and forecasts we require statistical tools of various kinds, which help us discover short, medium, and long term trends in

[7] Although he provides some data for the first seven decades of the 20th century, Sprague (1981) focuses on the period between 1930 and 1970. He suggests that in this particular era there is a stable level of Democratic control towards which the political system moves; he calls this level the "system telos."

[8] As an example of such forecasting, such as the expected utility model, see Bueno de Mesquita (1997, 264) and Organski (2000, 350).

our variables, as well as the causal relationships between them. We will learn how to use these tools in this book.

1.1 TIME SERIES ANALYSIS

Time series analysis explains the temporal dependencies within and between social processes. By temporal dependency *within* a social process, we mean that the current value of a variable is in part a function of previous values of that same variable. Temporal dependency *between* social processes, conversely, means that the current value of a variable is in part a function of previous values of both other variables and of that same variable. Time series analysis presumes that univariate and multivariate social systems are composed of short, medium, and long term processes of these dependencies, sometimes all at once.

Sorting out (decomposing) these different components of social processes is one of the primary objectives of time series analysis. In this way, it emphasizes the ordering of social behaviors over time. Because of its process orientation, time series analysis resists methods that treat slices of time as independent from one another especially within the same unit of study.[9] It traces the history of variables over time. In doing so, various modeling goals of time series analysis become clear. These include studying the dynamic structure of the series, investigating dynamic relationships, improving regression analysis when errors are serially correlated, and producing forecasts (Tsay, 2002).

Time series analysts test theories about social dynamics. Competing causal claims about the temporal composition of processes of representation, terrorism, suicide, crime, economic activity, and civil strife are some of what time series analysts evaluate. Hypotheses about causal relationships between the variables *in time* are what analysts test. Time series analysts eschew methods that chop time into slices and treat the resulting observations as independent. Instead, they consider that, for instance, we should be able to know more about the economic activity of a country in 2012 if we account for its level of economic activity in 2011, rather than treating economic activity in 2012 as an isolated data point. Time series analysis is a powerful approach in part because it treats these dynamics inherent in social processes as phenomena to be modeled and understood instead of as nuisances to be corrected or, worse, ignored.[10]

[9] While they are closely related, time series analysis and event history analysis are different forms of inquiry. In contrast to time series analysis, event history analyzes whether and how long it takes for events to occur (if at all). Event history can be thought of as extending discrete time models to consider time. That is, it is not only the study of whether or not an event occurs, which could be modeled with a logit or probit model, but when those events occur as well. The dependent variable in event history analysis is *duration* – how long until the event of interest occurs (Box-Steffensmeier and Jones, 2004).

[10] Time series analysis is also related to, but is distinct from, panel methods. In panel data observations are cross sections of some units (e.g., individuals, countries, states, firms) over multiple

The idea of social equilibrium is at the heart of time series analysis. As we will learn, time series analysts have well-developed concepts of fixed and moving equilibria. When social systems are at rest or in stable equilibrium we say they are static. Mathematical modeling of social systems often produces results about comparative statics or contrasting features of systems at rest. The contributions of much of rational choice theory are illustrative in this regard. A system in equilibrium is in a "steady state." That is, it settles to a certain value in the long run (fixed equilibria), or oscillates between a set of values in the long run (moving equilibria). Using the tools in this book, we can identify the equilibria of social processes, capturing their long-term temporal dynamics.

In all branches of social science, knowing these equilibria allows us to predict how systems will behave. The case of the U.S. House of Representatives is illustrative. If we can find the moving equilibrium that characterizes the alternation of power in that body, we can predict long-term swings in Democratic (Republican) Party control.[11]

In addition, time series analysis gives us insights into social systems' behaviors when they are *not* in equilibrium or are in their *out-of-equilibrium* paths. With time series methods, we can derive the paths variables will take from their initial levels to equilibria and/or how variables will return to their equilibria after the respective social system experiences a surprise one-time or repeated series of increases or decreases ("shocks") in a particular variable. On this basis we then can evaluate theoretical counterfactuals of various kinds. An example with respect to political accountability, shown in Figure 1.2, would be modeling the effect that a hypothetical, surprise, one-time surge of positive economic expectations would have on the time path of U.S. Congressional approval.

1.1.1 Contending Approaches and the Quasi-Experimental Tradition

Generally speaking, time series analysts take three approaches to model building. When they are confident they know that some variables are exogenous to others and also how long the delays are in the effects of these independent variables on their dependent variables, time series analysts use extensions of familiar time series regression methods to analyze social processes. The regression equations they employ thus are "strongly restricted." Scholars who use

time periods. Panel data, in a broad sense, refer to any data spanning multiple dimensions, usually, but not necessarily, time and space. Typically the N's (number of units) are larger than the T's (time periods). If T is "short" (that is, less than 40–50), panel data methods are more likely to be used than time series methods. See, for instance, Wawro (2002).

[11] Political scientists, sociologists, and other scholars have well-developed ideas of equilibrium. But most of these ideas are static in nature. Evolutionary game theory offers more dynamic notions of equilibrium (Samuelson, 1997; Young, 1993), but this branch of game theory is only beginning to be applied in these disciplines. For applications in political science see (Mebane, 2000, 2005).

these models emphasize the magnitudes of the effects of changes in independent variables on dependent variables. On the basis of analysis of the time series data in Figure 1.2, for instance, they produce results about exactly how large an increase in congressional approval would be produced in the short and long term by a one percentage point increase in economic expectations.

When they are less certain about the specification of their equations, time series scholars employ reduced form or "weakly restricted" models. Theory always informs the choice of variables in this kind of model. The reduced form is treated as a representation of what is an unknown, structural model. These modelers are willing to trade off quantitative precision for sounder inferences about the existence and direction of relationships. They emphasize the results of what are called Granger causality tests and the pattern of responses of their equation systems to hypothetical shocks in selected variables (impulse response functions), a method we discuss in Chapter 4.

Last, there is a quasi-experimental tradition in time series analysis. Some scholars who study social dynamics induce the functional forms of their models from their data. They devote their energy to characterizing the data-generating process (DGP) that gave rise to their data and to illuminating facts about the temporal composition and dynamic responses of variables in social systems. These efforts are theoretically informed in their choice of variables and in scholars' interpretations of their results. These results provide empirical challenges for theorists because the results must be explained by contending theories.

Our own approach utilizes each of these schools of thought. Throughout this book, we give examples of each, discussing the tradeoffs found in adopting one approach over another.

1.1.2 The Architecture of the Book

The structure of the book is as follows: The Appendix explains the calculus of finite differences, the key branch of mathematics on which much of time series analysis is based. In it, we explain in detail the idea of social equilibration and the different forms that equilibration might take. We show how political representation and other ideas from the social sciences are captured by these conceptions of equilibrium. Readers interested in working in time series methodology or the mathematical foundations for the rest of the book are strongly encouraged to read the Appendix before beginning Chapter 2. Chapter 2 introduces a particular statistical model that decomposes single time series into certain types of social memory: the autoregressive, moving average (ARMA) model. We use this univariate model to analyze homicides in the United States and Middle East politics. We then proceed in Chapters 3 and 4 to study multivariate models. Chapter 3 reviews single equation time series regression models. Again, readers may wish to consult the Appendix to better comprehend both the nature of the error processes for regression models and the implications of the functional forms of these models. Chapter 4 studies multiequation time