A Practical Guide to Time Series Analysis

Suzanna Linn
Department of Political Science
Pennsylvania State University
State College, PA 16802
Email: sld8@psu.edu

Matthew J. Lebo
Department of Political Science
Western University
London, ON N6G 0K9
Email: matthew.lebo@uwo

Clayton Webb
Department of Political Science
University of Kansas
Lawrence, KS 66045
Email: webb767@ku.edu

December 28, 2021

Contents

1 Why Should You Read This Text? **2**

2 Why Do You Need an Entire Text on Regression with Time Series Data? **5**

3 Getting your data together **6**

3.1 Choose a Sampling Window . . . 7
3.2 Choose a Sampling Interval . . . 8
3.3 Example: Sampling Decisions for Presidential Approval . . . 9
3.4 Things to Keep in Mind for Time Series Measurement . . . 11

4 Using this Book **13**

1 Why Should You Read This Text?

You cannot describe a movie by looking at a single frame of film, you cannot appreciate a poem by reading a single stanza, and you cannot experience an opera by listening to a single note; so why would you try to understand the complex processes that animate our society relying only on snippets of information plucked from single points in time? There are things you can learn from these fragments, but the stories they tell will always be incomplete. To develop a more comprehensive understanding of the world we need to know what lies between these fragments and how they are connected. Time series analysis is about building dynamic models of the dynamic processes that govern the social, economic, and political phenomena that make up the world around us.

The word dynamic is sometimes used to describe people or fields that are full of energy and new ideas. We are using the word dynamic in the more elementary sense. A dynamic process is a process characterized by change, activity, and progress. Economics, politics, and culture change over time. Many of these dynamic processes do not materialize independent of one another, but are embedded in dynamic systems; where economics can determine the direction of political events, politics can affect the kinetics of culture, and cultural changes can alter the flows of economic activity. The challenges arise because many of the rudimentary statistical techniques we are taught do not naturally accommodate the dynamic features of social processes. Many statistical tests are better suited to describing snapshots of information, frozen in time and isolated from one another. The methods presented in this book allow social scientists to capture a richer representation of the world, as a complex but comprehensible moving picture.

Time series analysis is the application of statistical tools to analyze time series data. Time series data are discrete observations ordered over time. Some of the processes captured in time series data are discrete. There is one presidential election in the United States every four years and American football teams only play one game a week. But many time series data sets are created by sampling continuous dynamic processes at discrete intervals. People are always walking around with feelings about the economy and the president in their heads but opinion time series are created by surveying the public at regular intervals; the supply and demand for different currencies is constantly changing based on the constant flow of goods and services across borders but we record the prices at specific points in time; and temperature levels are constantly fluctuating but we make decisions about what we are going to wear each day based on what we are told to expect before we leave the house. The time series analyst fits statistical models to these data to describe how the processes are changing over time and the variables that drive these processes. The models we select may not conform to the processes exactly because, after all, we can only model these continuous phenomena using the discrete data we have; but we can learn a lot about these phenomena if we can find the models that best represent the underlying processes given the data we have on hand.

The analysis of time series is not a methodology per se, but a number of approaches that have been devised for the purpose of analyzing time series data. These approaches include Box-Jenkins ARIMA (autoregressive integrated moving average) analysis (Box and Jenkins 1970), ordinary least squares dynamic regression (Banerjee et al. 1993; Hendry 1995), vector autoregression (VAR) (Sims 1980), state-space models (Kalman 1960), and others. It would be very difficult to provide a comprehensive text that covered all of these approaches. If you tried to do everything at once, the book would be sprawling and impractical. Instead, the authors of time series texts typically take one of two tacks.

Most books about time series analysis follow one of two designs: survey texts that offer short introductions to a variety of approaches or one-topic texts that take deep dives into a particular problem or approach. A survey text provides a series of chapter-long introductions to different methods. There might be a thirty-to-forty page introduction to ARIMA models in one part of the book and a treatment of VAR in another chapter. ARIMA modeling and VAR analysis are very different approaches to time series data. The models were originally developed for different purposes, the building of these models follow different philosophies, and the analyst can be required to use different tools to estimate the models; yet, a good survey text will introduce the reader to both. The authors of these texts provide the intuition behind the models along with a sense of when the different tools are more-or-less appropriate. Particularly useful survey texts for social scientists include Box-Steffensmeier et al. (2014) and Enders (2015).

The approach-specific texts are more focused. Rather than providing the reader with a general outline of different approaches, these texts are written as introductions to a specific model or as a reference on a particular method. In *Multiple Time Series Analysis* – part of the *Quantitative Applications in the Social Sciences* (QASS) series curated by *Sage Publications*, the little green books – Brandt and Freeman (2009) provide a detailed introduction to VAR models; describing the utility of the models and how they are built along with detailed examples for readers. Lütkepohl (2005) provides a much more in depth treatment, focusing less on the practical application of the models and more on the mathematical properties of the different multiple time series estimators. Both books focus on VAR analysis, but one is much shorter and much more accessible to beginners. Other examples of detailed treatments include Choi (2015) on unit root processes, Francq and Zakoian (2019) on Autoregressive Conditional Heteroscedasticity (ARCH-GARCH models), and Juselius (2006) on the cointegrated VAR model. Some of these books can be used in more advanced classes and short workshops on specific topics but many of them are most useful for applied researchers who need more detailed references than one they might find in a survey text. Our goal in writing this text is to provide a resource that provides more detail than a survey text but one that is as accessible to newcomers as short introductions and chapter-long treatments.

We provide a practical guide to applied time series analysis using ordinary least squares (OLS) regression. OLS regression is covered in most introductory statistics courses taken by graduate students and undergraduate students alike. Our expectation is that you (the reader) have taken a course that covers regression analysis before picking up this text. It may surprise you to find out that many of the most important tools in the time series analysts' tool kit are built on this basic scaffolding. Not only does our practical guide provide you with what you need to know about the application of these tools, this book gives you a bridge into the complex world of time series analysis that is built on material that you have already been exposed to, estimation and statistical inference of OLS regression models.

If you know how to estimate an OLS regression model and interpret a slope coefficient, this book is for you. If those terms seem exotic, you should probably wait to read this book until after you have taken a course on regression. This book can be used in a semester-long course on time series analysis, as part of a workshop on time series topics, or as a desk resource for social scientists who think they will (or may) need to estimate time series models for their research. We have endeavored to make the writing clear and direct and we have included a number of examples (along with the data and code you need to replicate the examples) that will allow you to teach yourself, if necessary. To our knowledge, this book is the first of its kind: One that draws connections between the various OLS-based approaches to time series analysis and provides insights into the practical application of these tools. And, in our view, it is one that is sorely needed.

There are two principles that guide our approach to time series analysis. These principles also motivated us to write this book and underlie our presentation of the material. We will take a moment to highlight these principles for two reasons: to help clarify why we think this book is necessary and to provide you with what we think are the most important principles that should govern the way you approach time series data.

Principle One: uncertainty is ubiquitous in applied time series analysis. There is a facile quality to many statistical texts. The authors present the equations for statistical estimators and describe when they should be used but fail to inform the reader about the challenges they are likely to encounter or what to do when their tests provide conflicting evidence. The standard approach to time series analysis begins with the classification of time series as stationary or non-stationary. Armed with this information, the analyst chooses an estimator, estimates the model, checks that the model assumptions are met, and interprets the results. Applied time series analysis is not so simple.

Time series analysis can be messy. The three of us often find that the dynamics characterizing the time series with which we work – shorter, intermediate frequency time series – aren't easily classified. In fact, classifying the dynamics that characterize individual time series is often a frustrating process in which multiple tests leave the analyst with competing inferences. Moreover, this uncertainty is typically hidden in subsequent analyses. Other challenges in working with time series data also engender uncertainty. What should you do when different diagnostic tests give you conflicting results about the number of lags to include in a model? What does it mean when your model fit criteria point to different dynamic specifications? What do you do when a reviewer tells you that the "right way" to do something is X but your gut tells you that you should do Y? These are the types of questions that commonly arise when analyzing time series data and advice is scarce in statistical texts about what you should do when they arise.

Much of *our* scholarship takes on these kinds of problems. We developed a bounds approach to inference that builds the uncertainty in univariate classification into hypothesis tests about the relationship between them (Webb, Linn, and Lebo 2019, 2020). The lessons we have learned lead us to stress that, whenever possible, uncertainty should be explicitly accounted for in the modeling process. When the tools to do so aren't available, analysts should endeavor to make that uncertainty a transparent part of the descriptions of analyses and reporting of results, especially if they are contradictory.

As a practical guide, this text is designed to help the reader (you) navigate each fork in the road from univariate classification through single-equation and multiple-equation regression models with the goal of building, estimating, and interpreting useful models of dynamic relationships. There are many forks on this road and we provide step-by-step (by-step) guides to help you confront the practical problems that analysts face, including:

- How to classify time series
- How to determine whether single or multiple equation analysis is necessary
- How to build an appropriate model specification including:
  - How to decide the appropriate deterministic features
  - How to model the effects of events
  - How to determine the structure of dynamic relationships, i.e. choosing the number of lags
- How to test whether your model satisfies the necessary assumptions
- How to draw detailed inferences (hypothesis testing and interpretation) from a variety of quantities.

Principle Two: theory and data interplay in important ways in time series analysis. This can be uncomfortable because social science disciplines often stress the importance of theory in model building. But theory seldom identifies the dynamic structure of relationships. We may understand consumer sentiment to be highly persistent, for example, but whether consumer sentiment varies in a predictable or undpredictable way is unclear. Theory may suggest that consumer sentiment affects public views of the president, but theory seldom – if ever – is developed enough to specify the number of lags that belong in the model. Depending on how the data are measured, different numbers of lags may be appropriate. Throughout the text, we determine the dynamic structure from the data and rely on the analyst to use theory to select the variables in the model. Even if you have a conjecture about what dynamic specification might be appropriate for a statistical model, you must subject this conjecture to formal testing before imposing it on the data. Otherwise, you risk misleading yourself and misleading your audience.

2 Why Do You Need an Entire Text on Regression with Time Series Data?

Although this book relies on linear regression analysis, the nature of temporal data demands that we approach model specification, diagnostics, inference, and interpretation differently than in the cross-sectional context. In some cases single-equation analysis will be inappropriate. At root, the complications of time series data with regression analysis lie in the fact that observations of a time series process are seldom, if ever, independent. History matters; previous values of a time series inform the present value. Put differently, the underlying process generating the data (the data generating process or DGP) exhibits time dependence. This complicates regression analysis, indeed necessitates a new nomenclature. But our goal remains the same: to estimate a statistical model that approximates the DGP so that we can make statements that generalize to the population.

In fact, time dependence has a number of implications for analysis. First, time dependence means we need to think differently about *model specification*. More often than not, the effect of a change in a regressor propagates over time. Models will need to incorporate lags. That is, we include regressors not just at time t but also at $t - s$, and we typically include lagged values of the dependent variable as regressors. This inclusion results in a dynamic regression, which, by definition, violates one of the Gauss-Markov assumptions – the zero conditional mean, or strict exogeneity, assumption. As a result, dynamic regression is not BLUE. Instead, an alternative set of assumptions are required for OLS with time series data. When met, these assumptions guarantee that OLS is consistent.

Second, the approach to *diagnostic* analysis of regression assumptions differs from the cross-sectional case. Because time dependence is a feature of time series data, we need to be particularly concerned with violations of the Gauss Markov assumption that the errors are independent. In fact, serially correlated errors pose the most common threat to consistent estimates in time series regression. But serially correlated errors are typically not a problem that needs to be "corrected for" with fixes like weighted least squares or robust standard errors. In fact, applying these "solutions" does not address the underlying problem, omitted variable(s), and obfuscates important features of dynamic relationships. Instead, serially correlated errors should be taken as evidence indicating dynamic misspecification – the omission of relevant lags from the model – and indicate the need to reconsider the lag structure of the model.

Third, different patterns of time dependence require different strategies for estimation and *inference*. Time series analysts need to characterize the patterns of time dependence that describe the data. In particular, analysts need to determine whether their time series are best characterized as stationary processes, which have a constant mean, variance, and autocovariance over time, or whether they are not stationary. Unit root processes, those that contain a random trend, present particular challenges since they are prone to the spurious regression problem. That is, independent unit root processes appear related when they are not. Even when they are related, model specification requires more care and critical values for hypothesis tests will tend to be non-standard.

Fourth, *interpretation* of dynamic regression models requires more than interpreting a single coefficient. A dynamic regression model provides information about how variables affect one another in the short-, medium-, and long-term; how changes in a regressor play out in a regressand over time; how those effects decay; and what types of equilibria characterize long-run relationships, if they exist. But this added "complication" is really an advantage, enabling us to provide rich interpretations of dynamic relationships.

Finally, time dependence means that time series analysts need to consider whether *single-equation or multiple-equation analysis* is necessary for valid inference. Single-equation analysis requires that no information about the relationship of interest is lost by ignoring the model generating the regressors, i.e., that we have weak exogeneity for the parameters of interest. If it is, we violate the weak exogeneity assumption needed for consistent inference. Most texts provide limited discussion of this assumption, instead treating it as a decision the analyst has made in advance of the analysis. Indeed there are a number of misconceptions about the conditions that permit single-equation analysis. So much so, that we devote a chapter to this topic. When weak exogeneity is violated, multiple-equation analysis provides an alternative means of describing dynamic relationships. But this modeling approach puts additional demands on the data and the analyst.

These complications mean time series analysts require a different *nomenclature*. The terms stationary and unit root processes figure prominently in the vocabulary, and it will become apparent quickly that a clear understanding of these terms is essential to robust time series analysis. But other terms like persistence, integration, deterministic and stochastic trends, structural breaks, cointegration and different varieties of exogeneity, to name a few, will also play an important role in model-building. Interpretation of the results from a model will require expanding your vocabulary to include impact multipliers, long-run multipliers, and impulse response functions. These terms likely don't mean a lot to you now. That's fine. We will introduce these terms and explain how and when they are relevant in the chapters to come. We also provide a glossary of terms.

In short, the opportunities that come with time series data also come with a number of challenges that are not faced in a cross-sectional context. We will walk you through those challenges, but first let's get your data in order.

3 Getting your data together

To make the best use of this book, you need a time series data set and a research question. Producing a time series data set can be challenging. In addition to finding sufficient data over time, you need to decide what unit of time to analyze and the period of time the analysis will cover. This section provides a brief roadmap that will help you get your time series data together as you prepare to conduct a time series analysis.

You begin your time series data collection effort the same way you begin any other statistical analysis, by posing a question. What do you want to study? Or, to be more concrete, what are the independent and dependent variables in your analysis? Many time series texts begin with a discussion of the univariate properties of time series: trend, seasonality, and autocorrelation. This can give the impression that the focus of your measurement decisions begins and ends with the properties of your dependent variable. To be sure, the univariate properties of your dependent variable are important. We will discuss them in Chapters 2 and 3 of this text but your dependent variable should not be the only thing at the front of your mind in the early stages of a time series analysis. Before you collect data, you need to know what your variables of interest are, how they might be related, and when they might be related. These questions are essential to the measurement decisions you will have to make: sampling window and sampling interval.

3.1 Choose a Sampling Window

There are two time span questions to consider. First, what is the *ideal* sampling window for your data? A *sampling window* is the period of time over which your data are collected, the period of time between the beginning and end of a time series. The word *ideal* is italicized to highlight the theoretical nature of this proposition. In the absence of any practical limitations, what is the period of time that would best allow you to observe the relationship you are interested in analyzing? If you are interested in the consequences of cyber-warfare on interstate conflict, it may not make sense to include militarized interstate disputes (MIDs) from the 19th and 20th centuries, even though the data collected as part of the international correlates of war (ICOW) project include observations as far back as 1816 (Palmer et al. 2015). If you are interested in the effects of presidential elections on consumer sentiment and presidential approval, you will need to consider how many elections to include in the data and whether the relationship changes across elections.

This final point, whether the relationship is constant or changes over the sampling window, bears emphasis. A conventional time series regression model assumes that the parameters of interest are constant across the sampling window. As with all regression models, a partial slope coefficient gives the average conditional effect of the regressor on the regressand across all the observations. In cross-sectional analysis, it has become increasingly common for researchers to include interaction terms to model how the marginal effect of one regressor may change over the levels of another. Marginal effects can change over time as well. A failure to account for these parameter changes can produce omitted variable biases. Change is the only constant in human behavior, so the longer your sampling window the more likely one or more of your parameters of interest are to change. You should be cognizant of this possibility and choose your sampling window accordingly.[1]

The next question about the sampling window is as important as the first: what sampling window is *available* to you. Where the question about the ideal sampling window is theoretical, this question raises practical considerations. You will often find that the practical bounds of your analysis are more limited than you might like.

In some rare circumstances, you may have the opportunity to collect the ideal data set; letting theory guide all of your measurement decisions. Through archival research or through field surveys, you will not only be able to collect your independent and dependent variables of interest, but also all the control variables you will need for your analysis. Should such a scenario come to pass, you are fortunate indeed. In most cases, we are not so lucky. You will often find yourself cobbling together data from different sources; pulling information from different data sets. In these circumstances, your sampling window will be dictated by the past work of others. You may need to make compromises in your measurement strategy based on what data are available for your variables of interest; you may also need to make hard decisions about the variables to include in your model as controls.

The sampling window for your time series analysis will be the same for all the variables in your analysis – those observations for which you have data for every time point for each series in your analysis. Regardless of what information you have available for each series, your window is only as wide as your shortest series. Your software will automatically remove observations for which values are missing. The intuition is simple. A partial slope coefficient is estimated as the covariance between the regressor and

[1]There is a caveat to this dictum, of course, for analysts interested in studying how relationships change over time. One can use interaction terms or specialized change-point models to test hypotheses about parameter changes over time. In these instances, the ideal sampling window is one that includes a sufficient number of observations before and after the change that the change in parameters can be discerned and the long term consequences of the change evaluated. We will return to the issue of parameter constancy later in the text; for now, we only suggest that you keep the idea of parameter constancy in mind when formulating potential sampling strategies.

the regressand less the covariation amongst the regressors. If observations are missing for some control variable, these covariations cannot be estimated. When you make choices about control variables, be sure that missingness in some variable does not remove relevant variation through casewise deletion. If there is a substantial amount of variation missing for a relevant control variable, you will need to collect more data or find a suitable substitute.

3.2 Choose a Sampling Interval

The next two questions pertain to your sampling interval. Similar to the decisions you make about your sampling window, you begin by identifying the *ideal* sampling interval for your analysis. When you go to collect your data, you will likely have to make compromises between what is theoretically ideal for your analysis and what is practical given your time and resources.

The *sampling interval* is the discrete period of continuous time you choose to measure your phenomenon of interest. Time is continuous because times is always moving but we measure time in discrete intervals: days, weeks, months, and years. The interval for a time series does not change over the sampling window and you create your time series by recording one observation per period (per interval).

The ideal sampling interval is the one where you are most likely to detect your relationship of interest. We are political scientists, so we tend to pose the question the following way: what is the rate at which your politics happen? The question seems straightforward enough but it is not always easy to answer. The wheels that propel interesting social, economic, and political phenomena are constantly turning, but we may not be able to observe meaningful changes in these phenomena continuously. Take changes in the partisanship of the United States Supreme Court. Every year the partisanship of the court has important effects on politics; determining which cases will be considered, how cases will be decided, and which laws will be enforced and overturned. But the Court's partisanship changes infrequently.

This lack of variation has important consequences. The partial slope coefficients we use to make inferences in multiple regression analyses are covariances ($cov(y, x)$) divided by a variances ($var(x)$), net $cov(y, z)$ and $cov(x, z)$ for sets of controls Z. In a model used to analyze the partisanship of the Court, every period where your variable of interest (x) varies but court partisanship (y) remains constant militates against detecting an effect; two variables cannot covary if one of the variables doesn't vary. For obvious reasons, choosing a monthly sampling interval might be impractical for such an analysis. At the other end of the spectrum, access to high frequency social media data allows us to capture Twitter activity on a second-by-second basis. But if you are looking at the relationship between some political variable and social media data, it may not make sense to measure both phenomena second-by-second if you do not believe that members of congress or members of the public update their political beliefs second-by-second. When you are thinking about the ideal sampling interval, you should use what you think you know about the relationship of interest to identify the measurement strategy that would give you the best time series to test your hypotheses.

Like the choice you made about your sampling window, there are also practical considerations which force important compromises between what is theoretically ideal and what is plausible for your applied analysis. You now have to identify what sampling interval is *available* to you. Again, without unlimited time and resources, your sampling choices may be dictated by the work done by others. Economists at the United States Federal Reserve measure exports to and from China on a monthly basis but you may need to measure exports on an annual basis if you want to include more than two countries in your analysis. Ideally, your study of US presidential approval and public perceptions of the economy would go back to the beginning of polling on approval. Yet, the Institute for Social Research (ISR) at the University of

Michigan only began its survey of consumer sentiment in 1960 and did not do so monthly until 1978. You will need to choose between a larger time window (1960-2021) or a higher-resolution sampling interval (monthly instead of quarterly).

The sampling interval should be consistent for all the variables included in the analysis. The sampling intervals of control variables are important here as well. For example, you may be able to measure your outcome and independent variable of interest at a quarterly level but can only measure a control at an annual level. You may want to resolve this problem by simply recording the annual observations of this control as though the variable were quarterly; by inserting the annual observations four times, one for each quarter. This has consequences – as before, every period where your variable of interest varies but the control variable does not is a period where your estimator is not achieving your intended end. In this case, if closely accounting for covariation between your variable of interest and the annual control variable is important, you will want that accounting to be complete. Or, you might try to turn a quarterly measure into a monthly one by assigning the values you have to January, April, July, and October and then interpolating data in equal steps for the months in between. This has consequences too – it imposes a dynamic structure on the data that might affect modeling later.

In some cases you will have the option of aggregating higher frequency data to a lower one. This is preferable to the situation just described, but it raises additional measurement questions that need careful consideration and present new opportunities to make critical errors.

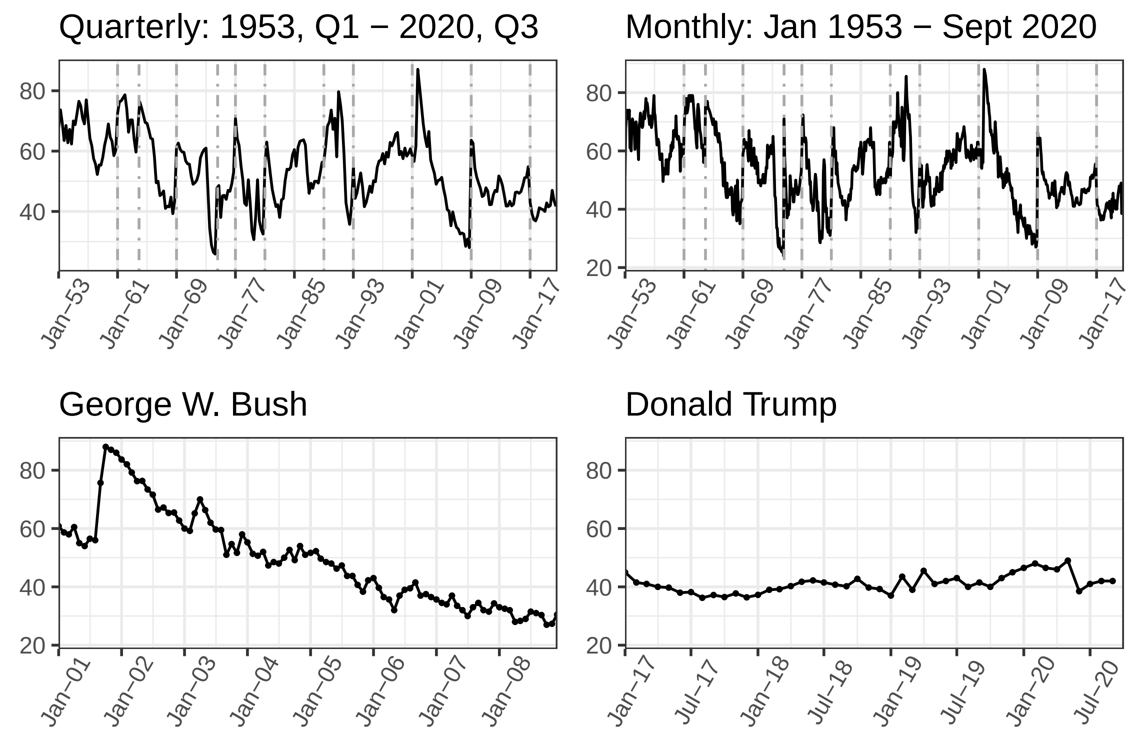
Temporal aggregation refers to the process of grouping data from higher frequency intervals to lower frequency intervals. There are many tacks you can take when aggregating time series data. Consider the process of moving from a monthly sampling interval to a quarterly sampling interval. You could average the observations across months, use the observation from the first or last month of the quarter, take the lowest or highest value from the quarter, or try a more complicated algorithm. Your choice can have implications for your analysis. If your process of interest is punctuated by major events and sudden changes in direction, removing variation through averaging may wash out the effects you are interested in observing. If your process is relatively gradual outside of a handful of outliers, taking the highest value in each period may produce significant noise in your data. Freeman (1989, p. 93) advises researchers to think hard about the "natural time unit of our theories" and consider how one's conclusions might change if different decisions about sampling and aggregation were made.

3.3 Example: Sampling Decisions for Presidential Approval

You can get a sense of some of these issues by examining what we think of as the "mother" of all political time series, presidential approval in the United States. Figure 1 presents four graphs: The top two panels show data going back to the beginning of Eisenhower's presidency in quarterly (left) and monthly (right) data with dashed vertical lines demarcating presidencies. Each point in the series represents the average of all the Gallup surveys in that month or quarter. The monthly time series has more observations but it covers less historical time and because surveys were not faithfully administered back to President Eisenhower, the monthly series includes many interpolated time points.

The series are in many ways quite similar, but a closer look suggests that the quarterly series evolves more smoothly. The monthly series contains more "bumps and wiggles," in part because of random sampling error that is not "smoothed" through aggregation to the quarterly level. The quarterly series is thus more persistent, a feature that has implications for the number of lags that might be needed to model its behavior. The choice of which of these levels of aggregation to use depends on the particular research question and on which other time series will be modeled alongside approval. For example, using the monthly series to increase the sample size would not make sense if other variables are only available quarterly or yearly. Note, too, that the level of aggregation has implications for how you think about (and model) the effects of any explanatory variables. In particular, an effect that proves significant in monthly data – e.g. consumer sentiment in one month affecting approval the following month – might appear to be contemporaneously related at the quarterly level.

Figure 1: Gallup Presidential Approval Ratings



The behavior of approval also presents stark differences across sampling windows. The bottom panels of Figure 1 show monthly data separately for the Bush II (left) and Trump (right) administrations. Note the differences: George W. Bush's approval benefited from two large rally effects – the first follows the terrorist attacks of 9/11 and the second occurs with the onset of the War with Iraq. Both rallies persisted for some time but, eventually, the president's popularity followed what Mueller (1970) referred to as inexorable decline. In many places in this book we will talk about "structural breaks" and it is noteworthy here that the 9/11 attacks – the biggest spike in the history of presidential approval – may be classified as either a temporary increase or a structural break depending upon the sampling window used.

In contrast, President Trump's tenure is remarkable for its lack of volatility. With strong partisanship, neither record-low unemployment nor impeachment(s), nor pandemic budged his approval out of the low-40s for very long. Thus, hypotheses that might prove correct during one administration might not in others – what is true in one window of data may not hold in a wider, narrower, or different window.

3.4 Things to Keep in Mind for Time Series Measurement

The decisions you make when getting your data set together will affect every stage of the analysis; so you need to think carefully! In the concluding part of this section, we want to highlight some things to keep in mind when making sampling decisions. These should help you clarify your decision-making.

You need a sufficient number of observations to conduct a time series analysis. A common mistake made by people who are new to time series analysis is asking too much from too few observations. You may have 35 years worth of data but if it is all measured annually, you may not have enough information to build a dynamic model. Or, more to the point, you may not have enough information to build a good dynamic model. When we move from a static regression model – a model that does not contain any lags – to a dynamic regression model – a model that includes lags of the regressors, the regressand, or both – we are moving from the world of *unbiased* and *efficient* estimators to the world of *consistent* estimators. Knowing the difference is important for avoiding mistakes.

Inferential statistics involves the use of sample statistics to make inferences about population parameters. If you had a data set that included the entire universe of cases that could be selected, you wouldn't need inferential statistics; you could just calculate your quantities of interest and report them. Most of the time you don't have the universe of cases, you have just a subset: your sample. In applied statistics we are often interested in using some sample statistic (e.g., $\hat{\beta}$) to make inferences about the parameters that govern the relationship of interest (e.g., β). The quality of these inferences depends on the properties of the estimators which, in turn, depend on the underlying model assumptions and the features of the data.

Unbiasedness and efficiency are small sample properties of estimators. Both depend on the idea of repeated sampling. If we draw 1,000 samples of equal size (say $T = 100$) from a population, calculate a sample statistic for each sample i ($\hat{\beta}_i$), and plot the 1,000 $\hat{\beta}_i$ values, we would have a sampling distribution of sample $\hat{\beta}$ values. An unbiased estimator is one where the mean ($\bar{\beta} = \sum_{i=1}^{1000} \hat{\beta}_i/1000$) of this sampling distribution is equal to the true population parameter ($\bar{\beta} = \beta$). Any individual $\hat{\beta}_i$ can be higher or lower than the population parameter β because of sampling error. But if $\bar{\beta} \neq \beta$ it is because the estimator is systematically over- or under-estimating the population parameter. That is, the estimator is biased. Efficiency pertains to the standard deviation of the sampling distribution of sample statistics. An estimator $\hat{\beta}$ is efficient compared to another estimater, $\tilde{\beta}$, when the standard deviation of the sampling distribution for $\hat{\beta}$ is smaller than the standard deviation of the samling distribution for $\tilde{\beta}$. Other things equal, we would like our estimators to be unbiased and efficient. These are the primary virtues of OLS we are taught in introductory regression courses.

Consistency is the large sample analogue to unbiasedness and efficiency – a consistent estimator is unreliable in small samples. To say that an estimator is asymptotically unbiased or efficient implies that a consistent estimator might be biased and / or inefficient in small samples but that the mean of the sampling distribution of a consistent estimator converges to the population parameter as the sample size increases. The standard deviation of the estimator shrinks as the sample size increases, as well. The key to using a consistent estimator, then, is to have enough information so that it is no longer biased nor inefficient.

The models we describe in this book, and almost all other time series models that we are aware of, are consistent estimators – not unbiased and efficient estimators. When you include a lagged value of your dependent variable in the model, y_{t-1}, you can no longer rely on the unbiasedness and efficiency of the OLS estimator. More data is required to get good estimates of the coefficient on y_{t-1} than is required to get similarly good estimates for the coefficient on x_t from a static model without lags. Thus, you should be reticent to conduct a time series analysis without a sufficient number of observations and you should be skeptical of applied dynamic models with short samples.

How big is big enough? Unfortunately there is no simple answer. In might depend on the type of model you are trying to estimate and on the features of your data. You may be able to produce a reliable estimate of a coefficient that describes the relationship between y_t and y_{t-1} (call it $\hat{\rho}$) with a sample of $T = 50$ if the population parameter is $\rho = .5$, but the reliability of the estimate could be lower if $\rho = .99$, and even more data could be required to conduct a reliable test that $\rho \neq 1$.

Further, the more variables in a model, the more time points you will need. Keele, Linn, and Webb (2016) reference Babyak (2004) on this point, noting that one should fit one parameter for each 10 observations when the data are independent and identically distributed. With time series data, the lack of independence necessitates more than 10. As ρ approaches one, for example, substantially more observations will be needed for each parameter estimated. Throughout this book, as we introduce new estimators and tests, we will highlight the sample sizes required for different procedures. Suffice to say, not every approach is practical for every time series data set that you might have. In some cases you will need to collect more data and in others you may need to find a different approach.

You can also have too much information. This seems to fly in the face of what we are normally taught to think about sampling: that more is better. The prudence of this point relates to the concept of statistical power which can be a double-edged sword for the time series analyst.

The power of a statistical test is the probability it will reject a false null hypothesis. A tests's power depends on the significance criterion, the reliability of the sample results, and the "effect size," the degree to which the phenomenon exists (Cohen 2013, 4). The significance criteria ($\alpha = .01$, $\alpha = .05$, and $\alpha = .10$) are dictated by the academic communities within which we operate, so there is no need to perseverate on them here. The reliability of the sample results and the effect size are more significant.[2]

The reliability of a sample value is a measure of how closely it approximates a population value (Cohen 2013, 6). Given that we don't usually know the population value, we have to estimate it. For multiple regression models, the reliability of the sample values (the regression coefficients $\hat{\beta}$) are estimated as the standard errors of the regression coefficient ($SE(\hat{\beta})$). These standard errors are proportional to the standard error of the regression, $\hat{\sigma}_\varepsilon = \sqrt{\sum (y_i - \hat{y}_i)^2/(n-k)}$. This formula highlights that the standard error of the regression, and hence the standard errors of the coefficients, decrease as the sample size in the denominator increases.

The formula for the standard error of regression has spawned a problematic bit of folk-wisdom that perpetuates a misunderstanding about statistical power. If the standard error of the estimates decrease as the standard error of the regression decreases, so the fable goes, then increasing the sample size necessarily increases the statistical power. You will often here smart people say things like, "well no wonder everything is significant in those dyadic models, the sample size is huge!" This logic ignores another important part of the formula, the numerator.

Statistical power depends on the signal-to-noise ratio in the sample. The standard error of regression increases as the ability of the model to accurately predict ($\hat{y}_i$) the values of the outcome variable ($\hat{y}_i$) falls. If the hypothesized effect is constant $\hat{\beta}$, an assumption implicit in any model that does not explicitly accommodate the possibility that it is not, then adding more data only increases the statistical power of the test if the amount of signal in the data (the effect) increases proportional to the amount of noise in the data (the variability). If increasing your sample size increases the number of irregular events that you need to account for or introduces the possibility that the signal will change because $\hat{\beta}$ changes over the sampling window, then increasing the sample size (T) will decrease the power of your test. Increasing the sample

[2] Pun intended.

size can make your estimates sharper in some circumstances but it may also cut against your ability to detect an effect in others.

This practical problem of sampling and power is why taking the time to consider the *idealized* version of your sample is so important. Before you collect your data, think hard about how you can maximize the ratio of signal to noise in your sample. You may be able to collect stock market data back to January 2, 1960 for Ford Motor company, but if you are interested in detecting the effect of a government loan to Ford in 2009, it may not make sense to choose a sampling window and sampling interval that requires you to control for fundamental changes in the stock market between 1960 and 2008 along with every event that ever affected the profitability of the firm over that period of time. Sure, it is possible, but it is wildly unnecessary and the model would be far from parsimonious.

In sum, you need to think before you sample. You need to select a sample that is large enough for your analysis but not so large that it makes your analysis more difficult. This requires making informed choices about your sampling interval and window which may require you to make compromises between what is theoretically ideal and practical in practice. These complex decisions and uneasy compromises are critical to the principles we outlined above.

Time series analysis involves a lot of uncertainty and theory is unlikely to completely resolve it. To build the best model, you will both leverage what you think you know about the process that produced the data and poke at the data you have for empirical clues to its origins. In the chapters that follow we will describe the best practices that have been developed to draw the right balance between theory and data in the empirical specification of our theoretical models.

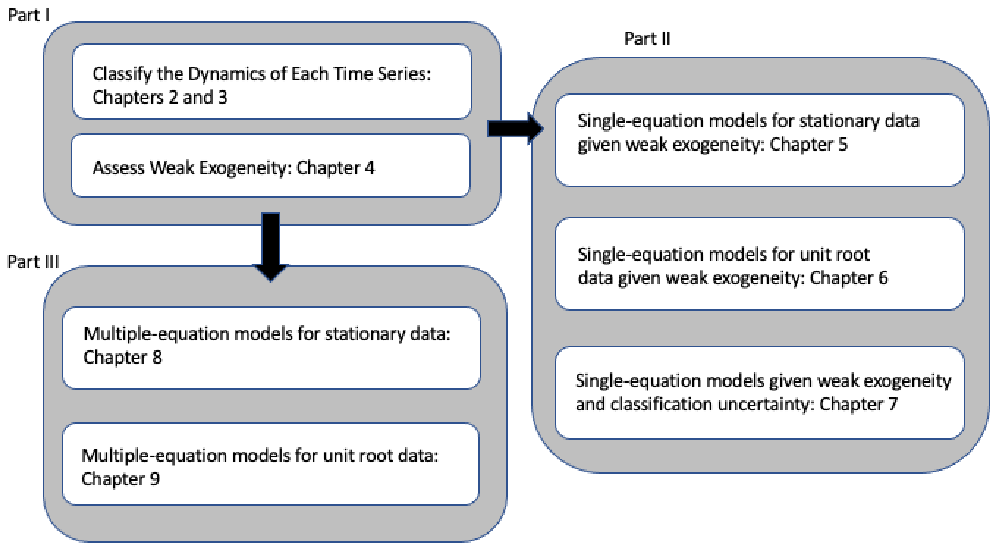
4 Using this Book

Once you have your data together, you are ready to conduct the analyses we cover in this book.

Figure 2 presents a map of the book. The text is broken into three parts. Part I outlines the preliminary questions that must be resolved before analysis. Are the data stationary or non-stationary (Chapters 2 and 3) and is weak exogeneity a valid assumption (Chapter 4)? The conclusions you draw from the analysis covered in Part I dictate how you conduct analysis of the relationships between time series. Readers who conclude that weak exogeneity is a reasonable assumption will find the single-equation methods covered in Part II appropriate for their analysis. We divide Part II of the text into regression models for stationary data (Chapter 5), unit root data (Chapter 6), and for cases in which you are unsure of the nature of the univariate dynamics (Chapter 7). Which chapter is relevant will depend on how you classify your time series in Part I. If you have determined that weak exogeneity is not a reasonable assumption, Part III covers the multiple equation models you need to use for stationary data (Chapter 8) and unit root data (Chapter 9). Part III ends with a discussion of a principled approach to time series analysis (Chapter 10).

Of course, there is often uncertainty that makes the roads to follow unclear in a given analysis. You may find that efforts to characterize your time series as stationary or unit root processes leaves you unsure about the best path forward. You might similarly be unwilling to state categorically that the variables you care about are weakly exogenous, even if you believe it plausible. Such is the nature of applied time series analysis. But the tools we provide will help you to navigate these dilemmas. Throughout, we use examples that highlight the problems you are likely to encounter. Data and code for the applications in this book can be found at `XXXXXXX`. While most of the analysis in this book can be conducted in Stata and R, if you anticipate being a regular user of time series methodology, you might find purchasing a statistical

Figure 2: Map of the Book



package developed specifically for time series analysis a worthwhile investment. Eviews, RATS, and CATS for RATS are packages that provide extensive tools for time series analysts that can make your work easier.

Time series analysis is an enormous field. Nevertheless, the contents of the following chapters should provide a reasonably comprehensive foundation upon which to learn more advanced techniques. If our collective experience with time series data has taught us anything, it is that learning new ways of approaching time series analysis, and statistical analyses in general, never ends. For our part, the chapters that follow are a good place to begin.

References

Babyak, Michael A. 2004. "What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models." *Psychosomatic medicine* 66(3): 411–421.

Banerjee, Anindyaerjee, Anindya, Juan Dolado, John W. Galbraith, and David F. Hendry. 1993. *Co-Integration, Error Correction, and the Econometric Analysis of Non-Stationary Data.* Oxford: Oxford University Press.

Box, George E.P., and Gwilym Jenkins. 1970. *Time Series Analysis: Forecasting and Control.* San Francisco, CA: Holden-Day.

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 John R. Freeman. 2009. "Modeling Macro-Political Dynamics." *Political Analysis* 17(1): 113–142.

Choi, In. 2015. *Almost All about Unit Roots.* Cambridge University Press.

Cohen, Jacob. 2013. *Statistical Power Analysis for the Behavioral Sciences.* New York, NY: Academic Press.

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

Francq, Christian, and Jean-Michel Zakoian. 2019. *GARCH models: structure, statistical inference and financial applications.* John Wiley & Sons.

Freeman, John R. 1989. "Systematic Sampling, Temporal Aggregation, and the Study of Political Relationships." *Political Analysis* 1(1): 61–98.

Hendry, David F. 1995. *Dynamic Econometrics.* Oxford: Oxford University Press.

Juselius, Katarina. 2006. *The cointegrated VAR model: methodology and applications.* Oxford university press.

Kalman, Rudolph Emil. 1960. "A new approach to linear filtering and prediction problems." *Journal of Basic Engineering* 82(1): 35–45.

Keele, Luke, Suzanna Linn, and Clayton McLaughlin Webb. 2016. "Treating time with all due seriousness." *Political Analysis* 24(1): 31–41.

Lütkepohl, Helmut. 2005. *New introduction to multiple time series analysis.* Springer Science & Business Media.

Mueller, John. 1970. "Presidential Popularity from Truman to Johnson." *American Political Science Review* 65(March): 18–34.

Palmer, Glenn, Vito d'Orazio, Michael Kenwick, and Matthew Lane. 2015. "The MID4 dataset, 2002–2010: Procedures, coding rules and description." *Conflict Management and Peace Science* 32(2): 222–242.

Sims, Christopher R. 1980. "Macroeconomics and Reality." *Econometrica* 48: 1–48.

Webb, Clayton, Suzanna Linn, and Matthew Lebo. 2019. "A Bounds Approach to Inference Using the Long Run Multiplier." *Political Analysis* .

Webb, Clayton, Suzanna Linn, and Matthew Lebo. 2020. "Beyond the Unit Root Question: Uncertainty and Inference." *American Journal of Political Sciencee* 64: 275–292.