Exponential smoothing and change point detection

Søren Højsgaard

January 17, 2013

1 Introduction

2 Simple exponential smoothing

Given is a time series $\{x_t; t = 0, 1, \dots, T\}$ observed at discrete equidistant times. Exponential smoothing or single exponential smoothing (hereafter SES) is a dynamic smoothing of data which for $0 < \alpha < 1$ is defined as

$$S_t = \alpha x_t + (1-\alpha)S_{t-1} = S_{t-1} + \alpha(x_t - S_{t-1}) \tag{1}$$

Suppose that the underlying process is "almost constant", that is $x_t = a$. Then the forecast of x at time $t+h$ based on data upto (and including) time t is

$$\hat{x}_{t+h|t} = S_t \tag{2}$$

Write $\hat{x}_{t|t} = S_t$ as $\hat{x}_t$ which we may think of as the current fitted value. Notice that (1) also reads $S_t = S_{t-1} + \alpha e_t$ where $e_t = x_t - S_{t-1} = x_t - \hat{x}_{t|t-1}$ is the forecast error.

We may write S_t given in (1) as a linear function of data. Letting $\beta = 1 - \alpha$ we get

$$S_t = \alpha \sum_{k=0}^{t-1} \beta^k x_{t-k} + \beta^t x(0) \tag{3}$$

Hence, the influence of the initial value becomes negligible for large t. Moreover, the sequence of weights $(\alpha, \alpha\beta, \alpha\beta^2, \dots, \alpha\beta^{t-1})$ decrease exponentially and will in the limit sum to one. Hence S_t can be regarded as a probability weighted average of data where recent observations carry the highest weight.

In practice a value for α must be chosen. A data–driven way is to choose α so as to minimize sum of squares of 1–step ahead forecasts, that is:

$$\sum_{t=2}^{T} (x_t - \hat{x}_{t|t-1})^2$$

3 Double exponential smoothing

Double exponential smoothing (hereafter DES) is SES applied to S_t, that is

$$S_t^{[2]} = \alpha S_t + (1-\alpha) S_{t-1}^{[2]} \tag{4}$$

Suppose that the true process is "almost linear", that is

$$x_t = a + bt$$

Then S_t will be

$$S_t = a + bt + \frac{\beta}{\alpha} b + O\left((1-\alpha)^t\right)$$

Then SES will be biased in the sense that

$$x_t - S_t = b\frac{\beta}{\alpha} \text{ for } t \to \infty \tag{5}$$

or $x_t \approx S_t + b\frac{\beta}{\alpha}$.
By the same argument we get that smoothing S_t again gives

$$S_t^{[2]} = (a - 2b\frac{\beta}{\alpha}) + bt \text{ for } t \to \infty$$

Hence $S_t^{[2]} - S_t \to b\frac{\beta}{\alpha}$. From these considerations and (5) we therefore have

$$\begin{aligned} b &= \{S_t - S_t^{[2]}\}\frac{\alpha}{\beta} \qquad (6) \\ x_t &= S_t - b\beta/\alpha = S_t + (S_t - S_t^{[2]}) = 2S_t - S_t^{[2]} \qquad (7) \end{aligned}$$

Hence natural estimates of levels and slopes become

$$\begin{aligned} \hat{x}_t &= S_t + \{S_t - S_t^{[2]}\} = 2S_t - S_t^{[2]} \qquad (8) \\ \hat{b}_t &= \{S_t - S_t^{[2]}\}\frac{\alpha}{\beta} \qquad (9) \end{aligned}$$

Similarly, the forecasts become

$$\hat{x}(t+h|t) = \hat{x}_t + \hat{b}_t h \tag{10}$$

4 Triple exponential smoothing

Triple exponential smoothing (hereafter TES) is SES applied to $S_t^{[2]}$, that is

$$S_t^{[3]} = \alpha S_t^{[2]} + (1-\alpha) S_{t-1}^{[3]} \tag{11}$$

Suppose the original proces is quadratic

$$x_t = a + bt + ct^2$$

Then S_t will be

$$S_t = a + bt + ct^2 + 2\,\frac{(-1+\alpha)\,ct}{\alpha} + \frac{-b\alpha - 3\,c\alpha + 2\,c + c\alpha^2 + b\alpha^2}{\alpha^2} + O\left((1-\alpha)^t\right)$$

or

$$x_t - S_t = 2\,\frac{(-1+\alpha)\,ct}{\alpha} + \frac{-b\alpha - 3\,c\alpha + 2\,c + c\alpha^2 + b\alpha^2}{\alpha^2} + O\left((1-\alpha)^t\right)$$

A Linear filters and the z–transform

We can look at (3) as a linear time invariant filter

$$S_t = \sum_{j=-\infty}^{\infty} h(j)x(t-j) \tag{12}$$

where $h(j) = \alpha\beta^j$ for $j \geq 0$ and zero otherwise. Hence $h(j)$ is the weight given to $x(t-j)$ in forming S_t or, equivalently, the effect of an observation x_t at time t will j time steps later be $x_t\alpha\beta^j$.

Let $\{h(j)\}$ for $j = -\infty \ldots, \infty$ be a collection of weights and define

$$y_t = \sum_{j=-\infty}^{\infty} h(j)x(t-j) \tag{13}$$

Then (13) is a linear time invariant filter. As an example, suppose that $h(0)$, $h(1)$ and $h(2)$ are non–zero and all other weights are zero. Then we get

$$y_t = h(0)x_t + h(1)x_{t-1} + h(2)x(t-2) \tag{14}$$

Equation (13) is called a convolution of $x()$ and $h()$ and we write this briefly as

$$y_t = x_t * h_t \tag{15}$$

We want to study what y_t looks like for different choices of x_t and h_t. In doing so we use the z–transform: Let $f(n)$ be a function defined for integer values of n. The z–transform is

$$F(z) = \sum_{n=-\infty}^{\infty} f(n)z^n \tag{16}$$

for complex values z. Notice that we write the function in lower case and the z–transform of the function in upper case. Tables exist for pairs of functions and their z–transform.

We apply the z–transform to (14). Change variables from as $m = t - j$ so that $j = t - m$ so that

$$y(n) = \sum_{m=-\infty}^{\infty} x(m)h(n-m) \tag{17}$$

Now apply the z–transform to $y(n)$ and we get (where all summations are from $-\infty$ to ∞):

$$\begin{aligned} Y(z) &= \sum_n z^n \sum_m x(m)h(n-m) \\ &= \sum_m \{z^m x(m) \sum_n z^{n-m} h(n-m)\} \\ &= \sum_m \{z^m x(m) \sum_p z^p h(p)\} \\ &= H(z)X(z) \end{aligned} \tag{18}$$

Hence, all we need to do is to find $H()$ and $X()$ (typically by looking into a table), carry out the multiplication in (18) to find $Y(z)$ and then find y_t by looking into the table again.

There are linearities to be exploited in this connection: If $f(n) = \sum_u \alpha_u f_u(n)$ then $F(z) = \sum_u \alpha_u F_u(z)$. Thus if $x_t = \sum_u \alpha_u(t)$ and $h(j) = \sum_v \beta_v h_v(j)$ then

$$Y(z) = \sum_{u,v} \alpha_u \beta_v X_u(z) H_v(z) = \sum_s \gamma_s Y_s(z), \tag{19}$$

say. Exploiting the linearity again then gives

$$y_t = \sum_s \gamma_s y_s(t) \tag{20}$$