Ch3 | The Forecasters Toolbox

Michael Rose

November 3, 2018

Chapter 3 | The Forecaster's Toolbox

In this chapter we will discuss:

- General tools useful for different forecasting situations
- benchmark forecasting methods
- transformations to make forecasting simpler
- methods for checking whether a forecasting method has utilized the available information
- techniques for computing prediction intervals

3.1 | Some simple forecasting methods

We will use the following four forecasting methods as benchmarks througout this book:

Average Method

All the forecasts of all future values are equal to the average of the historical data. Let the historical data be $y_1, ..., y_T$. Then

$$y_{T+h|T} = \bar{y} = (y_1 + ... + y_T)/T \text{ where } y_{T+h|T} \text{ is the estimate of } y_{t+h} \text{ based on the data } y_1, ..., y_T$$

```r
# y contains the time series
# h is the forecast horizon

# mean(y, h)
```

Naive Method

We can simply set all forecasts to be the value of the last observation.

$$y_{T+h|T} = y_T$$

This method works well for many economic and financial time series

```r
# naive(y, h)
# rwf(y, h)  equivalent alternative
```

Because a naive forecast is optimal when data follow a random walk, these are also called **Random Walk Forecasts**.

Seasonal Naive Method

When our data is highly seasonal, we can set our forecast to be equal to the last observed value from the same season of the year.

$y_{T+h|T} = y_{T+h-m(k+1)}$ where m is the seasonal period k is the integer part of $\frac{h-1}{m}$ (the number of complete years in the forecast period prior to time $T+h$)

```
# snaive(y, h)
```

Drift Method

A variation on the naive method is to allow the forecasts to increase or decrease over time, where the amount of change over time (called the **drift**) is set to be the average change seen in the historical data.

$$y_{T+h|T} = y_T + \frac{h}{T-1}\sum_{t=2}^{T}(y_t - y_{t-1}) = y_T + h(\frac{y_T - y_1}{T-1})$$

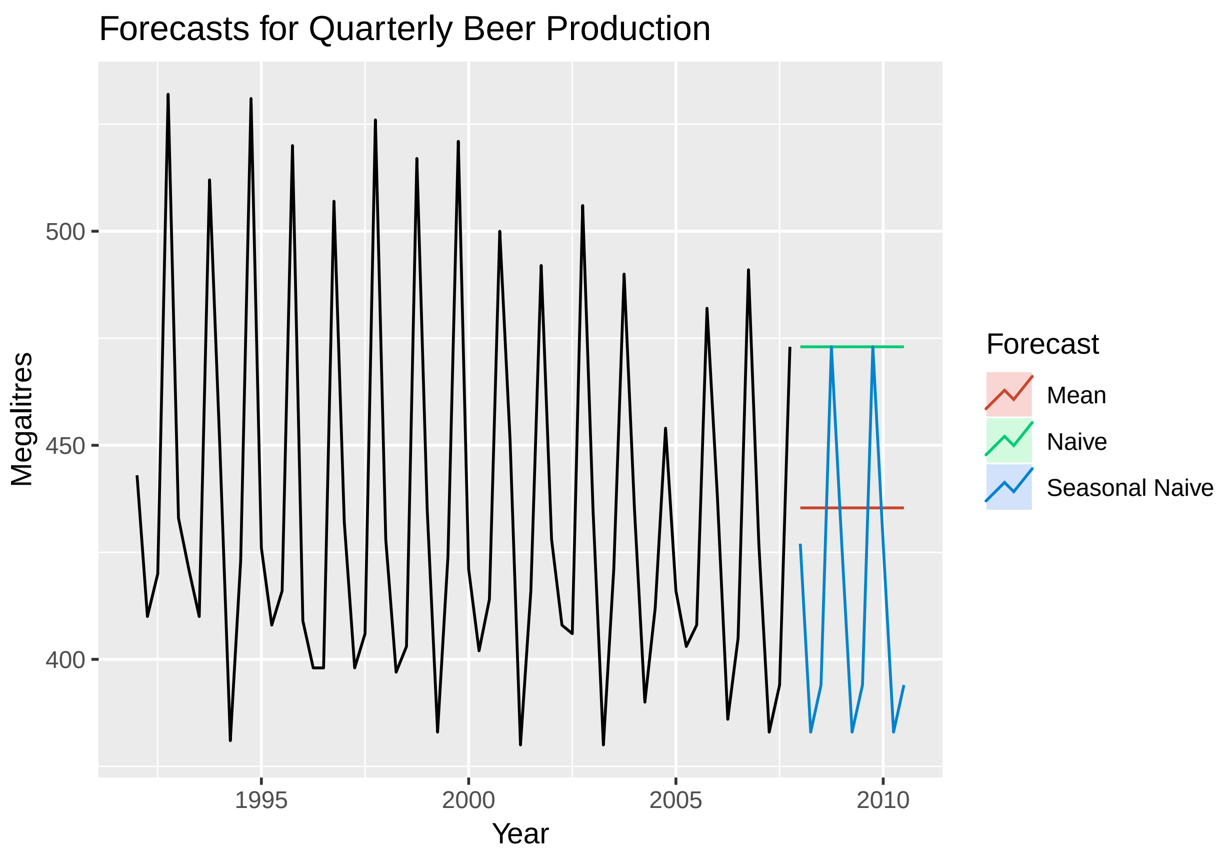
This is equivalent to drawing a line between the first and last observations, and extrapolating it into the future.

```
# rwf(y, h, drift = TRUE)
```

Examples

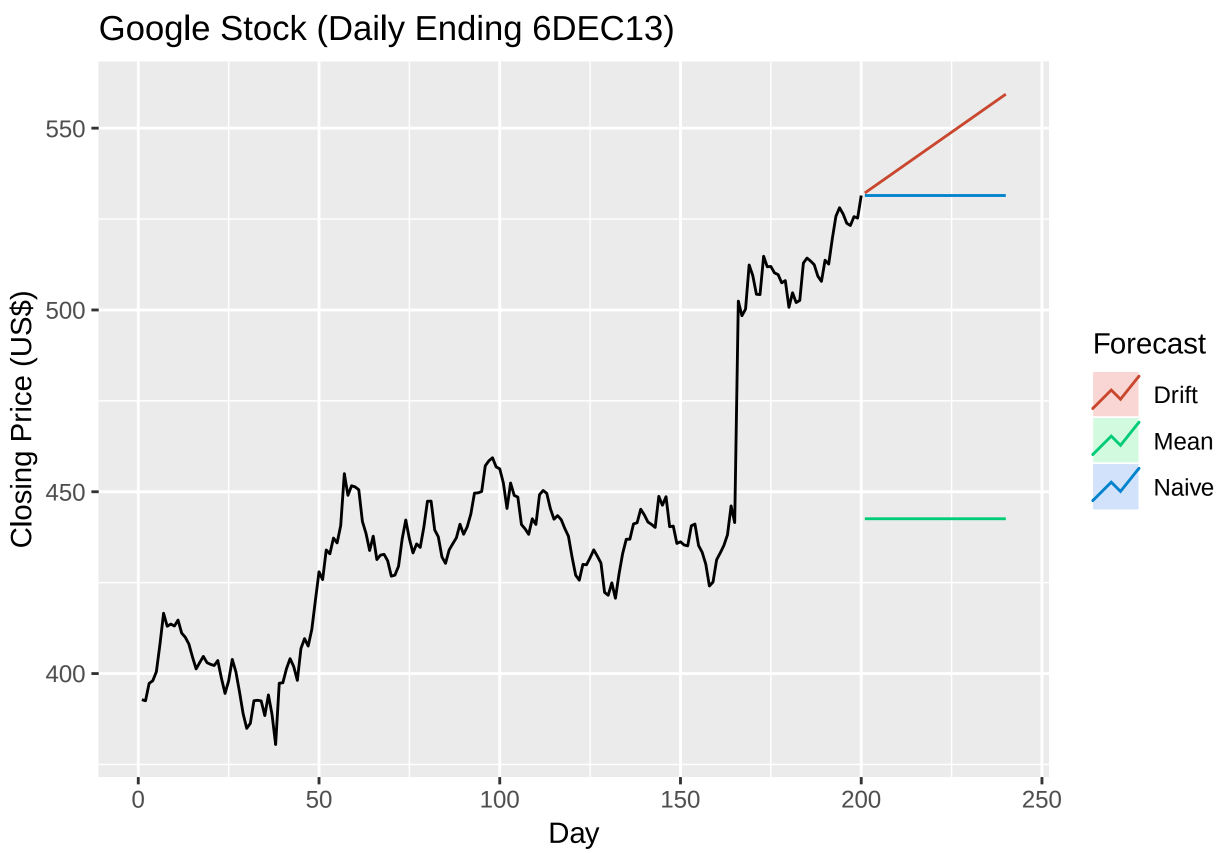
```
# set training data from 1992 to 2007
beer2 <- window(ausbeer, start = 1992, end = c(2007, 4))

# plot forecasts for first 3 methods
autoplot(beer2) +
  autolayer(meanf(beer2, h = 11), series = "Mean", PI = FALSE) +
  autolayer(naive(beer2, h=11), series = "Naive", PI=FALSE) +
  autolayer(snaive(beer2, h=11), series="Seasonal Naive", PI=FALSE) +
  ggtitle("Forecasts for Quarterly Beer Production") +
  xlab("Year") + ylab("Megalitres") +
  guides(colour = guide_legend(title = "Forecast"))
```



Here we apply the non-seasonal methods to a series of 200 days of Google daily closing stock price:

```
autoplot(goog200) +
  autolayer(meanf(goog200, h = 40), series = "Mean", PI=FALSE) +
  autolayer(rwf(goog200, h = 40), series = "Naive", PI=FALSE) +
  autolayer(rwf(goog200, h = 40, drift = TRUE), series = "Drift", PI=FALSE) +
  ggtitle("Google Stock (Daily Ending 6DEC13)") +
  xlab("Day") + ylab("Closing Price (US$)") +
  guides(colour = guide_legend(title = "Forecast"))
```



These methods are generally considered benchmarks, but they may be the most effective in some cases. When we develop a new method, they must be better than these benchmarks – else they aren't worth considering.

3.2 | Transformations and Adjustments