CAMBRIDGE
Cambridge University Press & Assessment
978-1-009-39629-5 — Time Series for Economics and Finance
Oliver Linton
Excerpt
More Information

1 Introduction

This book is about time series data, that is, data that are recorded in sequence. Time series have some special features due to the ordering in time. Our analytic framework is to suppose that the outcomes we observe are realizations from some population of random variables or stochastic process. For each t in some set $\mathcal{T}$, $y_t : \Omega \to \mathbb{R}$ is a random variable with realization $y_t(\omega)$, where $\omega \in \Omega$ is some underlying sample space. For each ω, the set $\{y_t(\omega),\ t \in \mathcal{T}\}$ is called the **sample path** or trajectory. For each $t \in \mathcal{T}$, the collection $\{y_t(\omega),\ \omega \in \Omega\}$ is the set of potential outcomes of the random variable y_t, of which we observe precisely one. We may define the distribution of the random variable y_t and its moments for each t with respect to $\omega \in \Omega$. The key thing here is to define the joint distribution of each sequence $y_{t_1}, \ldots, y_{t_n}$, where $t_i \in \mathcal{T}$ for $i = 1, \ldots, n$, but this requires some detail about the relationship between the random variables.

We may have observation times that are not equally spaced. For example, stock markets are closed at weekends and during holidays. For some data, such as intraday financial transaction prices, the observation times themselves can be considered the outcomes of some stochastic process, which can interact with the observations themselves. For the most part we deal with equally spaced observations where the observation times are assumed without loss of generality to be integers. We also for the most part deal essentially with the case where the random variables are continuous, meaning they take values in the real line rather than in a more restricted domain. There are special issues to do with, say, binary or integer-valued time series and we will consider these toward the end.

We may observe a trajectory or orbit of values $\{y_1, \ldots, y_T\}$, which is one draw from the stochastic process, but under certain conditions we are able to use this sample to learn about the population properties of the process, which concern all $\omega \in \Omega$. The analysis usually consists of modelling, that is, describing the laws of motion of the time series in a parsimonious fashion reflecting subject knowledge and data features, estimation of the parameters of the model based on the sample of data, testing hypotheses about parameters of the model, and forecasting future values of the series. The analysis may be motivated by the quest for understanding, or there may be a concrete objective to evaluate the effects or potential effects of a government policy or some other intervention.

Many time series books start by talking about the additive decomposition of a series y_t, where y_t may be some transformation of the raw data (such as the logarithm or logistic), into components, that is,

$$y_t = T_t + S_t + C_t + E_t, \tag{1.1}$$

where T_t is the trend component, S_t is the seasonal component, C_t is the cyclical component, and E_t is the error term. We have to define what makes T a trend, what makes

S a seasonal component, and what makes C a cyclical component, otherwise this is a meaningless decomposition; we will take this up later. Roughly speaking, a trend is a persistent upward or downward movement, a seasonal component is a regular periodic variation (of known period) coinciding with specific calendar features, such as days of the week, months of the year, and so on, and a cycle is a more nebulous concept involving quasi-periodic behavior of unknown horizons. In economics, there are several named cycles: the Kitchin cycle of around 3–5 years, the Juglar cycle of around 7–11 years, the Kuznets cycle of around 15–25 years, and the Kondratiev wave of 45–60 years. The National Bureau of Economic Research (NBER) dates US business cycle peaks and troughs, according to their definition.[1] In climate science, there are many known cycles of varying lengths, from 30 days to thousands of years, such as the El Niño southern oscillation (around 2–7 years) and the glacial cycles, the Brückner–Egerson–Lockyer cycle of length 30–40 years, and so on.

Economists often want to work with "deseasonalized" data, which amounts to estimating the component S_t and subtracting it from y_t; central banks and many others have developed sophisticated methodologies to do this. In some cases, economists want to work with "detrended" data, to abstract from whatever is causing the growth over time in a variable and to focus on the short-run fluctuations, which may be influenced by macroeconomic policy. Finally, there are some applications where it is common to work with "decycled" data. For example, cyclically adjusted government budget deficits are favored by some economists as better reflecting the true balance in public finances, taking account of automatic adjustments that occur through a business cycle. Private sector companies are not allowed to decycle their earnings and costs in their public announcements; there are accounting tools at their disposal that allow them to smooth earnings and costs to some extent. Campbell and Shiller (1988) introduced the cyclically adjusted price to earnings (CAPE) ratio that tries to remove the short-term cyclical variation in announced earnings to give a more appropriate measure of the state of the stock market. In climate science, the focus has been on the trend part of the process, or indeed whether there is a trend and how big it is; this trend is usually called the anomaly and defined as the departure of the temperature from a long-term average such as the twentieth-century global average temperature.

The decomposition in (1.1) raises identification issues: Can we distinguish a trend from a long cycle? Can we distinguish a seasonal component from a trend? To implement this we need further modelling assumptions. One approach is to use deterministic trends and seasonal components. A second approach, called structural time series modelling, is to use random walks over and over and over again to model the trend and the seasonal. Both approaches involve applying different linear transformations to the data in sequence to deliver the separate components.

I view the decomposition in (1.1) really as a metaphor, reminding us that the key features of a time series are its trend (or lack thereof) and its seasonality (or periodicity more generally); we may have a more complicated and holistic model. We first show a few datasets to illustrate some of the issues.

[1] That is, two consecutive quarters of negative GDP growth; see www.nber.org/research/business-cycle-dating.

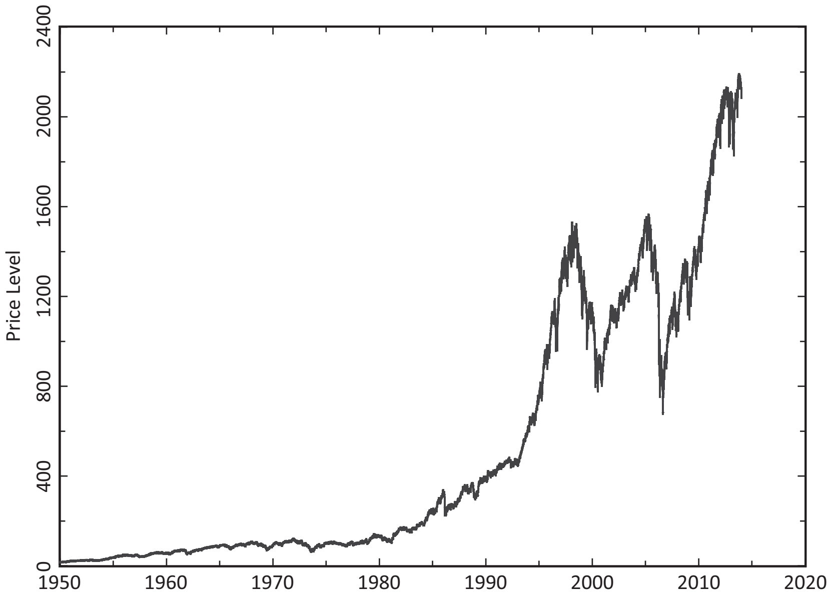

Figure 1.1 S&P500 daily stock closing price.

We next consider, in Figures 1.1–1.15, some empirical examples. The datasets are available from www.cambridge.org/lintontimeseries and are detailed in Appendix D.

The first dataset, `sp500`, is the Standard and Poor's (S&P) 500 index level, Figure 1.1, which shows a substantial upward trend with several visible reversals of nontrivial duration (bear markets). The key feature here is the trend.

The next series, Figure 1.2, is the daily observed short-term interest rate on US government securities, specifically the one-month T-bill, that is, the contract length is one month but the observation frequency is daily, taken from the dataset `ffdaily`. This series has some wandering up and down around its mean level but does not appear to have a very strong trend in comparison with stock prices; the level of this series was for a while very close to zero, which is an effective lower bound.

We next show the Chinese yuan / US dollar exchange rate from the series `cnyusd` along with its percentage change or return (Figure 1.3). The rate was fixed until 2005 and then effectively fixed again between 2008 and 2010, but otherwise shows some upward and downward variation in a modest range with no substantial trend in evidence. The return series (the time difference of the logarithm of price) shows the variation in more detail, as well as the occasional big movements associated with the depegging and other events.

The next series, Figure 1.4, is the daily closing price of the VIX futures contract, from the series `VIX`. This series also appears not to have a strong trend but rather certain cycles or waves of up and down motion along with occasional big moves such as during the financial crisis and at the beginning of the COVID-19 pandemic.

We next show the US unemployment rate (the percentage of the work force currently unemployed), which is reported monthly, taken from the series `UNRATENSA` (Figure 1.5).

Figure 1.2 Daily one-month maturity T-bill rate.

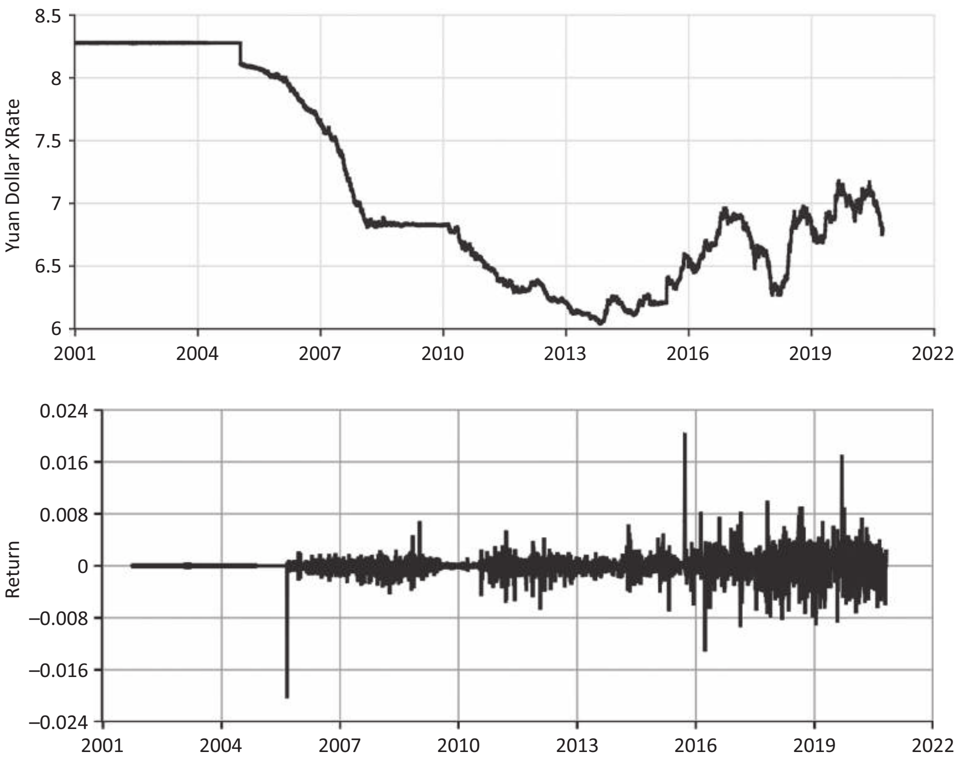

Figure 1.3 Daily yuan/dollar exchange rate and percentage change.

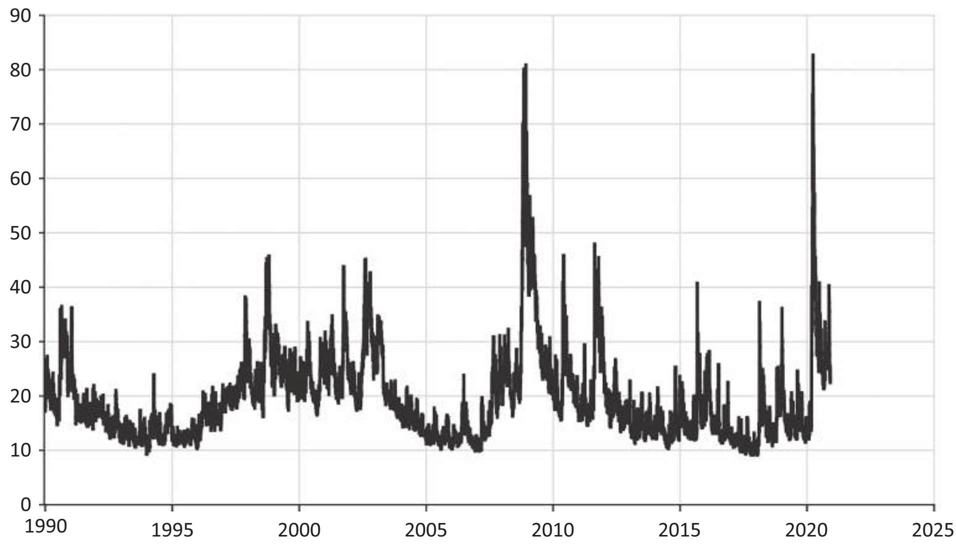

Figure 1.4 Daily level of VIX, 1990–2020.

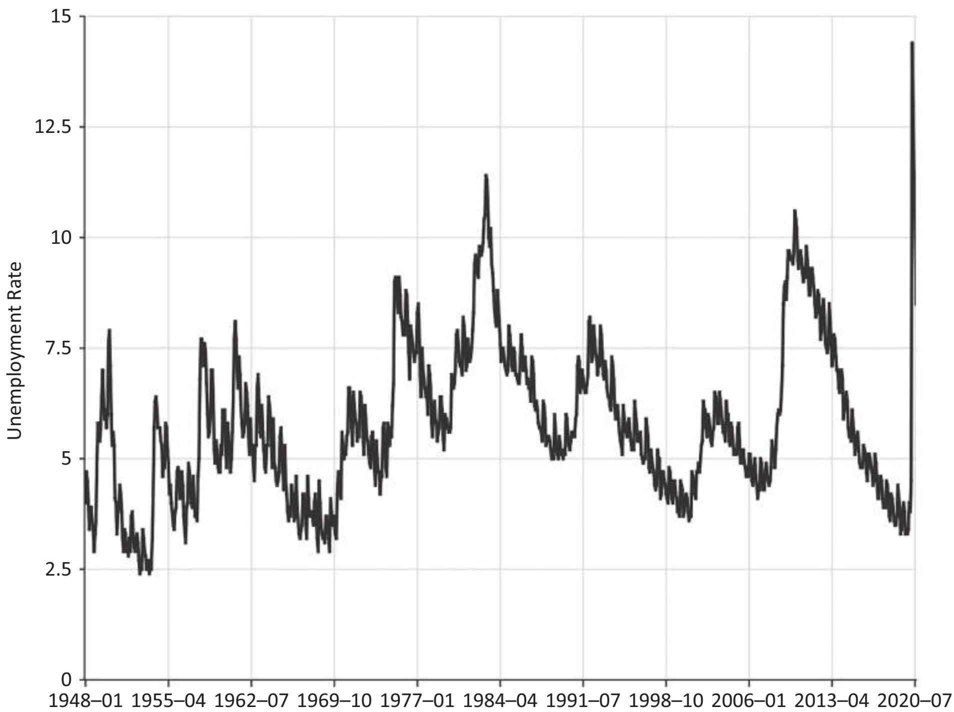

Figure 1.5 US monthly unemployment rate, not seasonally adjusted.

This series shows the recent COVID spike, along with other boom and bust periods. The series has a pronounced seasonality, because unemployment is lower in the summer and around Christmas time, which is why it is more common to show the seasonally adjusted rate.

Figure 1.6 US monthly industrial production, not seasonally adjusted.

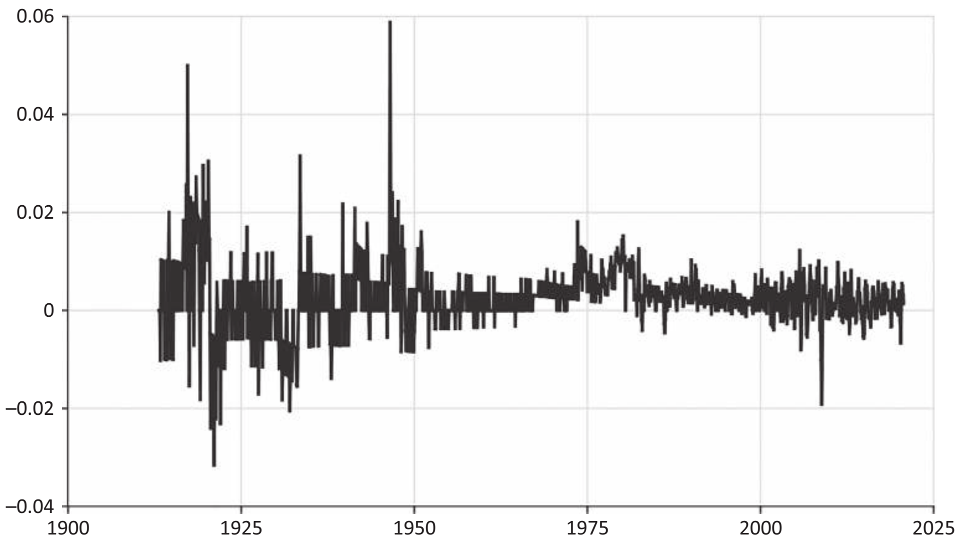

Figure 1.7 US monthly inflation rate.

Figure 1.6 shows the US monthly industrial production index, the series `INDPRO`. This series also shows the recent COVID spike, along with other boom and bust periods. The series has a pronounced upward trend due to economic growth and also seasonality similar to the unemployment series.

We next show the US monthly CPI inflation rate, calculated from the series `CPIAUCNS` (Figure 1.7). Clearly, before 1970, the series was essentially annual (apart from a few big spikes) and interpolated, in a not particularly clever way, to give a "monthly" series. The Phillips curve predicts an inverse relationship between unemployment rate and inflation. The raw correlation between contemporaneous values of the two monthly series over the

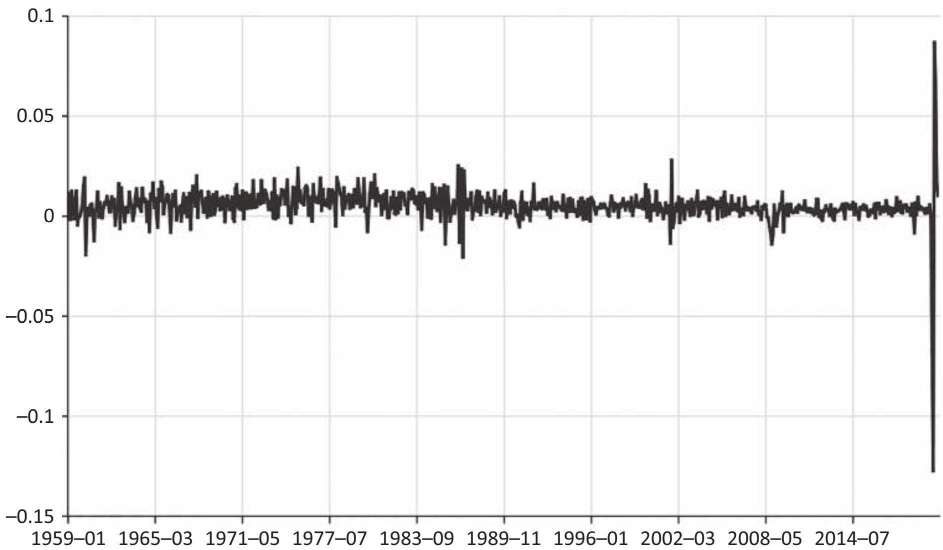

Figure 1.8 US consumption growth since 1959.

period 1948–2020 is about 0.04. One question is whether this can be improved by time series methods that adjust for seasonality and that bring in dynamics.

Figure 1.8 shows the monthly, seasonally adjusted, growth in personal consumption expenditure in the USA, the series `PCE`. The series is dominated by the most recent COVID-19 event. Hall (1978) argued that consumption should be a martingale, that is, the growth rate of consumption should be unpredictable.

We next show the monthly average daily maximum temperature at the Oxford weather station since 1850, from the dataset `OXMT` (Figure 1.9). This data shows an even more pronounced seasonality, with higher temperatures in the summer and lower temperatures in the winter. Of interest here is determining whether there is an upward trend in these temperatures due to climate change, but this is very difficult to see from the current plot, because it is dominated by the seasonal effect.

We next show, in Figure 1.10, the monthly average daily maximum temperature series for Toronto, Canada, since 1840, dataset `Toronto`. In this case we show the time series separately by month (January, February, etc.), the so-called seasonal subseries plot. The key thing about the graph is that every month is shown on the same vertical scale of temperature from -40 to $+40$ so that it is hard to perceive any change in the level of each series.

Figure 1.11 shows a higher-frequency temperature series, which is the temperature at the Cambridge University weather station recorded every 30 minutes since 1995, dataset `Cam30`. This data also has a pronounced seasonality, both within day (day is warmer than night) and within year (summer is warmer than winter). There does not appear to be a strong trend to this series over the time frame considered, although it is a little difficult to deal with the multiple seasonal patterns complicating things.

Next comes a weekly time series of Scottish mortality from the dataset `Deadscots` (Figure 1.12). This is the raw mortality unadjusted for population, which was around 5.25 million in 1974, and fell to 5.07 million in 2000 and thereafter rose to 5.46 million

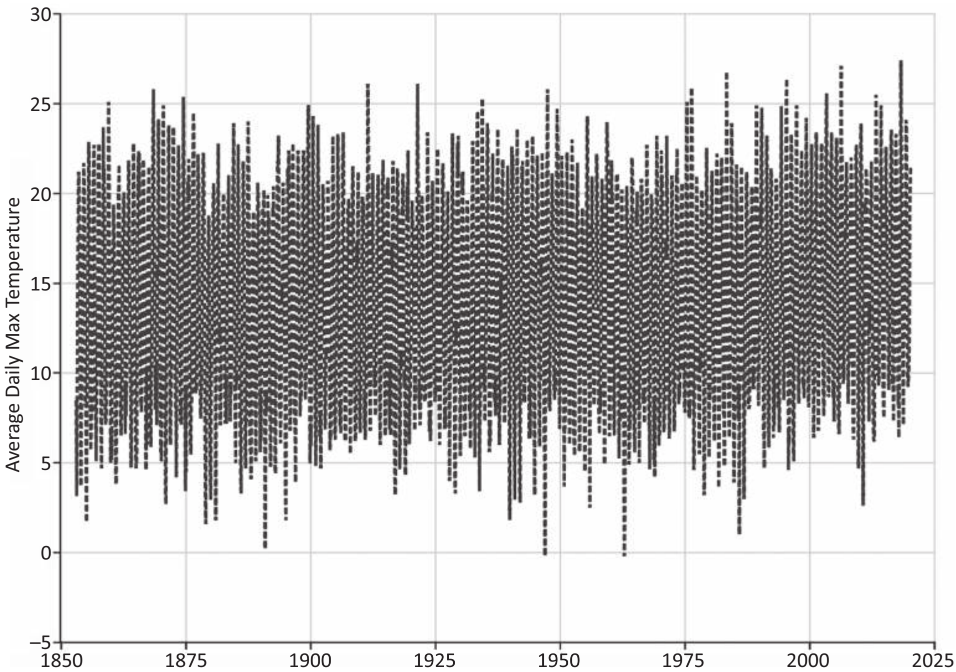

Figure 1.9 Oxford monthly average daily maximum temperature.

in 2019. This series was converted to a regular 52-week year by dividing the 53rd week between week 52 and week 1 of the following year. Modelling and forecasting mortality is important for insurance companies and public health bodies, not to mention funeral homes. There is a clear seasonality in this data since deaths are higher in the winter than in the summer. Demographers typically work with disaggregated (by age and sex) series and model these curves separately. For annual mortality, Denton *et al.* (2005) used an AR(2) process for the growth rate of Canada data from 1926 to 2000, see Chapter 3.

The next series, Figure 1.13. is of historical interest. It is data collected by John Arbuthnot on the annual number of boy and girl live births in London from 1629 to 1710, the dataset `Arbuthnot`. There appears to be first a downward trend between 1640 and 1660, which was the period of Oliver Cromwell and the Puritans, and a further, smaller, dip around 1665, caused by the Black Death, and then an upward trend in these raw numbers due to the expansion of the population of this city during the latter half of the century. The question Arbuthnot addressed was whether boys were more likely to be live-born than girls. He reported that for all 82 years there were more boys born than girls. We naturally think that the ratio should be 1, that is, boys and girls are equally likely, which we can think of as the null hypothesis. What is the probability that when you toss a coin you get heads 82 times in a row? This is $(1/2)^{82}$, which is a very small number (25 zeros). He concluded from this that boys are more likely to be born alive than girls. Can we say more? We plot in Figure 1.14 the ratio of boys to girls for each year. The ratio does not seem to have such a strong trend, but perhaps there are cycles in the ratio and short-term trends.

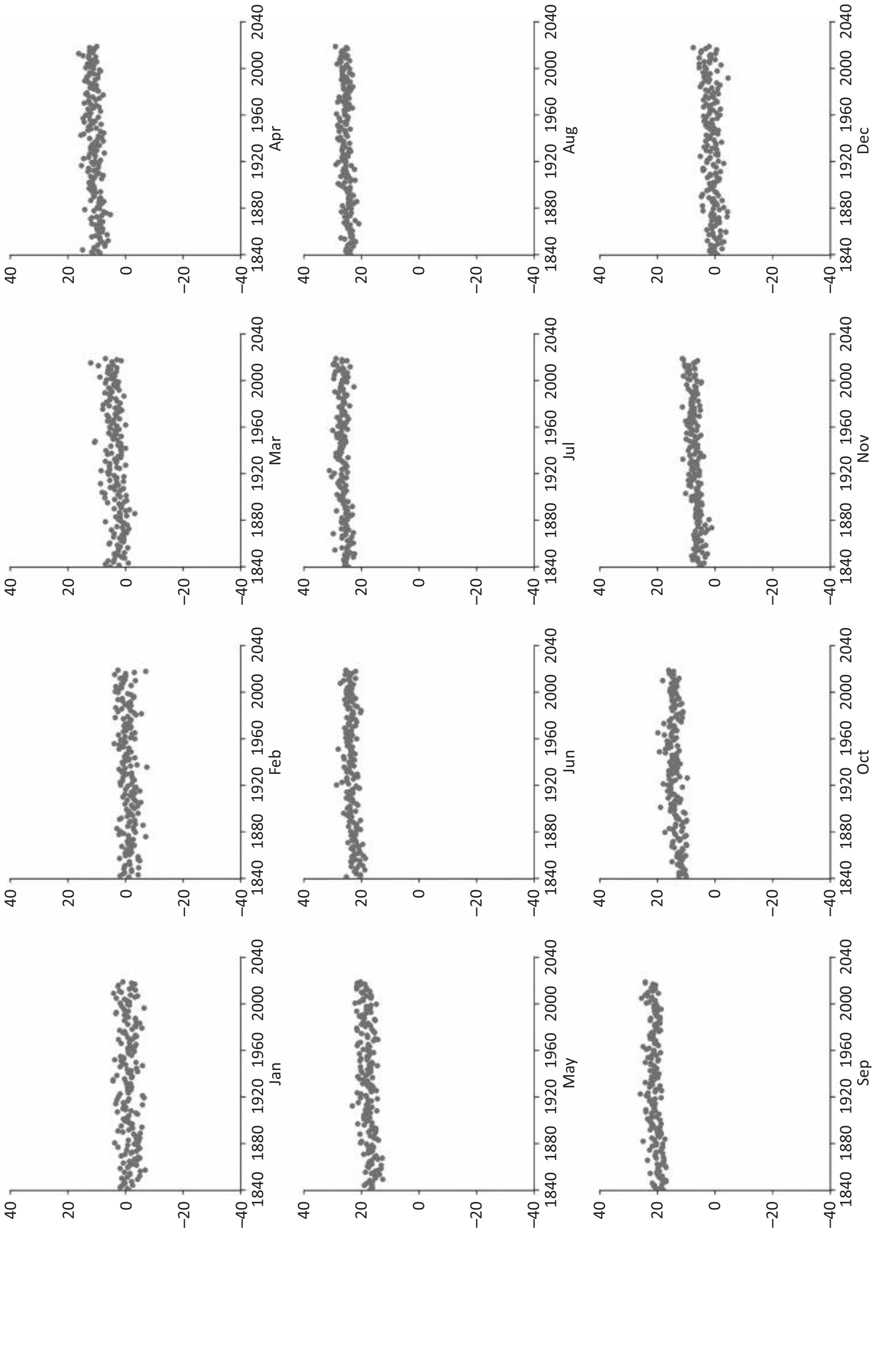

Figure 1.10 Toronto monthly average daily maximum time series by month.

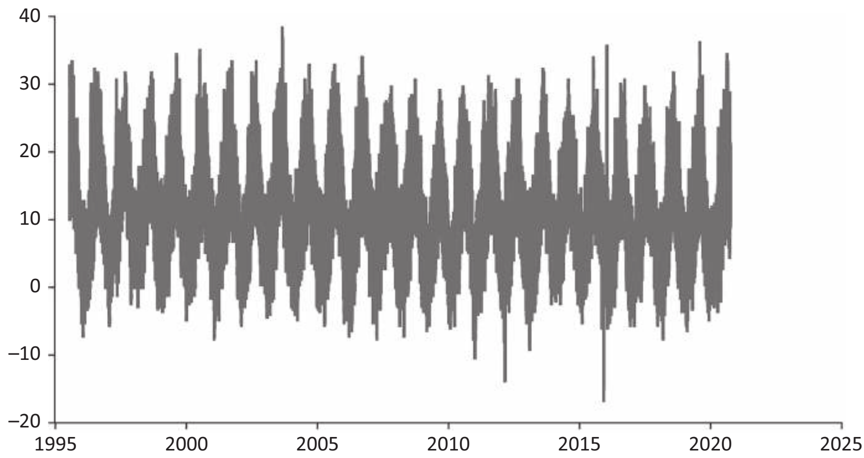

Figure 1.11 Cambridge half-hourly temperature.

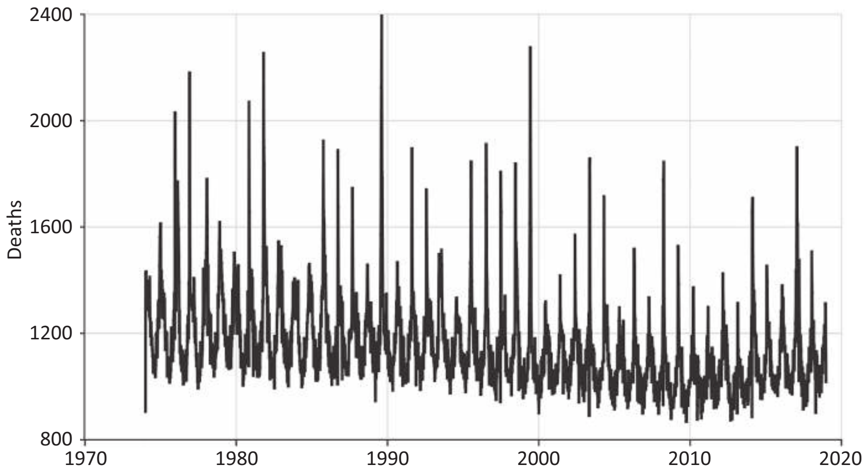

Figure 1.12 Weekly Scottish mortality, 1974–2019.

Finally, Figure 1.15 shows the daily number of new cases of COVID-19 reported in the UK during the first half of 2020, extracted from the dataset `Covid19`. This series has an unusual trend structure as it goes up, down, and up again; in both cases locally there is a strong trend. There is clearly a time series structure to this data as the number of new cases reported on a given day can be expected to depend on how many people were infected at the time, which itself depends on the recent numbers of new cases.

Where does economics come in? Economic theory typically involves solving some optimization problem defined by preferences, choices, information, and beliefs. This usually delivers some conditional moment restrictions that the data should satisfy. For example, the efficient markets theory says that stock returns should be unpredictable based on past information relative to a risk premium. This a fortiori suggests that stock returns should not have a seasonal component. But wait, if one is working with daily closing stock price data, then there are typically no transactions over the weekend,