Chapter 1

Ankit Gupta

26/11/2023

All R examples in the book assume you have loaded the *fpp3* package first.

```
library(fpp3)
```

```
## -- Attaching packages -------------------------------------------- fpp3 0.5 --

## v tibble      3.2.1     v tsibble     1.1.3
## v dplyr       1.1.3     v tsibbledata 0.4.1
## v tidyr       1.3.0     v feasts      0.3.1
## v lubridate   1.9.3     v fable       0.3.3
## v ggplot2     3.4.4     v fabletools  0.3.4

## -- Conflicts ------------------------------------------------- fpp3_conflicts --
## x lubridate::date()    masks base::date()
## x dplyr::filter()      masks stats::filter()
## x tsibble::intersect() masks base::intersect()
## x tsibble::interval()  masks lubridate::interval()
## x dplyr::lag()         masks stats::lag()
## x tsibble::setdiff()   masks base::setdiff()
## x tsibble::union()     masks base::union()
```

These include several *tidyverse* packages, and packages to handle time series and forecasting in a "tidy" framework.

We use the tsibble and fable packages rather than the forecast package. This allows us to integrate closely with the tidyverse collection of packages.

Forecasting has fascinated people for thousands of years, sometimes being considered a sign of divine inspiration, and sometimes being seen as a criminal activity.

(i) In ancient Babylon, forecasters would foretell the future based on the appearance of a sheep's liver.

(ii) A ban on forecasting occurred in England in 1736 when it became an offence to defraud by charging money for predictions.

(iii) The punishment was three months' imprisonment with hard labour.

(iv) The varying fortunes of forecasters arise because good forecasts can seem almost magical, while bad forecasts may be dangerous. Consider the following famous predictions about computing.

(A) I think there is a world market for maybe five computers. (Chairman of IBM, 1943)

(B) Computers in the future may weigh no more than 1.5 tons. (Popular Mechanics, 1949)

(C) There is no reason anyone would want a computer in their home. (President, DEC, 1977)

The last of these was made only three years before IBM produced the first personal computer. Not surprisingly, you can no longer buy a DEC computer. Forecasting is obviously a difficult activity, and businesses that do it well have a big advantage over those whose forecasts fail.

(1.1) What can be forecast?

Forecasting is required in many situations: deciding whether to build another power generation plant in the next five years requires forecasts of future demand; scheduling staff in a call centre next week requires forecasts of call volumes; stocking an inventory requires forecasts of stock requirements. Forecasts can be required several years in advance (for the case of capital investments), or only a few minutes beforehand (for telecommunication routing). Whatever the circumstances or time horizons involved, forecasting is an important aid to effective and efficient planning.

Some things are easier to forecast than others. The time of the sunrise tomorrow morning can be forecast precisely. On the other hand, tomorrow's lotto numbers cannot be forecast with any accuracy. The predictability of an event or a quantity depends on several factors including:

(1) how well we understand the factors that contribute to it;
(2) how much data is available;
(3) how similar the future is to the past;
(4) whether the forecasts can affect the thing we are trying to forecast.

For example, short-term forecasts of residential electricity demand can be highly accurate because all four conditions are usually satisfied.

(1) We have a good idea of the contributing factors: electricity demand is driven largely by temperatures, with smaller effects for calendar variation such as holidays, and economic conditions.
(2) There is usually several years of data on electricity demand available, and many decades of data on weather conditions.
(3) For short-term forecasting (up to a few weeks), it is safe to assume that demand behaviour will be similar to what has been seen in the past.
(4) For most residential users, the price of electricity is not dependent on demand, and so the demand forecasts have little or no effect on consumer behaviour.

On the other hand, when forecasting currency exchange rates, only one of the conditions is satisfied: there is plenty of available data. However, we have a limited understanding of the factors that affect exchange rates, the future may well be different to the past if there is a financial or political crisis in one of the countries, and forecasts of the exchange rate have a direct effect on the rates themselves.Forecasting whether the exchange rate will rise or fall tomorrow is about as predictable as forecasting whether a tossed coin will come down as a head or a tail. In both situations, you will be correct about 50% of the time, whatever you forecast. In situations like this, forecasters need to be aware of their own limitations, and not claim more than is possible.

Often in forecasting, a key step is knowing when something can be forecast accurately, and when forecasts will be no better than tossing a coin. Good forecasts capture the genuine patterns and relationships which exist in the historical data, but do not replicate past events that will not occur again. In this book, we will learn how to tell the difference between a random fluctuation in the past data that should be ignored, and a genuine pattern that should be modelled and extrapolated.

Many people wrongly assume that forecasts are not possible in a changing environment. Every environment is changing, and a good forecasting model captures the way in which things are changing. Forecasts rarely assume that the environment is unchanging. What is normally assumed is that the way in which the environment is changing will continue into the future. That is, a highly volatile environment will continue to be highly volatile; a business with fluctuating sales will continue to have fluctuating sales; and an economy that has gone through booms and busts will continue to go through booms and busts. A forecasting model is intended to capture the way things move, not just where things are. As Abraham Lincoln said, "If we could first know where we are and whither we are tending, we could better judge what to do and how to do it".

Forecasting situations vary widely in their time horizons, factors determining actual outcomes, types of data patterns, and many other aspects. Forecasting methods can be simple, such as using the most recent observation as a forecast (which is called the naïve method), or highly complex, such as neural nets and econometric systems of simultaneous equations. Sometimes, there will be no data available at all. For example, we may wish to forecast the sales of a new product in its first year, but there are obviously no data to work with. In situations like this, we use judgmental forecasting, discussed in Chapter 6. The choice of method depends on what data are available and the predictability of the quantity to be forecast.

(1.2) Forecasting, goals and planning

Forecasting: Forecasting is about predicting the future as accurately as possible, given all of the information available, including historical data and knowledge of any future events that might impact the forecasts.

An organisation needs to develop a forecasting system that involves several approaches to predicting uncertain events. Such forecasting systems require the development of expertise in identifying forecasting problems, applying a range of forecasting methods, selecting appropriate methods for each problem, and evaluating and refining forecasting methods over time.

(1.3) Determining what to forecast

In the early stages of a forecasting project, decisions need to be made about what should be forecast. For example, if forecasts are required for items in a manufacturing environment, it is necessary to ask whether forecasts are needed for:

(1) every product line, or for groups of products?

(2) every sales outlet, or for outlets grouped by region, or only for total sales?

(3) weekly data, monthly data or annual data?

It is also necessary to consider the forecasting horizon. Will forecasts be required for one month in advance, for 6 months, or for ten years? Different types of models will be necessary, depending on what forecast horizon is most important.

(1.4) Forecasting data and methods

The appropriate forecasting methods depend largely on what data are available.

If there are no data available, or if the data available are not relevant to the forecasts, then **qualitative forecasting** methods must be used. These methods are not purely guesswork—there are well-developed structured approaches to obtaining good forecasts without using historical data. These methods are discussed in Chapter 6.

Quantitative forecasting can be applied when two conditions are satisfied:

numerical information about the past is available; it is reasonable to assume that some aspects of the past patterns will continue into the future.

There is a wide range of quantitative forecasting methods, often developed within specific disciplines for specific purposes. Each method has its own properties, accuracies, and costs that must be considered when choosing a specific method.

Most quantitative prediction problems use either time series data (collected at regular intervals over time) or cross-sectional data (collected at a single point in time). In this book we are concerned with forecasting future data, and we concentrate on the time series domain.

Time series forecasting Examples of time series data include:

(1) Annual Google profits
(2) Quarterly sales results for Amazon

(3) Monthly rainfall
(4) Weekly retail sales
(5) Daily IBM stock prices
(6) Hourly electricity demand
(7) 5-minute freeway traffic counts
(8) Time-stamped stock transaction data

Anything that is observed sequentially over time is a time series. In this book, we will only consider time series that are observed at regular intervals of time (e.g., hourly, daily, weekly, monthly, quarterly, annually). Irregularly spaced time series can also occur, but are beyond the scope of this book.

The simplest time series forecasting methods use only information on the variable to be forecast, and make no attempt to discover the factors that affect its behaviour. Therefore they will extrapolate trend and seasonal patterns, but they ignore all other information such as marketing initiatives, competitor activity, changes in economic conditions, and so on.

Predictor variables and time series forecasting Predictor variables are often useful in time series forecasting. For example, suppose we wish to forecast the hourly electricity demand (ED) of a hot region during the summer period. A model with predictor variables might be of the form

$$ED = f(\text{current temperature, strength of economy, population,time of day, day of week, error})$$

The relationship is not exact — there will always be changes in electricity demand that cannot be accounted for by the predictor variables. The "error" term on the right allows for random variation and the effects of relevant variables that are not included in the model. We call this an explanatory model because it helps explain what causes the variation in electricity demand.

Because the electricity demand data form a time series, we could also use a time series model for forecasting. In this case, a suitable time series forecasting equation is of the form

$$ED_{t+1} = f(ED_t, ED_{t-1}, ED_{t-2}, ..., error)$$

where t is the present hour, t+1 is the next hour, t−1 is the previous hour,t−2 is two hours ago, and so on. Here, prediction of the future is based on past values of a variable, but not on external variables that may affect the system. Again, the "error" term on the right allows for random variation and the effects of relevant variables that are not included in the model.

There is also a third type of model which combines the features of the above two models. For example, it might be given by

$$ED_{t+1} = f(ED_t, \text{current temperature,time of day, day of week, error})$$

These types of "mixed models" have been given various names in different disciplines. They are known as dynamic regression models, panel data models, longitudinal models, transfer function models, and linear system models (assuming that f is linear).

An explanatory model is useful because it incorporates information about other variables, rather than only historical values of the variable to be forecast. However, there are several reasons a forecaster might select a time series model rather than an explanatory or mixed model. First, the system may not be understood, and even if it was understood it may be extremely difficult to measure the relationships that are assumed to govern its behaviour. Second, it is necessary to know or forecast the future values of the various predictors in order to be able to forecast the variable of interest, and this may be too difficult. Third, the main concern may be only to predict what will happen, not to know why it happens. Finally, the time series model may give more accurate forecasts than an explanatory or mixed model.

The model to be used in forecasting depends on the resources and data available, the accuracy of the competing models, and the way in which the forecasting model is to be used.

(1.5) Some case studies

Case 1

The client was a large company manufacturing disposable tableware such as napkins and paper plates. They needed forecasts of each of hundreds of items every month. The time series data showed a range of patterns, some with trends, some seasonal, and some with neither. At the time, they were using their own software, written in-house, but it often produced forecasts that did not seem sensible. The methods that were being used were the following:

(1) average of the last 12 months data;

(2) average of the last 6 months data;

(3) prediction from a straight line regression over the last 12 months;

(4) prediction from a straight line regression over the last 6 months;
(5) prediction obtained by a straight line through the last observation with slope equal to the average slope of the lines connecting last year's and this year's values;
(6)prediction obtained by a straight line through the last observation with slope equal to the average slope of the lines connecting last year's and this year's values, where the average is taken only over the last 6 months.

They required us to tell them what was going wrong and to modify the software to provide more accurate forecasts. The software was written in COBOL, making it difficult to do any sophisticated numerical computation.

(1.6) The basic steps in a forecasting task

A forecasting task usually involves five basic steps.

Step 1: Problem definition.

Often this is the most difficult part of forecasting. Defining the problem carefully requires an understanding of the way the forecasts will be used, who requires the forecasts, and how the forecasting function fits within the organisation requiring the forecasts. A forecaster needs to spend time talking to everyone who will be involved in collecting data, maintaining databases, and using the forecasts for future planning.

Step 2: Gathering information.***

There are always at least two kinds of information required: (a) statistical data, and (b) the accumulated expertise of the people who collect the data and use the forecasts. Often, it will be difficult to obtain enough historical data to be able to fit a good statistical model. In that case, the judgmental forecasting methods of Chapter 6 can be used. Occasionally, old data will be less useful due to structural changes in the system being forecast; then we may choose to use only the most recent data. However, remember that good statistical models will handle evolutionary changes in the system; don't throw away good data unnecessarily.

Step 3: Preliminary (exploratory) analysis.

Always start by graphing the data. Are there consistent patterns? Is there a significant trend? Is seasonality important? Is there evidence of the presence of business cycles? Are there any outliers in the data that need to be explained by those with expert knowledge? How strong are the relationships among the variables available for analysis? Various tools have been developed to help with this analysis. These are discussed in Chapters 2 and 3.

Step 4: Choosing and fitting models.

The best model to use depends on the availability of historical data, the strength of relationships between the forecast variable and any explanatory variables, and the way in which the forecasts are to be used. It is common to compare two or three potential models. Each model is itself an artificial construct that is based on a set of assumptions (explicit and implicit) and usually involves one or more parameters which must be estimated using the known historical data. We will discuss regression models (Chapter 7), exponential smoothing methods (Chapter 8), Box-Jenkins ARIMA models (Chapter 9), Dynamic regression models (Chapter 10), Hierarchical forecasting (Chapter 11), and several advanced methods including neural networks and vector autoregression (Chapter 12).

Step 5: Using and evaluating a forecasting model

Once a model has been selected and its parameters estimated, the model is used to make forecasts. The performance of the model can only be properly evaluated after the data for the forecast period have become available. A number of methods have been developed to help in assessing the accuracy of forecasts. There are also organisational issues in using and acting on the forecasts. A brief discussion of some of these issues is given in Chapter 5. When using a forecasting model in practice, numerous practical issues arise such as how to handle missing values and outliers, or how to deal with short time series. These are discussed in Chapter 13.

(1.7) The statistical forecasting perspective

The thing we are trying to forecast is unknown (or we would not be forecasting it), and so we can think of it as a random variable. For example, the total sales for next month could take a range of possible values, and until we add up the actual sales at the end of the month, we don't know what the value will be. So until we know the sales for next month, it is a random quantity.

Because next month is relatively close, we usually have a good idea what the likely sales values could be. On the other hand, if we are forecasting the sales for the same month next year, the possible values it could take are much more variable. In most forecasting situations, the variation associated with the thing we are forecasting will shrink as the event approaches. In other words, the further ahead we forecast, the more uncertain we are.

When we obtain a forecast, we are estimating the middle of the range of possible values the random variable could take. Often, a forecast is accompanied by a prediction interval giving a range of values the random variable could take with relatively high probability. For example, a 95% prediction interval contains a range of values which should include the actual future value with probability 95%.

Rather than plotting individual possible futures, we usually show these prediction intervals instead. Figure 1.3 shows 80% and 95% intervals for the future Australian international visitors. The blue line is the average of the possible future values, which we call the point forecasts. (refer fig in book)

We will use the subscript "t" for time. For example, y_t will denote the observation at time "t". Suppose we denote all the information we have observed as "I" and we want to forecast y_t. We then write $y_t|I$ meaning "the random variable y_t given what we know in I". The set of values that this random variable could take, along with their relative probabilities, is known as the "probability distribution" of $y_t|I$. In forecasting, we call this the forecast distribution.

When we talk about the "forecast", we usually mean the average value of the forecast distribution, and we put a "hat" over y to show this. Thus, we write the forecast of y_t as $\hat{y}_t$, meaning the average of the possible values that y_t could take given everything we know.

It is often useful to specify exactly what information we have used in calculating the forecast. Then we will write, for example, $\hat{y}_{t|t-1}$ to mean the forecast of y_t taking account of all previous observations $(y_1, ..., y_{t-1})$. Similarly, $\hat{y}_{T+h|T}$ means the forecast of y_{T+h} taking account of $y_1, ..., y_T$ (i.e., an h-step forecast taking account of all observations up to time T).