Chapter 1

Overview

1.1 Introduction

There are many events that recur. These are evident either directly in time series or through their effects on economic outcomes. Examples would be business and financial cycles, crises, high and low levels of volatility and sentiment, seasonal patterns, floods and droughts, and high and low levels of temperature. These differ in many ways, but a principal one concerns their predictability based on a limited information set that includes only calendar time. Events such as seasonal patterns are highly predictable on the basis of calendar time and are therefore said to be *periodic* events. Of course, even if the event is periodic, its effects on economic outcomes may not be predictable, so we are ascribing predictability to the recurrent event itself. Other events are not strictly predictable but close to it. Thus, on a monthly or quarterly frequency, there are limits on the times that winter and summer can occur, and therefore these are mostly treated as being periodic. This leaves recurrent events such as the business cycle. These cannot be predicted with high probability conditional on just calendar time, leading them to be classified as *nonperiodic*. Such events are the subject of these lectures.

We might ask why there has always been interest in recurrent events such as the contraction and expansion *phases* of the business cycle. By focusing on the phases, one has moved from a comprehensive account of economic outcomes to a concise summary of them. In doing so, we might miss some feature that would be apparent from a detailed examination of the series. Of course this is no different to presenting the mean and variance of a series (and perhaps a measure of skewness) rather than either the series itself or its density function. Any cut of the data may be particularly interesting or informative to us and often becomes a key way of speaking about economic outcomes. In many ways this is true of the business cycle, where a great deal of attention is paid to the possibility and nature of a recession.

Often indicators of the events that are assembled concerning them represent a compression of information, and this can aid communication, just as citing a mean or a variance may suffice when talking about certain outcomes. Humans seem to be in favor of compressing information into manageable forms. It is also the case that in some instances the recurrent events we are looking at represent extreme outcomes, for example, crises, and a great deal of attention gets paid to such extremes owing to the potential for large losses when they occur.

There are three key issues we will need to deal with when discussing recurrent events. These are:

1. The description of the event via a set of statistics.
2. The uses that can be made of these statistics.
3. The possibility of predicting these events, in particular by using information sets that contain more information than just calendar time.

It pays to consider the basic issues involving recurrent events in the context of a simple example, and that is the modus operandi of this overview chapter. In later chapters modifications need to be made to derive more informative descriptions of events than are presented in this chapter, and these lead to additional complexities that need to be dealt with. But many of the fundamental issues are evident in the simple examples we work with here.

1.2 Describing the Events

To summarize the events we will need some *rules* that map the data we observe into a set of indicators that can then be used to construct statistics which describe the recurrent event in a succinct way. Within the literature there are two types of rules—those that are *prescribed* and those that are *model based*. We therefore consider each of these in turn, utilizing some simple examples.

1.2.1 Prescribed Rules

Consider first a *business cycle* in the level of economic activity. A business cycle involves periods of expansions and contractions in the *level* of economic activity.[1] If one viewed a graph of the level of economic activity one would see that a contraction begins when the activity reaches a *peak*, while an expansion begins with a *trough*. These *turning points* then provide a description of the business cycle. By their nature we are led to describe them by locating local maxima and minima in a series.

By far the simplest rule that would locate these features in a series y_t (the log of the level of economic activity Y_t) would be that a peak occurs at t if $y_{t-1} < y_t$ <u>and</u> $y_{t+1} < y_t$, while a trough is signaled if $y_{t-1} > y_t$ <u>and</u> $y_{t+1} > y_t$. Because log is a monotonic operation, the peaks and troughs in Y_t are the same as in y_t. Moreover, instead of using y_t to define a peak, we could use Δy_t, and say that a peak occurs at t if $\Delta y_t > 0$ <u>and</u> $\Delta y_{t+1} < 0$. In this special case an expansion happens in $t+1$ if $\Delta y_{t+1} > 0$, that is, there is positive growth in Y_t at $t+1$. In the same way a contraction at $t+1$ would involve negative growth ($\Delta y_{t+1} < 0$). With such rules there is a one-to-one relationship in the sample path between peaks and troughs and expansions and contractions. This is an example of what we refer to as a *prescribed rule*. It maps data on growth rates into cycle phases and points to the need to consider the nature of the data generating process (DGP) of Δy_t when assessing the likelihood of a peak or trough.

---

[1]That it is about the level of activity is clear from Zarnowitz and Ozyildirium (2006) who say "early studies which defined business cycles as sequences of expansions and contractions in a large array of series representing the levels of total output, employment."

Suppose therefore that Δy_t has the structure

$$\Delta y_t = \mu + \sigma \varepsilon_t, \tag{1.1}$$

where ε_t is normally and independently distributed with expected value of zero and variance of unity. Our abbreviation for this will be $n.i.d.(0, 1)$. Then we would have a recession in t if $\Delta y_t < 0$ and an expansion if $\Delta y_t > 0$. Hence, defining the binary variable S_t as being unity in expansions and zero in contractions, we would have $S_t = 1(\Delta y_t > 0)$, where $1(\cdot)$ has the value unity when $\Delta y_t > 0$, and zero when $\Delta y_t < 0$.[2] The binary variable S_t then summarizes when expansions and contractions occur, and we refer to it as the *cycle* in y_t. Given the model for Δy_t in (1.1) it is clear that

$$\begin{aligned} \Pr(S_t = 0) &= \Pr\left[\Delta y_t < 0\right] \\ &= \Pr\left[\varepsilon_t < -\frac{\mu}{\sigma}\right] \\ &= \Phi\left(-\frac{\mu}{\sigma}\right), \end{aligned}$$

where Φ is the c.d.f. of an $N(0, 1)$ random variable. Consequently, the probability of a recession will depend on the mean growth rate (μ) of Δy_t and its volatility σ.

In the foregoing, a cycle is a binary series which shows when contractions and expansions occur. Thus there is no continuous variable y'_t that is a "cycle," although we might try to construct such a series from y_t so that it had the same peaks and troughs as in y_t. In that instance y'_t would generally be called a *coincident indicator*. Of course it may not be possible to find a y'_t that has this property, and often coincident indicators are constructed so that their turning points are as close as possible to those in y_t, leading to the need to define some metric for measuring "closeness."

Occasionally, interest focuses not on (say) peaks and troughs in the level of the series but on whether variables such as confidence or volatility exceed a particular level. For

---

[2]As Δy_t is continuous in (1.1) we don't need to be concerned about how to deal with $\Delta y_t = 0$. If one is working with data where $\Pr(\Delta y_t = 0) > 0$ then we could define S_t as unity if $\Delta y_t \leq 0$.

example, Bloom (2009) and Caggiano et al. (2014) construct a dummy variable that takes the value 1 when a measure of financial market volatility (υ_t) exceeds a critical level. The specific rule employed by the latter is $\xi_t = 1\,(\upsilon_t > 1.65\sigma)$, where σ is the standard deviation of υ_t. It is clear that this also involves the construction of a binary random variable (ξ_t) based on an underlying variable υ_t through a rule, after which ξ_t can be analyzed or used. As stated, ξ_t would be an "exceedance" measure, and these are very popular in studies involving financial contagion.

In history there have been many ways to describe recurrent events in some succinct way. In Chapter 2 we look at three key ways of describing the "ups and downs" of an economy. These involve oscillations, fluctuations, and cycles. Each has a different way of describing the phenomenon being investigated and summarizing it. These vary depending on the frequency of the data one is working with. Moreover, often many series rather than a single one are used to describe the recurrent pattern. This raises some extra compression issues which will be examined in Chapter 3, leaving Chapter 2 to engage in a comparative study of the relationships between oscillations, fluctuations, and cycles.

1.2.2 Model-Based Rules

It is clear that the turning points in y_t found using any prescribed rule will depend on the nature of Δy_t. For this reason it is probably not surprising that a literature has evolved in which a model is assumed for Δy_t that depends in some way on *regimes*. When there are just two regimes, they can be represented by a binary variable ξ_t. To capture the flavor of these models, assume there that are just two regimes, with the model for Δy_t being a mixture of two normals

$$\Delta y_t = \xi_t N(\mu_1, \sigma^2) + (1 - \xi_t) N(\mu_0, \sigma^2). \tag{1.2}$$

Thus the first regime has a growth rate of μ_0 while the second is μ_1. Then when $\xi_t = 1$ is realized Δy_t would be drawn from an $N(\mu_1, \sigma^2)$ density, whereas if $\xi_t = 0$ is realized, Δy_t would be drawn from an $N(\mu_0, \sigma^2)$ density. In comparison with the model (1.1) where Δy_t is $N(\mu, \sigma^2)$, the *switching model* (1.2) is non-normal with fatter tails.

We might then ask what the probability of observing $\xi_t = 0$ would be given a *realization* of y_t equal to y_t^*? As Hamilton (2011) notes we would have the following probabilities for this model.

$$\Pr(\xi_t = 0, \Delta y_t) = N(\mu_0, \sigma^2) \Pr(\xi_t = 0) \tag{1.3}$$

$$\Pr(\xi_t = 0|\Delta y_t) = \frac{\Pr(\xi_t = 0, \Delta y_t)}{\Pr(\xi_t = 0, \Delta y_t) + \Pr(\xi_t = 1, \Delta y_t)} \tag{1.4}$$

Now, given an unconditional probability of π_0 for the event $\xi_t = 0$, the numerator of (1.4) would be $N(\mu_0, \sigma^2)\pi_0$, while the denominator will be the density for Δy_t. The latter is the following combination of normals:

$$N(\mu_0, \sigma^2)\pi_0 + N(\mu_1, \sigma^2)(1 - \pi_0). \tag{1.5}$$

Once the values for the parameters in each regime are specified, as well as the unconditional probability for $\xi_t = 0$, we can describe what the densities would be at a value for Δy_t of Δy_t^*, and so can compute a value for $\Pr(\xi_t = 0|\Delta y_t)$ from (1.4). In the event that $\mu_0 = \mu_1$, $\Pr(\xi_t = 0|\Delta y_t = \Delta y_t^*)$ would be .5.

Adding on a restriction that $\mu_1 > \mu_0$ would mean that the regimes could be characterized as involving low and high growth rates. Moreover, if one had observed a large positive value for Δy_t^*, it would most likely indicate that $\xi_t = 1$ had been realized at t, whereas negative values would imply that $\xi_t = 0$. Now while this indicates which growth regime might hold, it doesn't describe whether one is an expansion or a recession at time t. To produce the requisite mapping between regimes and business cycle phases researchers need to *define a new binary variable* ζ_t taking the value unity in expansions and zero in contractions. A rule that does this is to set $\zeta_t = 1$ if $\Pr(\xi_t = 1|\Delta y_t^*)$ exceeds some prescribed value c—say $c = .5$. Thus the indicator of expansions and contractions would be $\zeta_t = 1[\Pr(\xi_t = 1|\Delta y_t^*) - c]$ making it clear that although it is tempting to think of the regime variables ξ_t as expansions and contractions (and often this is the loose terminology that is adopted), it is ζ_t that capture those phases and not ξ_t. As we will see in a number of chapters, the realizations of ξ_t and ζ_t can be very different, so it is easy to make mistakes by conflating them.

This regime-switching model is interesting. The parameters to be estimated are μ_j, σ^2, and π_0. This means four parameters, so we need to use four moments of Δy_t to estimate them, that is, more than just the mean and variance are required. One always needs to ask about how many moments are needed for estimation of the parameters of any regime-switching model and what would they be? It is also clear that this model produces a *rule* for determining what ζ_t is. It takes data on Δy_t (Δy_t^*) and elicits a decision about whether ζ_t is unity or zero by computing $\Pr(\xi_t = 1|\Delta y_t = \Delta y_t^*)$ and then comparing it to the value of c. Because the probability, and hence the rule, comes from a model of Δy_t, we refer to it as a *model-based rule*. It is clear from the simple example that the rule is a nonlinear function of the data and depends on more than the sign of Δy_t^*, all that was used by the prescribed rule when defining S_t. Thus there may be differences between the S_t coming from prescribed rules and the ζ_t coming from model-based rules, and a section of Chapter 4 looks at this.

An estimation problem occurs with this model. From (1.5) the log likelihood for realizations $\{\Delta y_t^*\}_{t=1}^T$ will be

$$L = \sum_{t=1}^{T} \log[\phi_{0t}\pi_0 + \phi_{1t}(1 - \pi_0)],$$

where $\phi_{jt} = \frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{1}{2\sigma^2}(\Delta y_t^* - \mu_j)^2)$. First, the likelihood is unbounded. As Kiefer (1978) notes this can be shown by setting $\mu_1 = \Delta y_1^*$ (say) and then letting $\sigma_1^2 \to 0$. This means one has to look for interior values of the parameters that set the first-order conditions to zero, that is, to find a local rather than global maxima for L. Second, consider setting $\mu_0 = 1, \mu_1 = 2, \pi_0 = .4, \pi_1 = 1 - \pi_0 = .6, \sigma = 1$, and using these to evaluate the likelihood. By inspection of the last, switching the parameter values for the regimes to $\mu_1 = 1, \mu_0 = 2$, $\pi_0 = .6, \pi_1 = .4, \sigma = 1$ will produce exactly the same value of the likelihood, that is, there are multiple maxima in L. This is what is referred to as the "labeling problem," since the regimes can be interchanged without changing the likelihood, leading to multiple maxima in L. Because such a phenomenon can create numerical problems, some way of avoiding it is desirable. One way is to to impose a constraint such as $\mu_1 > \mu_0$, since we can then rule out the interchange of parameters between regimes. Doing so, regime 1 is always the high-growth regime and regime 0 is the low-growth one. In practice one often sees researchers who use model-based rules deciding which is the high-growth and which is the low-growth regime *after* estimation is performed, that is, if $\hat{\mu}_1 > \hat{\mu}_0$ it is assumed that the high-growth regime is characterized by a mean $\hat{\mu}_1$. Of course this does not solve the multiple maxima problem. Moreover, in the event that there are more parameters in the model which are regime dependent, for example, suppose that volatility takes different values of σ_0^2 and σ_1^2 in the regimes, ex post classification must be problematic, unless one knows what the relative volatilities are in each regime. This is where it becomes very important to state prior information, and it points to a Bayesian approach when this information is good.

We return to the theme above in Chapter 4. Chapter 4 will also look at many extensions of the switching regression model, mostly under the soubriquet of Markov switching. One question that will repeatedly occur in the extensions is whether the corresponding models have too many parameters to estimate. In the foregoing example there are four parameters, and so four moments of Δy_t are needed for estimation. But in some extensions there are huge number of parameters and, with just a single series Δy_t, it is hard to believe that one can estimate them precisely. Chapters 4 and 6 will use some examples from the literature to show that this is an issue, at least when y_t is a scalar. It will emerge that all the issues with the switching regression model in its simplest context will recur. The danger is that with the greater complexity of the models being formulated and estimated, these issues sometimes get lost, so it is worth seeing them in the simplest environment.

1.2.3 *Differences between Prescribed and Model-Based Rules*

There are two broad ways of describing recurrent events. Prescribed rules work directly with the data. The only decision that had to be made with the version above was how to define a turning point in y_t. In contrast, model-based rules require a model for Δy_t, as this is used to construct the rules. Both work with Δy_t, even though it is the behavior of y_t that is under investigation. Although a turning point is a relatively simple concept to define, one might ask whether the model adopted for Δy_t when formulating model-based rules is a correct description of Δy_t. Suppose it failed to fit the data using standard diagnostic tests? Would we use it then? There seems no answer to this query, since ideally we want to judge the model by its ability to replicate the recurrent patterns. But because we don't directly observe these, they need to be measured in some way. If there was a set of measures that were widely agreed on, and for U.S. business cycles these are the National Bureau of Economic Research (NBER) dates, it seems to make sense to compare the latter with ζ_t when judging the adequacy of the model. In countries where such external measures are not available, it often happens that a comparison is made between the S_t constructed from a prescribed rule and the ζ_t from a model. In that instance one wonders why one fitted a model. The only arguments advanced seem to be that there is an advantage to having an equation for generating the ζ_t, and it is implied that this is not available when the S_t from prescribed rules are used. We argue in this book that there is no such advantage. It is generally possible to work out an approximation that captures some of the main features of the DGP of S_t associated with recurrent events, at least when discussing items like cycles. When one turns to recurrent events involving high and low values of some variable, it is often hard to find a satisfactory prescribed rule that would produce a useful classification. In those instances regime-based models might be useful for giving one adequate rules for defining "high" and "low." Whether they do so is an empirical question.

1.3 Using the Event Indicators ("States")

Once a description is available of when the recurrent events occur, it is natural to inquire into questions concerning how long they persist, the size of their effects, and whether the probability of occurrence depends on some observable variables. These questions are augmented when there are recurrent events in multiple series, for example, one often wants to ask whether the events are synchronized. Chapter 5 makes a start on the agenda of measuring recurrent event features through concepts such as duration, amplitude, and variability of the events. As well as defining and measuring these with some statistics, we need to pay attention to the distributional properties of the latter. We examine the work that has been done on this and suggest some extensions. Chapter 7 uses the measures of Chapter 5 to look at cycles in a range of series and answer some questions that have often driven the literature, for example, on the asymmetric nature of the business cycle.

Often questions are asked either about what would explain the recurrent events or their influence on economic outcomes. An example of the former would be whether the probability of a recession has changed over time or whether it depends on economic and political institutions. To answer the latter, interest might be in whether recessions lead to higher or lower volatility in stock prices. Once either S_t (or ζ_t) are available, it is possible to set about examining these questions. To look at the first set we could "regress" S_t or ζ_t against some variables x_t, while the latter would require a regression of x_t against S_t (or ζ_t).

To answer whether such regressions make sense, it is necessary to know what the statistical properties of S_t and ζ_t are, that is, the main features of their DGPs. Some of their properties clearly originate from the nature of Δy_t. In the models of (1.1) and (1.2), Δy_t is an identically and independently distributed (i.i.d.) process, and therefore both S_t and ζ_t are independently distributed (i.d.). To investigate the impact of some x_t on S_t and ζ_t we would therefore need to allow Δy_t to depend on x_t in some way. Starting with the model

(1.1) used when discussing the prescribed rule $S_t = 1(\Delta y_t > 0)$, assume that $\Delta y_t = \mu + x_t'\beta + \sigma\varepsilon_t$, where ε_t is $n.i.d.(0, 1)$. Then

$$\begin{aligned}\Pr(S_t = 0|x_t) &= \Pr(\Delta y_t < 0|x_t)\\ &= \Pr(\mu + x_t'\beta + \sigma\varepsilon_t < 0|x_t)\\ &= \Phi\left(-\frac{\mu + x_t'\beta}{\sigma}\right).\end{aligned}$$

This can be recognized as the probit functional form. Moreover, because S_t is independently distributed, $\frac{\mu}{\sigma}$ and $\frac{\beta}{\sigma}$ can be estimated using probit model software. Just as for the probit model only the ratio of parameters can be identified. In Chapter 8 we examine whether this equivalence holds for other types of prescribed rules. In general the answer will be in the negative, since Chapter 2 shows that S_t will not be independently distributed, and so the implications of that fact need to be canvassed. In the case where S_t is i.d., using these indicators as either a regressor or a regressand is straightforward. But when S_t has some serial correlation, it is necessary to make adjustments to deal with that fact, and these are described in various points in Chapters 5–8.

The situation is less clear for model-based rules yielding ζ_t. Because the ζ_t are constructed using a model, the simple augmentation of (1.2) to allow for an influence of the x_t on ζ_t would be

$$\Delta y_t = \xi_t N(\mu_1, \sigma^2) + (1 - \xi_t)N(\mu_0, \sigma^2) + x_t'\beta. \qquad (1.6)$$

This would mean that a new set of event indicators ζ_t, ζ_t', would be implied, and the mapping between ζ_t' and x_t would be unlikely to be normal. So the two-stage approach used with S_t—first determine S_t, and then investigate the relationship between S_t and x_t—doesn't really apply when model-based rules are used. One needs to describe how x_t affects Δy_t, and after this model is fitted, one can determine the relationship between ζ_t' and x_t. There are a number of ways x_t could affect Δy_t when regime-switching models are involved. One is just through the simple augmentation used in (1.6) where the influence of x_t on Δy_t does not depend on which regime holds at a point in time, but it would also be possible to allow either the regime means (μ_j) or the probability function for ξ_t to depend on x_t. All of these alternatives have been used and will be mentioned in Chapter 4.

1.4 Prediction of Recurrent Events

In the analysis of the preceding subsection we looked at items like $\Pr(S_t = 0|x_t)$. This shows how the probability of *being in a recession at t* varies with x_t. That event needs to be distinguished from the probability of *going into a recession at t*, that is, of encountering a peak in y_t. Often the literature has not made this distinction clear. Below we will point out why it is important, and Chapter 9 will examine it in more detail.

The comment above draws attention to two issues. One relates to the information available when making a prediction. We might be interested in predicting what cycle phase the economy is at $t+1$ using information Ω_t available at time t. Such information could either be a set of observed variables x_t or even S_t and its history, including the elapsed time since the event last occurred (duration). The distinction between calendar time and duration of the event is of some importance. This book is focused on events that are unpredictable based only on calendar time, so that conditioning on a particular month, quarter, or year yields no information about the event. Nevertheless, there is the possibility that the events could be predictable based on other quantities, such as the time spent in a phase. The popular phrase "at this stage of the business cycle" expresses the idea that the beginning and end of events such as recessions could be predictable based on elapsed duration. But other variables have often been proposed as ways of predicting events such as recessions and turning points.

As seen in Chapter 2 whether S_t is known, and also what part of its history is known, will depend on the rules being used. In the simple prescribed rule dealt with in Section 1.2, once Δy_t is known so is S_t. Hence, in that context it would make sense to assume that the information available was x_t and S_t. Of course S_{t+j} will depend on x_{t+j} so some assumption will also be needed about the nature of x_t. In our examples above, Δy_t was *i.i.d.* so we would need to also have x_t being *i.i.d.* Then, because in that case S_t is independent of $S_{t-k}(k > 0)$, we must have $E(S_{t+j}|\Omega_t) = \mu_S = E(S_{t+j})$, and so one would use the unconditional mean as the best predictor.

As mentioned earlier an alternative item to forecast which has a long history in macroeconomics is whether there is a *turning point* at t. To examine this it is necessary to describe a turning point. We therefore define two binary variables $\wedge_t$ and $\vee_t$, where $\wedge_t$ takes the value unity if a peak occurs at t, and zero otherwise, while $\vee_t$ indicates a trough. Then $\wedge_t = S_t(1 - S_{t+1})$ and $\vee_t = S_{t+1}(1 - S_t)$. Thus

$$\begin{aligned}\Pr(\wedge_t = 1|\Omega_t) &= E(\wedge_t|\Omega_t)\\ &= E(S_t(1 - S_{t+1})|\Omega_t).\end{aligned}$$

When the S_t are independently distributed $E(S_t(1 - S_{t+1})|\Omega_t) = \mu_S(1 - \mu_S)$. Hence $\Pr(\wedge_t = 1|\Omega_t) \leq (1 - \mu_S) = \Pr(S_{t+1} = 0|\Omega_t)$, so that there may be a high probability of being in a recession at $t + 1$ (given x_t), but a low probability of predicting that one will move into a recession (encounter a peak) at time t. If a different set of rules is used and/or Δy_t is not independently distributed, the prediction problem is much more complex and is analyzed in Chapter 9.

1.5 Conclusion

The chapter has introduced many of the concepts and methods that will occupy us in the remainder of the book. A key element is that turning points in a series y_t are defined by a set of rules. Sometimes these rules are prescribed and sometimes they are based on a model for Δy_t. It is not true to say, as Diebold and Rudebusch (1996, 69) do, that "Yet it is only within a regime-switching framework that the concept of a turning point has intrinsic meaning." A turning point gets its meaning from the rules that are applied to locate it. A second item of concern that was bought up in this opening chapter was connected to this, namely, that there has been a confusion between the regimes present in many nonlinear models and the phases of the recurrent events that are isolated with a set of rules. Mixing the two different ideas is something that will lead to many difficulties in later chapters.