Getting started

Forecasting: principles and practice

book by Rob Hyndman and George Athanasopoulos
slides by Peter Fuleky, modified by Joseph ALBA

History of forecasting

- Divine inspiration or criminal activity?
- 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:
  - *I think there is a world market for maybe five computers. (Chairman of IBM, 1943)*
  - *Computers in the future may weigh no more than 1.5 tons. (Popular Mechanics, 1949)*
  - *There is no reason anyone would want a computer in their home. (President, DEC, 1977)*

Forecasting as a science

Forecasting is obviously a difficult activity, and businesses that do it well have a big advantage over those whose forecasts fail.

In this course, we will explore the most reliable methods for producing forecasts. The emphasis will be on methods that are replicable and testable, and have been shown to work.

The forecasting is required in many situations

Should we:

- build another power plant;
- hire staff for call center;
- stock more inventory?

Some things are easier to forecast than others. The predictability of an event or a quantity depends on several factors including:

- how well we understand the factors that contribute to it;
- how much data are available;
- whether the forecasts can affect the thing we are trying to forecast.

Examples

- Electricity demand
  - *Driven largely by temperatures, with smaller effects for calendar variation such as holidays, and economic conditions.*
  - *With sufficient history of data, and the skills to develop a good model linking electricity demand and the key driver variables, the forecasts can be remarkably accurate.*
- Currency exchange rates
  - *Plenty of data.*
  - *We have a very limited understanding of the factors that affect exchange rates, and forecasts of the exchange rate have a direct effect on the rates themselves.*

Signal and noise

- 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.
- Forecasts do not assume an unchanging environment. What is normally assumed is that *the way in which the environment is changing* will continue into the future.
- A forecasting model is intended to capture the way things move, not just where things are.
- 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”.

Choice of method

- Forecasting situations vary widely in their
  - *time horizons,*
  - *factors determining actual outcomes,*
  - *types of data patterns,*
  - *data availablility, and many other aspects.*
- Forecasting methods can be very simple such as using the most recent observation as a forecast (which is called the "naive method"), or highly complex.
- The choice of method depends on what data are available and the predictability of the quantity to be forecast.

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.

Short-term forecasts

are needed for the scheduling of personnel, production and transportation. As part of the scheduling process, forecasts of demand are often also required.

Medium-term forecasts

are needed to determine future resource requirements, in order to purchase raw materials, hire personnel, or buy equipment.

Long-term forecasts

are used in strategic planning. Such decisions must take account of market opportunities, environmental factors, and internal resources.

Forecasting vs. planning and goals

Forecasting helps to inform decisions about the scheduling of production, transportation and personnel, and provides a guide to long-term strategic planning. It is different from planning and having goals.

Goals

: are what you would like to have happen. Goals should be linked to forecasts and plans, but this does not always occur. Too often, goals are set without any plan for how to achieve them, and no forecasts for whether they are realistic.

Planning

: is a response to forecasts and goals. Planning involves determining the appropriate actions that are required to make your forecasts match your goals.

What to forecast?

For example, if forecasts are required for items in a manufacturing environment, it is necessary to ask whether forecasts are needed for:

- every product line, or for groups of products?
- every sales outlet, or for outlets grouped by region, or only for total sales?
- weekly data, monthly data or annual data?
- one month, six months, or ten years in advance?
- daily use, updated frequently?

Next: collect data.

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.

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.

Types of data

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 forecasting problems use either *time series data* (collected at regular intervals over time) or *cross-sectional data* (collected at a single point in time).

Cross-sectional forecasting

With cross-sectional data, we want to predict the value of something we have not observed, using the information on the cases that we have observed. Examples of cross-sectional data include:

- House prices for all houses sold in 2011 in a particular area. We are interested in predicting the price of a house not in our data set using various house characteristics: position, no. bedrooms, age, etc.
- Fuel economy data for a range of 2009 model cars. We are interested in predicting the carbon footprint of a vehicle not in our data set using information such as the size of the engine and the fuel efficiency of the car.

Cross-sectional models are used when the variable to be forecast exhibits a relationship with one or more other predictor variables.

Time series forecasting

Time series data are useful when you are forecasting something that is changing over time (e.g., stock prices, sales figures, profits, etc.). Examples of time series data include:

- Daily IBM stock prices
- Monthly rainfall
- Quarterly sales results for Amazon
- Annual Google profits

Anything that is observed sequentially over time is a time series. 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 course.

We focus on time series data for this course.

Example of a simple time series forecast

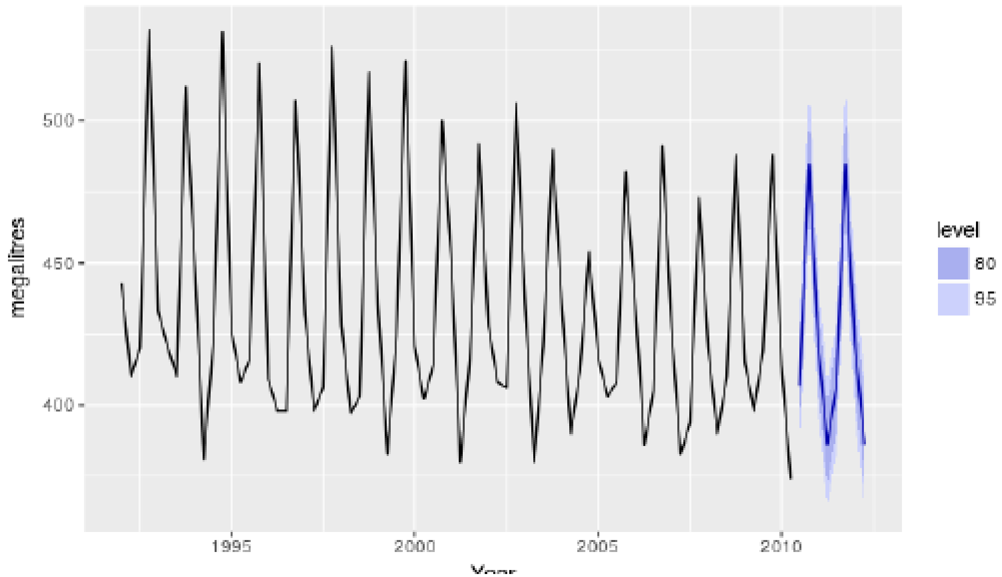


Figure 1.1: Australian quarterly beer production: 1992Q1–2010Q2, with two years of forecasts.

Predictor variables and time series forecasting

Predictor variables can also be used in time series forecasting. For example, suppose we wish to forecast the hourly electricity demand (ED) of a hot region during the summer period by

*ED = f(current temperature, strength of economy, population, time of day, day of week, **error**).*

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

$ED_{t+1} = f(ED_t, ED_{t-1}, ED_{t-2}, ED_{t-3}, \ldots, \textbf{error})$,

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.

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*$_t$*, current temperature, time of day, day of week,…,* ***error****).*

These types of mixed models have been given various names in different disciplines. In economics, we refer to these models as dynamic regression models.

Considerations for model choice

An explanatory model is very 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 model.

1. 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 behavior.
2. It is necessary to know or forecast the various predictors in order to be able to forecast the variable of interest, and this may be too difficult.
3. The main concern may be only to predict what will happen, not to know why it happens.
4. 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 how the forecasting model is to be used.

Typical notations in the literature :

- For cross-sectional data, subscript i is used to indicate a specific observation, e.g. y_i. N is used to denote the total number of observations in the data set.
- For time series data, subscript t is used instead of i: y_t will denote the observation at time t. T is used to denote the number of observations in a time series.
- As mentioned, we will focus on methods that use time series data.

Case studies

There are four case studies in the book (We will refer to these cases in future discussions) .

The basic steps in a forecasting task

1. Problem definition.

- understanding of the way the forecast will be used
- knowing who requires the forecast
- how the function of forecasting fits within the organization
- definition will involve identifying who will be involved in
  - *collecting the data*
  - *maintaining the database*
  - *using the forecast for future planning*

2. Gathering information

- statistical data
- expertise of those who collect the data and use the forecast

3. Preliminary (exploratory) analysis. **Graphs!**

- patterns
- trends
- seasonality
- cycles
- outliers

4. Choosing and fitting models. 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.

5. Using and evaluating a forecasting model.

The statistical forecasting perspective

- The thing we are trying to forecast is unknown (or we wouldn't be forecasting it), and so we can think of it as a *random 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 *point forecast*, we are estimating the middle of the range of possible values the random variable could take.

- Very often, a forecast is accompanied by a *prediction interval* giving a range of values the random variable could take with relatively high probability.

Information set in time series forecasting

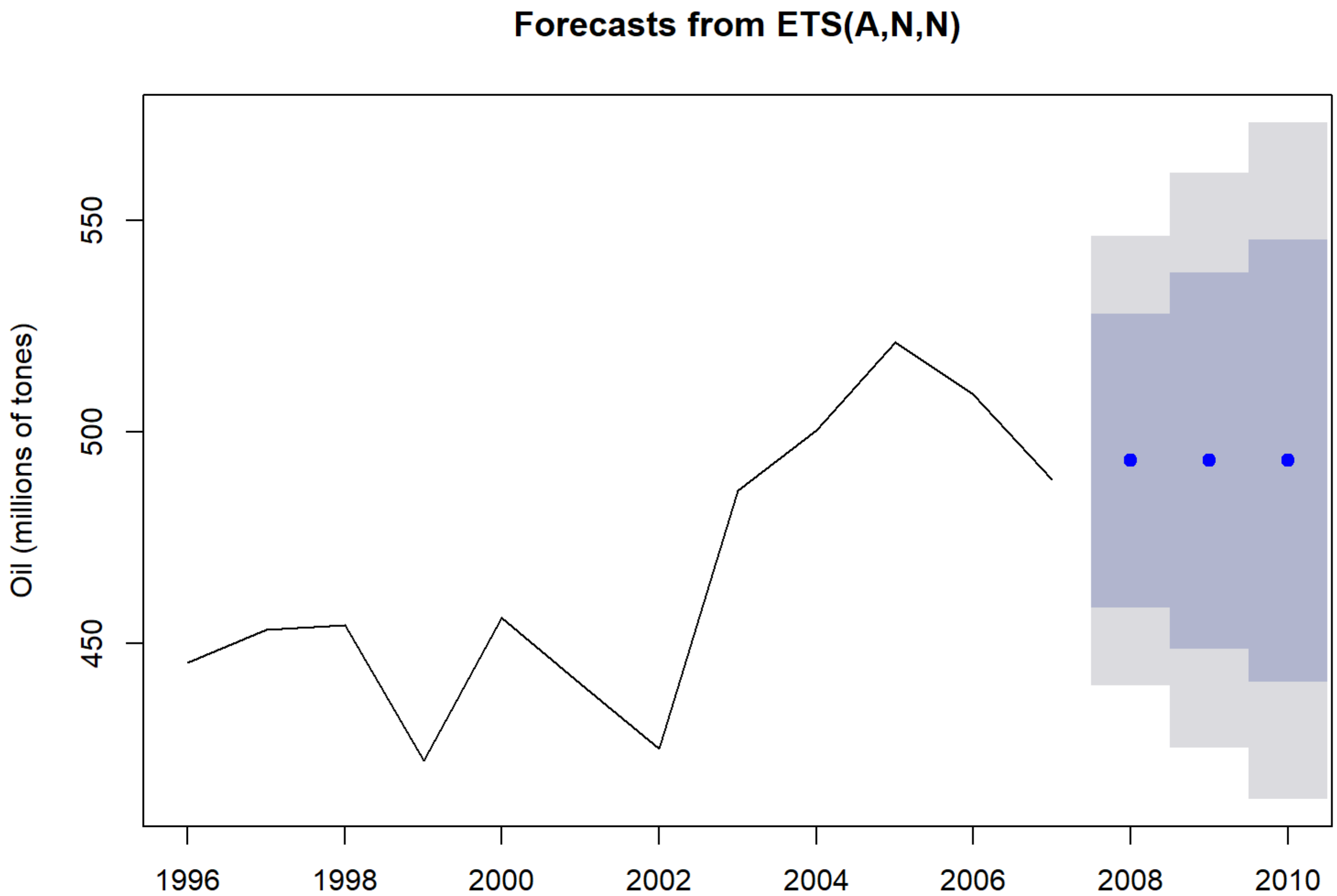
- A forecast is always based on some observations. Suppose we denote all the information we have observed as $\mathcal{I}$ and we want to forecast y_t. We then write $y_t|\mathcal{I}$ meaning "the random variable y_t given what we know in $\mathcal{I}$".

- The set of values that this random variable (y_t) could take, along with their relative probabilities, is known as the *probability distribution* of $y_t|\mathcal{I}$. 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.

- With time series forecasting, it is often useful to specify exactly what information we have used in calculating the forecast. $\hat{y}_{T+h|T}$ means the forecast of y_{T+h} taking account of $y_1, \ldots, y_T$ (i.e., an h-step forecast taking account of all observations up to time T).

Example


Forecasts from ETS(A,N,N)
Oil (millions of tones)
450
500
550
1996
1998
2000
2002
2004
2006
2008
2010

```
##       alpha           l
##   0.7958197 446.7849382
```

End of Lecture 1

Thank you for your attention