Lesson 3.1

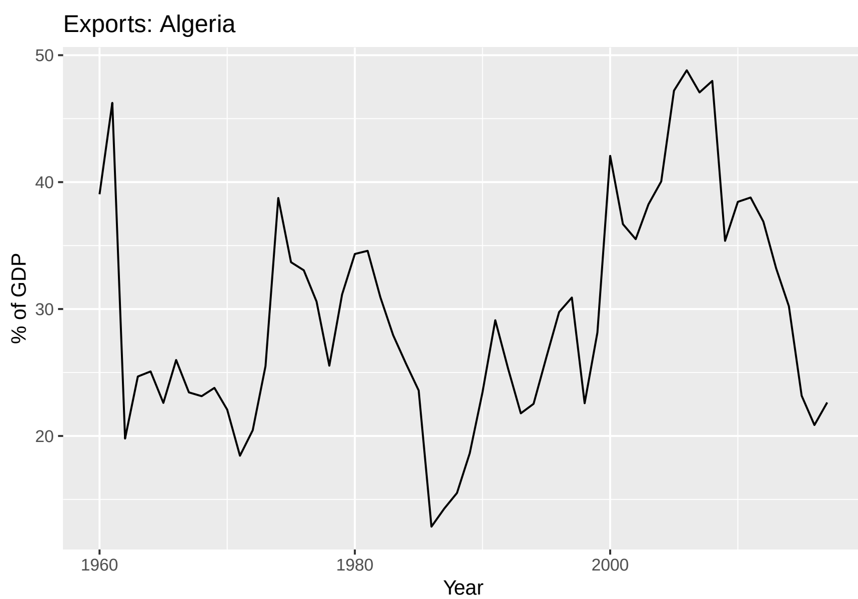
Simple Exponential Smoothing

Introduction

Exponential smoothing was proposed in the late 1950s (Brown, 1959; Holt, 1957; Winters, 1960), and has motivated some of the most successful forecasting methods. Forecasts produced using exponential smoothing methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older. In other words, the more recent the observation the higher the associated weight. This framework generates reliable forecasts quickly and for a wide range of time series, which is a great advantage and of major importance to applications in industry.

Simple Exponential Smoothing

The simplest of the exponentially smoothing methods is naturally called **simple exponential smoothing (SES)**. This method is suitable for forecasting data with no clear trend or seasonal pattern.



Recall the naive forecasting method wherein the forecast at time $T + h$ is given by

$$\hat{y}_{T+h|T} = y_T$$

for forecast horizon $h = 1, 2, 3, \ldots$. Hence, the naive method assumes that the most recent observation is the only important one, and all previous observations provide no information for the future. This can be thought of as a weighted average where all of the weight is given to the last observation.

Another forecasting method is the average method where all future forecasts are equal to a simple average of the observed data, that is,

$$\hat{y}_{T+h|T} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

for forecast horizon $h = 1, 2, 3, \ldots$. Hence, the average method assumes that all observations are of equal importance, and gives them equal weights when generating forecasts.

We often want something between these two extremes. For example, it may be sensible to attach larger weights to more recent observations than to observations from the distant past. This is exactly the concept behind simple exponential smoothing. Forecasts are calculated using weighted averages, where the weights decrease exponentially as observations come from further in the past — the smallest weights are associated with the oldest observations:

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots \tag{1}$$

where α is a constant between 0 and 1 called the smoothing constant (parameter). The one-step-ahead forecast for time $T + 1$ is a weighted average of all of the observations in the series $y_1, y_2, \cdots, y_T$. The rate at which the weights decrease is controlled by the parameter α.

The table below shows the weights attached to observations for four different values of α when forecasting using simple exponential smoothing. Note that the sum of the weights even for a small value of α will be approximately one for any reasonable sample size.

	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$
y_T	0.2000	0.4000	0.6000	0.8000
y_{T-1}	0.1600	0.2400	0.2400	0.1600
y_{T-2}	0.1280	0.1440	0.0960	0.0320
y_{T-3}	0.1024	0.0804	0.0384	0.0064
y_{T-4}	0.0819	0.0518	0.0154	0.0013
y_{T-5}	0.0655	0.0311	0.0061	0.0003

For any α between 0 and 1, the weights attached to the observations decrease exponentially as we go back in time, hence the name **exponential smoothing**. If α is small (i.e., close to

0), more weight is given to observations from the more distant past. If α is large (i.e., close to 1), more weight is given to the more recent observations. For the extreme case where $\alpha = 1$, $\hat{y}_{T+1|T} = y_T$, and the forecasts are equal to the naïve forecasts.

We present two equivalent forms of simple exponential smoothing.

Weighted average form

The forrecast at tome $T + 1$ is equal to the weighted average between the most recent observation y_T and the previous forecast $\hat{y}_{T|T-1}$:

$$\hat{y}_{T+1}|T = \alpha y_T + \alpha(1-\alpha)\hat{y}_{T|T-1}$$

where $0 \leq \alpha \leq 1$ is the smoothing parameter. Similarly, we can write the fitted values as

$$\hat{y}_{t+1|t} = \alpha y_t + \alpha(1-\alpha)\hat{y}_{t|t-1}$$

for $t = 1, 2, \cdots, T$. The fitted values are simply one-step forecasts of the training data.

The process has to start somewhere, so we let the first fitted value at time 1 be denoted by l_0 (which will be estimated). Then

$$\begin{aligned}
\hat{y}_{2|1} &= \alpha y_1 + (1-\alpha) l_o \\
\hat{y}_{3|2} &= \alpha y_2 + (1-\alpha)\hat{y}_{2|1} \\
\hat{y}_{4|3} &= \alpha y_3 + (1-\alpha)\hat{y}_{3|2} \\
&\vdots \\
\hat{y}_{T|T-1} &= \alpha y_{T-1} + (1-\alpha)\hat{y}_{T-1|T-2} \\
\hat{y}_{T+1|T} &= \alpha y_T + (1-\alpha)\hat{y}_{T|T-1}
\end{aligned}$$

Substituting each equation into the following equation, we obtain

$$\begin{aligned}
\hat{y}_{3|2} &= \alpha y_2 + (1-\alpha)[\alpha y_1 + (1-\alpha) l_o] \\
&= \alpha y_2 + \alpha(1-\alpha) y_1 + (1-\alpha)^2 l_0 \\
\hat{y}_{4|3} &= \alpha y_3 + (1-\alpha)[\alpha y_2 + \alpha(1-\alpha) y_1 + (1-\alpha)^2 l_0] \\
&= \alpha y_3 + \alpha(1-\alpha) y_2 + \alpha(1-\alpha)^2 y_1 + (1-\alpha)^3 l_0 \\
&\vdots \\
\hat{y}_{T+1|T} &= \sum_{j=0}^{T-1} \alpha(1-\alpha)^j y_{T-j} + (1-\alpha)^T l_0
\end{aligned}$$

The last term becomes tiny for large T. So, the weighted average form leads to the same forecast Equation 1.

Component form

An alternative representation is the component form. For simple exponential smoothing, the only component included is the level, l_t. Other methods which are considered later in this chapter may also include a trend b_t, and a seasonal component s_t. Component form representations of exponential smoothing methods comprise a forecast equation and a smoothing equation for each of the components included in the method. The component form of simple exponential smoothing is given by:

$$\begin{aligned} \text{Forecast equation:} \quad \hat{y}_{t+h|t} &= l_t \\ \text{Smoothing equation:} \quad l_t &= \alpha y_t + (1-\alpha) l_{t-1} \end{aligned}$$

where l_t is the level (or the smoothed value) of the series at time t. Setting $h = 1$ gives the fitted values, while setting $t = T$ gives the true forecasts beyond the training data.

The forecast equation shows that the forecast value at time $t + 1$ is the estimated level at time t. The smoothing equation for the level (usually referred to as the level equation) gives the estimated level of the series at each period t.

If we replace l_t with $\hat{y}_{t+1|t}$ and l_{t-1} with $\hat{y}_{t|t-1}$ in the smoothing equation, we will recover the weighted average form of simple exponential smoothing.

The component form of simple exponential smoothing is not particularly useful, but it will be the easiest form to use when we start adding other components.

The application of every exponential smoothing method requires the smoothing parameters and the initial values to be chosen. In particular, for simple exponential smoothing, we need to select the values of α and l_0. All forecasts can be computed from the data once we know those values. For the methods that follow there is usually more than one smoothing parameter and more than one initial component to be chosen.

In some cases, the smoothing parameters may be chosen in a subjective manner — the forecaster specifies the value of the smoothing parameters based on previous experience. However, a more reliable and objective way to obtain values for the unknown parameters is to estimate them from the observed data.

The unknown parameters and the initial values for any exponential smoothing method can be estimated by minimising the SSE.

$$\text{SSE} = \sum_{t=1}^{T} (y_t - \hat{y}_{t|t-1})^2 = \sum_{t=1}^{T} e_t^2$$

This is done using a non-linear minimization.

Example

In this example, simple exponential smoothing is applied to forecast exports of goods and services from Algeria.

```r
library(fpp3)
library(forecast)
library(tsibble)
algeria_exports <- global_economy %>%
  filter(Country == "Algeria") %>%
  as.data.frame() %>%
  select(Exports) %>%
  ts(start = 1960, frequency = 1)

fc <- ses(algeria_exports, h=5)
fc$model$par #Prints the estimated smoothing coefficient and initial state
```

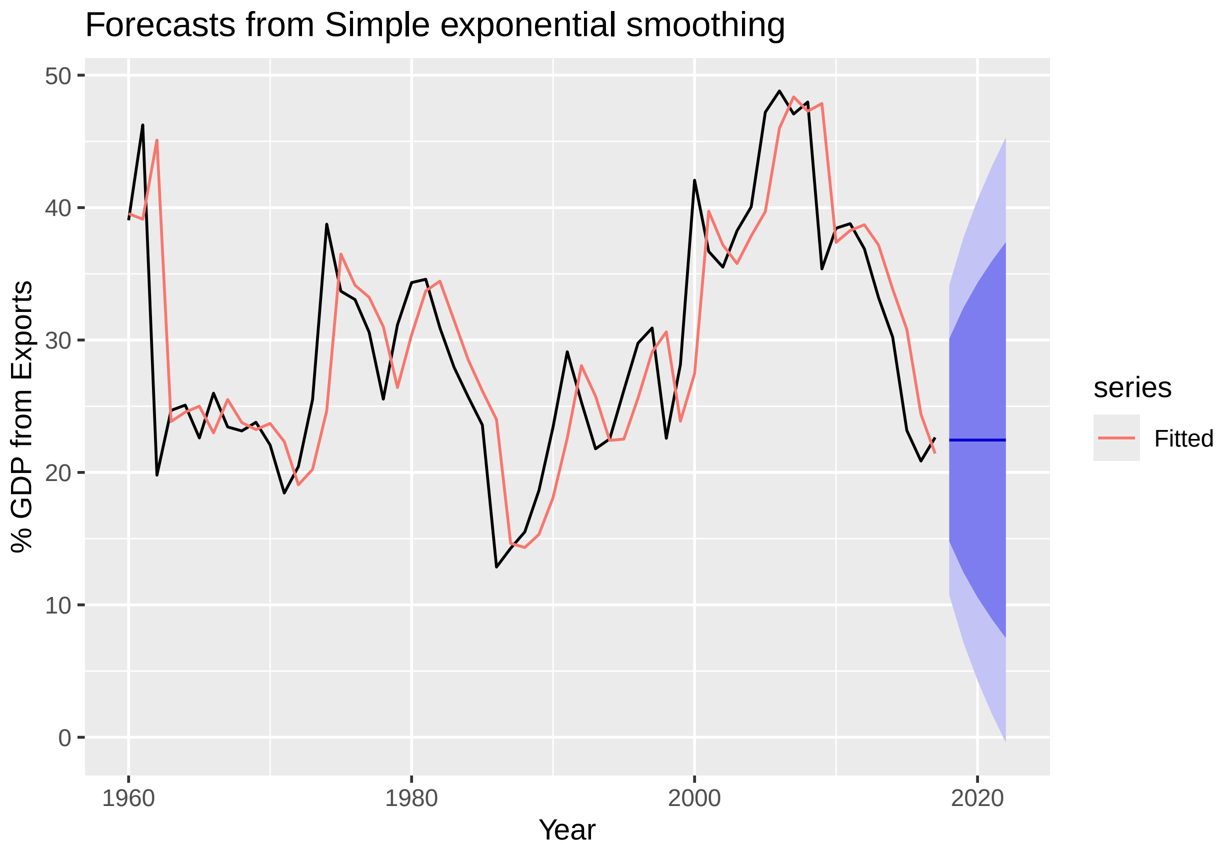
```
##     alpha         l
##  0.839812 39.540126
```

This gives parameter estimates $\alpha \approx 0.84$ and $l_0 \approx 39.54$, obtained by minimizing SSE over periods $t = 1, 2, \cdots, 58$ subject to the restriction $0 \leq \alpha \leq 1$.

The calculations using these parameters are shown below. The second last column shows the estimated level for times $t = 0$ to $t = 58$; the last few rows of the last column show the forecasts for $h = 1, 2, 3, 4, 5$.

Year	Time	Observation	Level	Forecast
	t	y_t	ℓ_t	$\hat{y}_{t\|t-1}$
1959	0		39.54	
1960	1	39.04	39.12	39.54
1961	2	46.24	45.10	39.12
1962	3	19.79	23.84	45.10
1963	4	24.68	24.55	23.84
1964	5	25.08	25.00	24.55
1965	6	22.60	22.99	25.00
1966	7	25.99	25.51	22.99
1967	8	23.43	23.77	25.51
	⋮	⋮	⋮	⋮
2014	55	30.22	30.80	33.85
2015	56	23.17	24.39	30.80
2016	57	20.86	21.43	24.39
2017	58	22.64	22.44	21.43
	h			$\hat{y}_{T+h\|T}$
2018	1			22.44
2019	2			22.44
2020	3			22.44
2021	4			22.44
2022	5			22.44

```r
autoplot(fc) +
  autolayer(fitted(fc), series="Fitted") +
  labs(x = "Year", y = "% GDP from Exports")
```



The forecasts for the period 2018–2022 are plotted in Figure 8.2. Also plotted are one-step-ahead fitted values alongside the data over the period 1960–2017. The large value of α in this example is reflected in the large adjustment that takes place in the estimated level l_t at each time. A smaller value of α would lead to smaller changes over time, and so the series of fitted values would be smoother.

The prediction intervals show that there is considerable uncertainty in the future exports over the five-year forecast period.