Chapter 4: Exponential Smoothing Methods

Rong Fan

South Dakota State University

rong.fan@sdstate.edu

A data set consists of signal component and noise component.

The constant process is represented as

$$y_t = \beta_0 + \epsilon_t$$

where μ represents the constant level and ϵ_t is the noise at time t.

Smoothing can be viewed as a process to separate the signal and the noise. We have discussed some smoothers, e.g. linear filters which include simple moving average of span N.

$$M_T = \frac{1}{N} \sum_{t=T-N+1}^{T} y_t$$

The smoother reconstruct the signal to some extend.

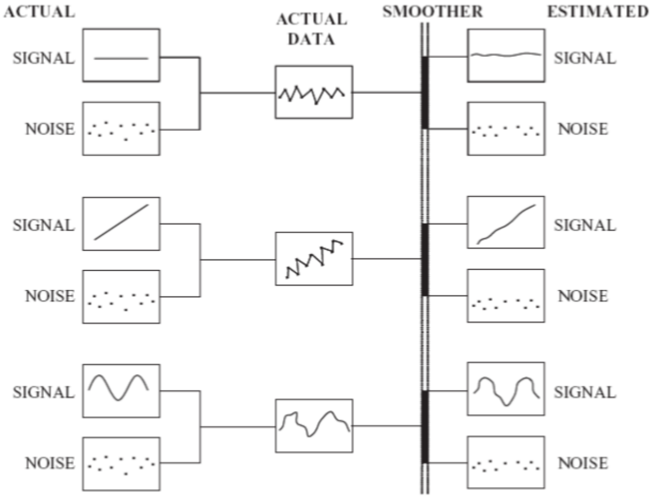


FIGURE 4.1 The process of smoothing a data set.

4.1 Introduction III

We will only talk about backward looking smoothers meaning the smoothers replace observation y_T with a combination of observations at and before time T.

On the other hand, there is forward looking smoothers.

A choice for the backward looking smoothing is to replace the current observation y_T with the average of all observations up to time T.

$$\tilde{y}_T = \frac{1}{T} \sum_{t=1}^{T} y_t.$$

This is particularly good for a constant process.
In a long run, there will be a change happening in the process, so earlier data no longer carry the information about the change, but they still contribute to the smoother at an equal proportion.

4.1 Introduction IV

We can either change the weights so earlier observations are weighted less, or we can totally remove them, e.g. the simple moving average only depends on the observations as early as $t = T - N + 1$, where N is the span.

Then the selection of N becomes important. Next picture shows a comparison.

4.1 Introduction V

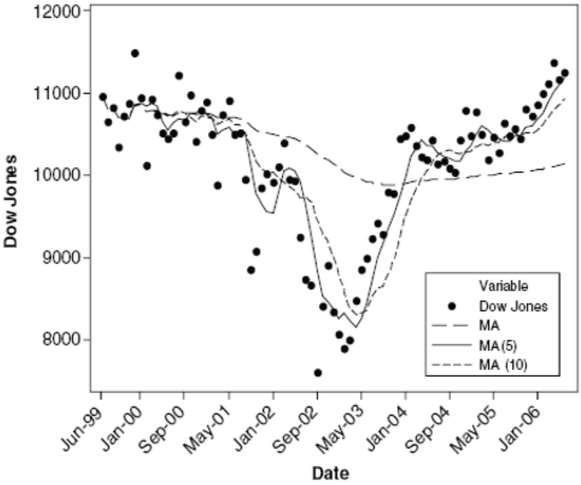


FIGURE 4.4 The Dow Jones Index from June 1999 to June 2006 with moving averages of span 5 and 10.

4.2 First-Order Exponential Smoothing

Consider a **smoother with weights** that are decreasing geometrically to the past observations.

$$\sum_{t=0}^{T-1} \theta^t y_{T-t} = y_T + \theta y_{T-1} + \theta^2 y_{T-2} + \ldots + \theta^{T-1} y_1$$

Here θ is a discount factor that satisfies $|\theta| < 1$. In order to make the sum of all weights equal 1, we multiply each term by $(1-\theta)$ to obtain

$$\tilde{y}_T = (1-\theta) \sum_{i=0}^{T-1} \theta^i y_{T-i}$$

which is called a **simple** or **first-order exponential smoother**.

4.2 First-Order Exponential Smoothing

It can be shown that this first-order exponential smoother can be defined equivalently as

$$\tilde{y}_T = (1 - \theta) y_T + \theta \tilde{y}_{T-1}$$

This recursive relationship allows us to update the smoother at time T by just using the values on the previous time period $T - 1$.

By letting $\lambda = 1 - \theta$, the smoother can be represented as

$$\tilde{y}_T = \lambda y_T + (1 - \lambda) \tilde{y}_{T-1}$$

4.2.1 The initial value $\tilde{y}_0$ I

Since in the beginning of the process, you have

$$\tilde{y}_1 = \lambda y_1 + (1 - \lambda)\tilde{y}_0,$$

which includes a value $\tilde{y}_0$ that is not observable, there are two common ways to assign a value to it:

1. Set $\tilde{y}_0 = y_1$.
2. Set $\tilde{y}_0 = \bar{y}$, which is the average of all observations available.

In any case, for large data sets, the initial value of $\tilde{y}_0$ has little effect as T gets large.

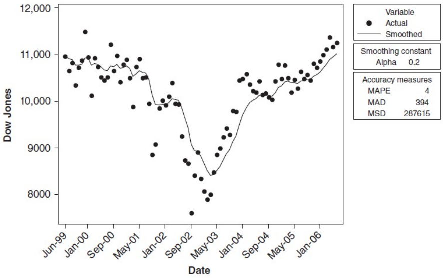


FIGURE 4.5 The Dow Jones Index from June 1999 to June 2006 with first-order exponential smoothing with $\lambda = 0.2$.

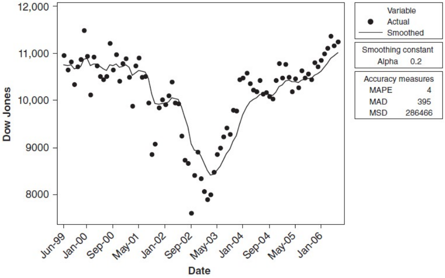


FIGURE 4.7 The Dow Jones Index from June 1999 to June 2006 with first-order exponential smoothing with $\lambda = 0.2$ and $\tilde{y}_0 = (\sum_{t=1}^{25} y_t/25)$ (i.e., initial value equal to the average of the first 25 observations).

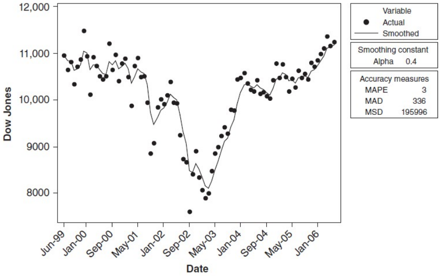


FIGURE 4.6 The Dow Jones Index from June 1999 to June 2006 with first-order exponential smoothing with $\lambda = 0.4$.

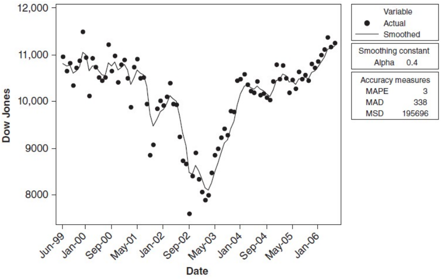


FIGURE 4.8 The Dow Jones Index from June 1999 to June 2006 with first-order exponential smoothing with $\lambda = 0.4$ and $\tilde{y}_0 = \left(\sum_{t=1}^{25} y_t/25\right)$ (i.e., initial value equal to the average of the first 25 observations).

4.2.2 The value of λ I

Different choices of λ values result in different behaviours of the smoother. For example, a comparison between $\lambda = 0.2$ and $\lambda = 0.4$ is as follows:

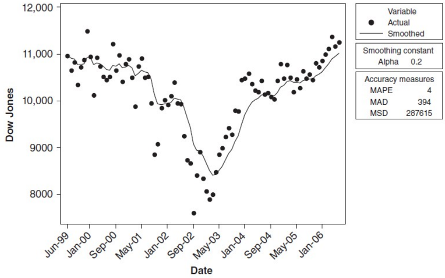


FIGURE 4.5 The Dow Jones Index from June 1999 to June 2006 with first-order exponential smoothing with $\lambda = 0.2$.

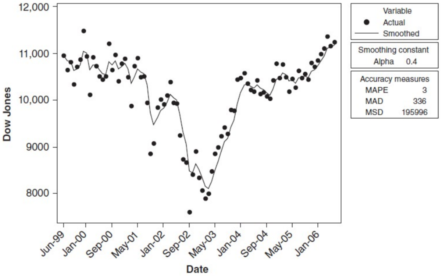


FIGURE 4.6 The Dow Jones Index from June 1999 to June 2006 with first-order exponential smoothing with $\lambda = 0.4$.

4.2.2 The value of λ II

$$\tilde{y}_T = \lambda y_T + (1 - \lambda)\tilde{y}_{T-1} \tag{1}$$

- If $\lambda = 0$, the smoothed values will be equal to a constant. The constant line has the "smoothest" version of whatever pattern the actual time series follows.
- If $\lambda = 1$, we have $\tilde{y}_T = y_T$ and this will represent the "least" smoothed (or unsmoothed) version of the original time series.
- When λ is getting closer to 1,(which means θ is getting closer to 0) one can see that the smoothed values will be closed to the original observations, therefore the smoothed values are more variable, and vice versa.

4.2.2 The value of λ III

Under the independence ($y_1, y_2, \cdots, y_T$ are independent) and constant variance ($var(y_1) = var(y_2) = \cdots = var(y_T)$) assumptions, we have the variance of the smoothed value $\tilde{y}_T$:

$$Var(\tilde{y}_T) = \frac{\lambda}{2-\lambda} Var(y_T)$$

Therefore the λ determines the degree of smoothness. In the literature, λ values between 0.1 and 0.4 are often recommended and do perform well in practice. We will go back to the selection of λ in later sections.

4.2.2 The value of λ IV

One can see when $\lambda = 0.4$ the smoothed data are following the original observations more "closely".

This can be seen quantitatively. The MAPE, MAD, and MSD defined as follows:

$$MAPE = \frac{\sum_{t=1}^{T} |(y_t - \tilde{y}_{t-1})/y_t|}{T} \times 100, \quad (y_t \neq 0).$$

$$MAD = \frac{\sum_{t=1}^{T} |y_t - \tilde{y}_{t-1}|}{T}.$$

$$MSD = \frac{\sum_{t=1}^{T} (y_t - \tilde{y}_{t-1})^2}{T}.$$

1. for case $\lambda = 0.2$, 4, 395, and 286466 respectively.
2. for case $\lambda = 0.4$, 3, 338, and 195696 respectively.

But following original observation too closely could result in "fitting the noise".

Define a Function in R

If you want to run a collection of computation/command many times in your programming, it is a good idea to define a function. The structure of a function is given as

Format of the definition of a function

myfunction $<-$ function(argument1, argument2, ...) {

statements

return(object)

}

The object returned can be any data type.

Define a Function in R

An easy example is

Function

myfunction$<-$function(x,y)
{
$z <- x^2 + y^2$
$t <- x - y$
return(cbind(z,t))
}

Then you can use it as:
myfunction(1:5,2:6) which gives output

```
> myfunction(1:5,2:6)
      z  t
[1,]  5 -1
[2,] 13 -1
[3,] 25 -1
[4,] 41 -1
[5,] 61 -1
```

4.3 Modeling Time Series Data I

Recall a constant process is represented as

$$y_t = \beta_0 + \epsilon_t.$$

The **exponential smoothers can be used to estimate the model**.

Recall:The sum of squared errors for the constant process is given by

$$SS_e = \sum_{t=1}^{T} (y_t - \beta_0)^2. \tag{2}$$

If we argue about that not all observations should have equal influence on the sum and decide to introduce a string of weights that are geometrically decreasing in time, we introduce the sum of weighted squared errors for the constant process:

$$SS_E^* = \sum_{t=0}^{T-1} \theta^t (y_{T-t} - \beta_0)^2 \tag{3}$$

where $|\theta| < 1$. To find the least squares estimator for β_0, we take the derivative of Eq.(3) with respect to β_0 and set it to zero:

$$\left.\frac{dSS_E^*}{d\beta_0}\right|_{\beta_0} = -2\sum_{t=0}^{T-1}\theta^t(y_{T-t} - \hat{\beta}_0) = 0 \tag{4}$$

The solution to (4), $\hat{\beta}_0$, which is the least squares estimate of β_0 is

$$\hat{\beta}_0 \sum_{t=0}^{T-1}\theta^t = \sum_{t=0}^{T-1}\theta^t y_{T-t} \tag{5}$$

From Eq. (5), we have

$$\hat{\beta}_0 = \frac{(1-\theta)}{(1-\theta^T)}\sum_{t=0}^{T-1}\theta^t y_{T-t}. \tag{6}$$

4.3 Modeling Time Series Data IV

For large T, θ^T goes to zero, therefore we set

$$\hat{\beta}_0 = (1-\theta)\sum_{t=0}^{T-1}\theta^t y_{T-t} = \lambda\sum_{t=0}^{T-1}(1-\lambda)^t y_{T-t}. \tag{7}$$

Recall:

$$\tilde{y}_T = (1-\theta)\sum_{t=0}^{T-1}\theta^t y_{T-t} = \lambda\sum_{t=0}^{T-1}(1-\lambda)^t y_{T-t}. \tag{8}$$

Therefore for a constant process,

$$\hat{\beta}_0 = \tilde{y}_T.$$

Therefore the model for a constant process is:

$$\hat{y}_T = \tilde{y}_T \tag{9}$$

4.3 Modeling Time Series Data V

A more general model for a time series process is

$$y_t = f(t, \beta) + \epsilon_t.$$

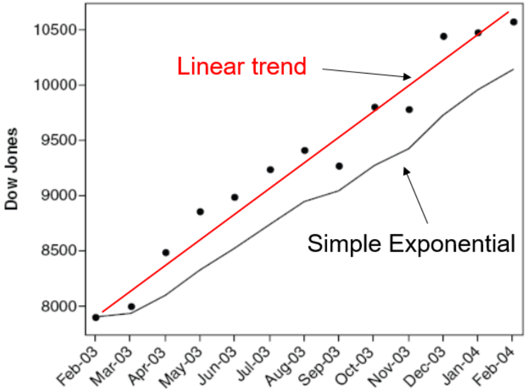
- $f(t, \beta) = \beta_0$, this is just a constant process. We have $\hat{\beta}_0 = \tilde{y}_T$.
- $f(t, \beta) = \beta_0 + \beta_1 t$, this is linear trend model.How to estimate β_0 and β_1?
- $\cdots$

4.4 Second-Order Exponential Smoothing I

Recall a linear trend model is represented as

$$y_t = \beta_0 + \beta_1 t + \epsilon_t, \quad \epsilon_t \sim NID(0, \sigma_\epsilon^2).$$

We can use the simple exponential smoothing procedure to smooth the linear trend. For example, Dow Jones Index data from Feb 2003 to Feb 2004 shows a linear(bullish) trend, then it is smoothed with $\lambda = 0.3$


Linear trend
Simple Exponential
Dow Jones
10500
10000
9500
9000
8500
8000
Feb-03
Mar-03
Apr-03
May-03
Jun-03
Jul-03
Aug-03
Sep-03
Oct-03
Nov-03
Dec-03
Jan-04
Feb-04

4.4 Second-Order Exponential Smoothing I

Notice the smoother captures the slope and shows some bias. In fact, whenever the data exhibit a linear trend, the simple exponential smoother seems to over or under-estimates the data consistently.

4.4 Second-Order Exponential Smoothing II

By computing the expectations, we have

$$
\begin{aligned}
E(\tilde{y}_T) &= E(\lambda \sum_{t=0}^{\infty} (1-\lambda)^t y_{T-t}) && (10)\\
&= \lambda \sum_{t=0}^{\infty} (1-\lambda)^t E(y_{T-t}) && (11)\\
&= \lambda \sum_{t=0}^{\infty} (1-\lambda)^t (\beta_0 + \beta_1 (T-t)) && (12)\\
&= (\beta_0 + \beta_1 T) - \frac{1-\lambda}{\lambda} \beta_1 && (13)\\
&= E(y_T) - \frac{1-\lambda}{\lambda} \beta_1 && (14)
\end{aligned}
$$

which implies the bias is $-\frac{1-\lambda}{\lambda}\beta_1$. For example, when $\beta_1 > 0$ as in Dow Jones example, the bias is negative meaning an underestimate.

4.4 Second-Order Exponential Smoothing

Also notice that when λ is closed to 1, the bias is closed to 0. The following picture shows a comparison between $\lambda = 0.3$ and $\lambda = 0.99$.

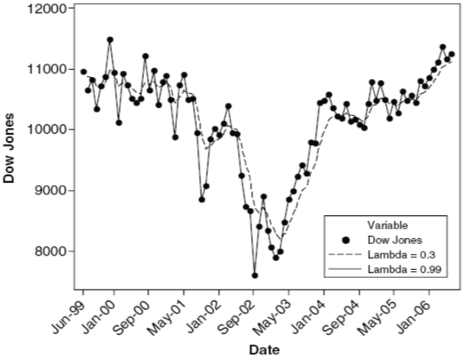


FIGURE 4.11 The Dow Jones Index from June 1999 to June 2006 using exponential smoothing with $\lambda = 0.3$ and 0.99.

Although it has little bias, the smoother with large λ follows the data too closely.

4.4 Second-Order Exponential Smoothing I

We introduce the second order exponential smoothing by applying simple exponential smoothing on $\tilde{y}_T$,

$$\tilde{y}_T^{(2)} = \lambda \tilde{y}_T^{(1)} + (1 - \lambda) \tilde{y}_{T-1}^{(2)} \tag{15}$$

where $\tilde{y}^{(1)}$ and $\tilde{y}^{(2)}$ denote the first and second order smoothed exponential respectively.

4.4 Second-Order Exponential Smoothing II

The second order smoother, which is a first order exponential smoother of the first order exponential smoother, should also be biased.

$$E(\tilde{y}_T^{(2)}) = E(\tilde{y}_T^{(1)}) - \frac{1-\lambda}{\lambda}\beta_1 \tag{16}$$

From (16), an estimator for β_1 at time T is:

$$\hat{\beta}_{1,T} = \frac{\lambda}{1-\lambda}\left(\tilde{y}_T^{(1)} - \tilde{y}_T^{(2)}\right). \tag{17}$$

(Go to section 4.6.2 equation (54).)

4.4 Second-Order Exponential Smoothing III

Recall:

$$E(\tilde{y}_T) = (\beta_0 + \beta_1 T) - \frac{1-\lambda}{\lambda}\beta_1 \tag{18}$$

From (18), we have $\tilde{y}_T^{(1)} = (\hat{\beta}_{0,T} + \hat{\beta}_{1,T} T) - \frac{1-\lambda}{\lambda}\hat{\beta}_{1,T}$. Therefore,

$$\hat{\beta}_{0,T} = \tilde{y}_T^{(1)} - \hat{\beta}_{1,T} T + \frac{1-\lambda}{\lambda}\hat{\beta}_{1,T} \tag{19}$$

Combining (17) and (19), we have a predictor for y_T as

$$\hat{y}_T = \hat{\beta}_{0,T} + \hat{\beta}_{1,T} T \tag{20}$$

$$= 2\tilde{y}_T^{(1)} - \tilde{y}_T^{(2)} \tag{21}$$

Amazingly, $\hat{y}_T$ is an unbiased predictor of y_T, i.e.

$$E(\hat{y}_T) = \beta_{0,T} + \beta_{1,T} T. \tag{22}$$

4.4 Second-Order Exponential Smoothing IV

1. The choice of the initial values, $\tilde{y}_0^{(1)}$ and $\tilde{y}_0^{(2)}$.
   - An easy solution would be letting $\tilde{y}_0^{(1)} = \tilde{y}_0^{(2)} = y_1$.
   - A more rigorous solution would be fitting the least square estimates of the trend line, then using the trend line to fit the initial values.
2. The choice of λ will be discussed in chapter 4.6.1.

4.4 An example for Second-Order Exponential Smoothing I

We use an example of US CPI from Jan 1995 to Dec 2004 shows a clearly linear trend to illustrate the whole process.
We first apply the **first order** exponential smoother with $\lambda = 0.3$ to the data. Not surprisingly, the smoother consistently underestimates the actual values.

4.4 An example for Second-Order Exponential Smoothing II

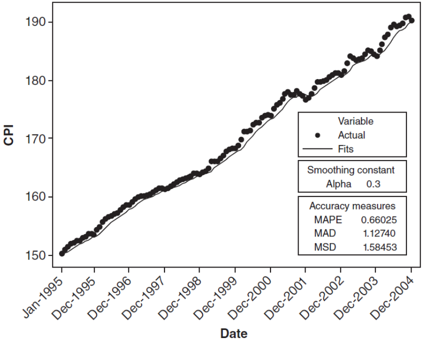


FIGURE 4.14 Single exponential smoothing of the US Consumer Price Index (with $\tilde{y}_0 = y_1$).

4.4 An example for Second-Order Exponential Smoothing III

We then use the **second order** exponential smoother with $\lambda = 0.3$ to obtain $\tilde{y}_T^{(2)}$.

TABLE 4.3 Second-Order Exponential Smoothing of the US Consumer Price Index (with $\lambda = 0.3, \tilde{y}_0^{(1)} = y_1$, and $\tilde{y}_0^{(2)} = \tilde{y}_0^{(1)}$)

Date	y_t	$\tilde{y}_T^{(1)}$	$\tilde{y}_T^{(2)}$	$\hat{y}_T = 2\tilde{y}_T^{(1)} - \tilde{y}_T^{(2)}$
Jan-1995	150.3	150.300	150.300	150.300
Feb-1995	150.9	150.480	150.354	150.606
Mar-1995	151.4	150.756	150.475	151.037
Apr-1995	151.9	151.099	150.662	151.536
May-1995	152.2	151.429	150.892	151.967
Nov-2004	191.0	190.041	188.976	191.106
Dec-2004	190.3	190.119	189.319	190.919

The last column shows the second order exponential smoothing $\hat{y}_T = 2\tilde{y}_T^{(1)} - \tilde{y}_T^{(2)}$.

4.4 An example for Second-Order Exponential Smoothing IV

$$
\begin{aligned}
\tilde{y}_0^{(1)} &= y_1 = 150.3 && (23)\\
\tilde{y}_1^{(1)} &= \lambda y_1 + (1-\lambda)\tilde{y}_0^{(1)} = 0.3 * 150.3 + 0.7 * 150.3 = 150.3 && (24)\\
\tilde{y}_2^{(1)} &= \lambda y_2 + (1-\lambda)\tilde{y}_1^{(1)} = 0.3 * 150.9 + 0.7 * 150.3 = 150.48 && (25)\\
\tilde{y}_3^{(1)} &= 0.3 * 151.4 + 0.7 * 150.48 = 150.756 && (26)\\
&\vdots && (27)
\end{aligned}
$$

$$
\begin{aligned}
\tilde{y}_0^{(2)} &= \tilde{y}_0^{(1)} = y_1 = 150.3 && (28)\\
\tilde{y}_1^{(2)} &= \lambda \tilde{y}_1^{(1)} + (1-\lambda)\tilde{y}_0^{(2)} = 0.3 * 150.3 + 0.7 * 150.3 = 150.3 && (29)\\
\tilde{y}_2^{(2)} &= \lambda \tilde{y}_2^{(1)} + (1-\lambda)\tilde{y}_1^{(2)} = 0.3 * 150.48 + 0.7 * 150.3 = 150.354 && (30)
\end{aligned}
$$

4.4 An example for Second-Order Exponential Smoothing V

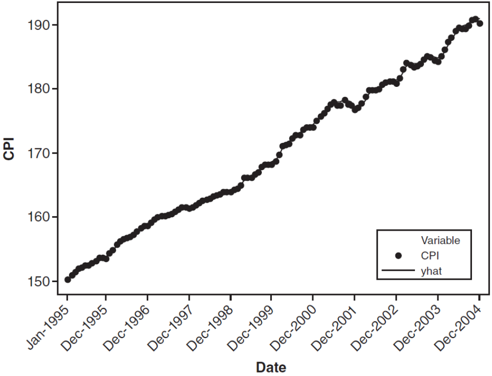


FIGURE 4.15 Second-order exponential smoothing of the US Consumer Price Index (with $\lambda = 0.3$, $\tilde{y}_0^{(1)} = y_1$, and $\tilde{y}_0^{(2)} = \tilde{y}_1^{(1)}$).

The new smoother not only captures the trend but also shows unbiasedness as shown in the picture.

Holt's Method(1957) I

Holt's method divides the time series data into two components: the level, L_t, and the trend T_t, i.e. $\hat{y}_t = L_t + T_t$. These two components can be calculated from

$$L_t = \alpha y_t + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{31}$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta) T_{t-1} \tag{32}$$

We could use the Holt-Winters function from the stats package to obtain the second-order exponential smoothing. This method is also called **double exponential smoothing**.

fit1 <- HoltWinters(cpi.data[,2],alpha=0.3,beta=0.3, gamma=FALSE).

The beta corresponds to the second-order smoothing (or the trend term) and gamma is for the seasonal effect.

Holt's Method(1957) II

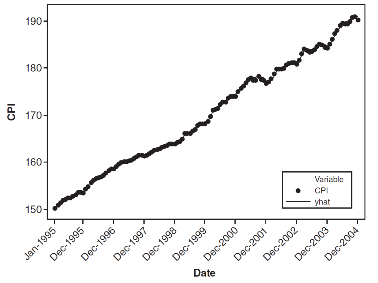


FIGURE 4.15 Second-order exponential smoothing of the US Consumer Price Index (with $\lambda = 0.3$, $\tilde{y}_0^{(1)} = y_1$, and $\tilde{y}_0^{(2)} = \tilde{y}_1^{(1)}$).

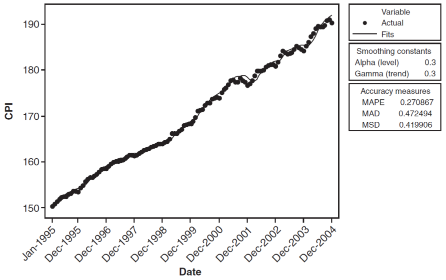


FIGURE 4.16 The double exponential smoothing of the US Consumer Price Index (with $\alpha = 0.3$ and $\gamma = 0.3$).

4.5 Higher-Order Exponential Smoothing

We have seen that the first-order exponential smoothing is used to estimate the constant process model, and the second order exponential smoothing is used to estimate the linear trend process model.

It can be shown that the model in the form of a nth-degree polynomial can be estimated by (n+1)-order exponential smoothers which are defined inductively:

$$\tilde{y}_T^{(n)} = \lambda \tilde{y}_T^{(n-1)} + (1-\lambda)\tilde{y}_{T-1}^{(n)}$$

But when the degree ≥ 2, the computations for getting the estimates becomes quite complicated. Other methods will be considered in later chapters.

4.6.1 Constant Process:model I

Now let us talk about forecast by using the exponential smoothers of first and second order.

For the constant process, our forecast for the future observation is simply equal to the current value of the smoother:

$$\hat{y}_{T+\tau}(T) = \tilde{y}_T \tag{33}$$

This shows that standing at any time period T, the forecast for all future values are the same. But we will keep updating our forecast. For example, if data at T+1 become available, the forecast becomes

$$\tilde{y}_{T+1} = \lambda y_{T+1} + (1-\lambda)\tilde{y}_T \tag{34}$$

or

$$\hat{y}_{T+1+\tau}(T+1) = \lambda y_{T+1} + (1-\lambda)\hat{y}_{T+\tau}(T). \tag{35}$$

4.6.1 Constant Process:model II

We can rewrite (35) for $\tau = 1$, as

$$\begin{aligned} \hat{y}_{T+2}(T+1) &= \lambda y_{T+1} + (1-\lambda)\hat{y}_{T+1}(T) && (36) \\ &= \hat{y}_{T+1}(T) + \lambda(y_{T+1} - \hat{y}_{T+1}(T)) && (37) \\ &= \hat{y}_{T+1}(T) + \lambda e_{T+1}(1) && (38) \end{aligned}$$

where $e_{T+1}(1) = y_{T+1} - \hat{y}_{T+1}(T)$ is called the one-step-ahead forecast or prediction error.

The interpretation of (38) makes it easier to understand the forecasting process using exponential smoothing: our forecast for the next observation is simply our previous forecast for the current observation plus a fraction of the forecast error we made in forecasting the current observation.

4.6.1 Constant Process:model III

The fraction in this summation of (38) is determined by λ. Hence how fast our forecast will react to the forecast error depends on the discount factor. A large discount factor will lead to fast reaction to the forecast error but it may also make our forecast react fast to random fluctuations. This once again brings up the issue of the choice of the discount factor.

4.6.1 Constant Process: Choice of λ

To select the value for λ, we recall the sum of squared one-step-ahead forecast errors and its standard deviation are given below:

$$SS_E(\lambda) = \sum_{t=1}^{T} e_t^2(1), \tag{39}$$

$$\hat{\sigma}_e = \sqrt{\frac{1}{T} SS_E} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} e_t^2(1)}. \tag{40}$$

For a given historic data, we can in general calculate SS_E values for various values of λ, and pick the value of λ that gives the smallest SS_E.

4.6.1 Constant Process: Prediction Intervals I

The $100(1-\frac{\alpha}{2})$ percent prediction Intervals for any lead time τ is given by

$$\tilde{y}_T \pm Z_{\alpha/2}\hat{\sigma}_e \tag{41}$$

where $\tilde{y}_T$ is the first-order exponential smoother, $Z_{\frac{\alpha}{2}}$ is the $100(1-\frac{\alpha}{2})$ percentile of the standard normal distribution, and $\hat{\sigma}_e = \sqrt{\frac{1}{T}\sum_{t=1}^{T} e_t^2(1)}$ is the estimate of the standard deviation of the forecast errors.

The prediction interval is constant for all lead times.

Issue:
This can be quite unrealistic. As it will be more likely that the process goes through some changes as time goes on, we would correspondingly expect to be less and less "sure" about our predictions for large lead times (or large τ values). Hence we would anticipate prediction intervals that are getting wider and wider for increasing lead times.

4.6.1 Constant Process: Example I

Example 4.4

We are interested in the average speed on a specific stretch of a highway during nonrush hours. For the past year and a half (78 weeks), we have available weekly averages of the average speed in miles/hour between 10 AM and 3 PM. The data are given in the textbook.

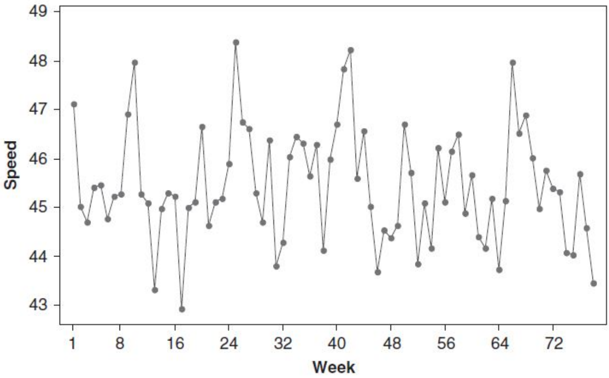


FIGURE 4.18 The weekly average speed during nonrush hours.

Figure 4.18 shows that the time series data follow a constant process.

4.6.1 Constant Process: Example III

The sum of the squared one-step-ahead prediction errors for various values is given in Table 4.6.

TABLE 4.6 SS_E **for Different** λ **Values for the Average Speed Data**

λ		0.1		0.2		0.3		0.4		0.5		0.9	
Week	Speed	Forecast	$e(t)$	Forecast	$e(t)$	Forecast	$e(t)$	Forecast	$e(t)$	Forecast	$e(t)$	Forecast	$e(t)$
1	47.12	47.12	0.00	47.12	0.00	47.12	0.00	47.12	0.00	47.12	0.00	47.12	0.00
2	45.01	47.12	−2.11	47.12	−2.11	47.12	−2.11	47.12	−2.11	47.12	−2.U	47.12	−2.11
3	44.69	46.91	−2.22	46.70	−2.01	46.49	−1.80	46.28	−1.59	46.07	−1.38	45.22	−0.53
4	45.41	46.69	−1.28	46.30	−0.89	45.95	−0.54	45.64	−0.23	45.38	0.03	44.74	0.67
5	45.45	46.56	−1.11	46.12	−0.67	45.79	−0.34	45.55	−0.10	45.39	0.06	45.34	0.11
6	44.77	46.45	−1.68	45.99	−1.22	45.69	−0.92	45.51	−0.74	45.42	−0.65	45.44	−0.67
7	45.24	46.28	−1.04	45.74	−0.50	45.41	−0.17	45.21	0.03	45.10	0.14	44.84	0.40
8	45.27	46.18	−0.91	45.64	−0.37	45.36	−0.09	45.22	0.05	45.17	0.10	45.20	0.07
9	46.93	46.09	0.84	45.57	1.36	45.33	1.60	45.24	1.69	45.22	1.71	45.26	1.67
10	47.97	46.17	1.80	45.84	2.13	45.81	2.16	45.92	2.05	46.07	1.90	46.76	1.21
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
75	44.02	45.42	−1.40	45.30	−1.28	45.12	−1.10	44.93	−0.91	44.75	−0.73	44.20	−0.18
76	45.69	45.28	0.41	45.05	0.64	44.79	0.90	44.56	1.13	44.39	1.30	44.04	1.65
77	44.59	45.32	−0.73	45.18	−0.59	45.06	−0.47	45.01	−0.42	45.04	−0.45	45.52	−0.93
78	43.45	45.25	−1.80	45.06	−1.61	44.92	−1.47	44.84	−1.39	44.81	−1.36	44.68	−1.23
SS_E			124.14		118.88		117.27		116.69		116.95		128.98

4.6.1 Constant Process: Example IV

For $\lambda = 0.1$, Let's calculate $\tilde{y}_t, t = 0, 1, \cdots, T$.

$$\tilde{y}_0 = y_1 = 47.12 \tag{42}$$
$$\tilde{y}_1 = \lambda y_1 + (1-\lambda)\tilde{y}_0 = 0.1 * 47.12 + 0.9 * 47.12 = 47.12 \tag{43}$$
$$\tilde{y}_2 = \lambda y_2 + (1-\lambda)\tilde{y}_1 = 0.1 * 45.01 + 0.9 * 47.12 = 46.91 \tag{44}$$
$$\tilde{y}_3 = \lambda y_3 + (1-\lambda)\tilde{y}_2 = 0.1 * 44.69 + 0.9 * 46.91 = 46.69 \tag{45}$$
$$\vdots \tag{46}$$

For $\lambda = 0.1$, Let's calculate $e_t(1) = y_t - \hat{y}_t = y_t - \tilde{y}_{t-1}, t = 1, \cdots, T$.

$$e_1(1) = y_1 - \hat{y}_1 = y_1 - \tilde{y}_0 = 47.12 - 47.12 = 0 \tag{47}$$
$$e_2(1) = y_2 - \hat{y}_2 = y_2 - \tilde{y}_1 = 45.01 - 47.12 = -2.11 \tag{48}$$
$$e_3(1) = y_3 - \hat{y}_3 = y_3 - \tilde{y}_2 = 44.69 - 46.91 = -2.22 \tag{49}$$
$$\vdots \tag{50}$$

4.6.1 Constant Process: Example V

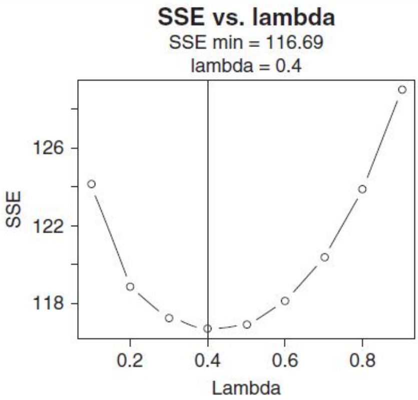


FIGURE 4.19 Plot of SS_E for various λ values for average speed data.

Figure 4.19 shows that the minimum SS_E is obtained for $\lambda = 0.4$.

4.6.1 Constant Process: Example VI

Let us assume that we are also asked to make forecasts for the next 12 weeks at week 78, where $\tilde{y}_{78} = 44.84$.

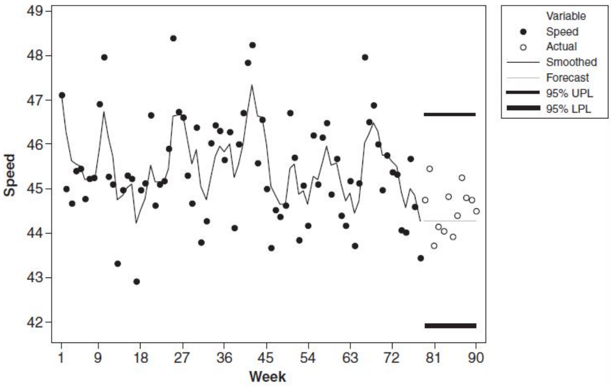


FIGURE 4.20 Forecasts for the weekly average speed data for weeks 79–90.

4.6.1 Constant Process: Example VII

Figure 4.20 shows the smoothed values for the first 78 weeks together with the forecasts for weeks 79 - 90 with prediction intervals. It also shows the actual weekly speed during that period. Note that since the constant process is assumed, the forecasts for the next 12 weeks are the same. Similarly, the 95% prediction intervals are constant for that period.

$$\tilde{y}_{78} \pm z_{0.05/2}\hat{\sigma}_e = (44.84 \pm 1.96(\sqrt{116.69/78}) = (42.44, 47.24) \tag{51}$$

4.6.1 Constant Process: Example VIII

R code for calculating the optimal λ for Example 4.4. The average speed data are in the second column of the array called speed.data in which the first column is the index for week.

- First function:

```
firstsmooth<-function(y,lambda,start=y[1]){
      ytilde<-y
      ytilde[1]<-lambda*y[1]+(1-lambda)*start
      for (i in 2:length(y)){
            ytilde[i]<-lambda*y[i]+(1-lambda)*ytilde[i-1]
      }
ytilde
}
```

4.6.1 Constant Process: Example IX

- Second function:

```
measacc.fs<- function(y,lambda){
            out<- firstsmooth(y,lambda)
            T<-length(y)
            #Smoothed version of the original is the c
              ahead prediction
            #Hence the predictions (forecasts) are giv
            pred<-c(y[1],out[1:(T-1)])
            prederr<- y-pred
            SSE<-sum(prederr^2)
            MAPE<-100*sum(abs(prederr/y))/T
            MAD<-sum(abs(prederr))/T
            MSD<-sum(prederr^2)/T
            ret1<-c(SSE,MAPE,MAD,MSD)
            names(ret1)<-c("SSE","MAPE","MAD","MSD")
            return(ret1)
}

measacc.fs(dji.data[,2],0.4)
         SSE          MAPE            MAD            MSD
1.665968e+07 3.461342e+00 3.356325e+02 1.959962e+05
```

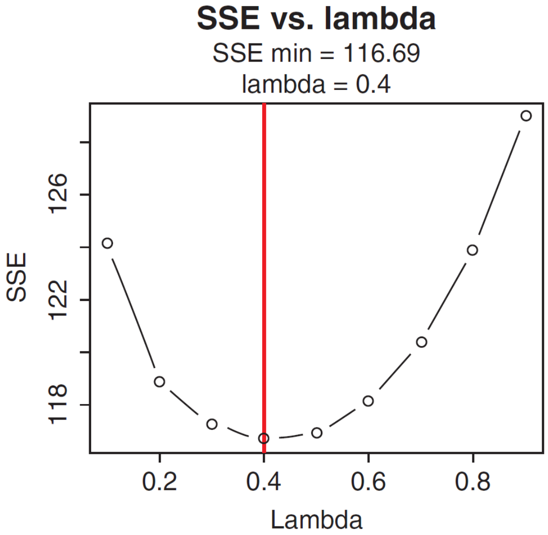
4.6.1 Constant Process: Example X

- Third function

```
lambda.vec<-seq(0.1, 0.9, 0.1)
sse.speed<-function(sc){measacc.fs(speed.data[1:78,2],sc)[1]}
sse.vec<-sapply(lambda.vec, sse.speed)
opt.lambda<-lambda.vec[sse.vec == min(sse.vec)]
plot(lambda.vec, sse.vec, type="b", main = "SSE vs. lambda\n",
  xlab='lambda\n',ylab='SSE')
abline(v=opt.lambda, col = 'red')
mtext(text = paste("SSE min = ", round(min(sse.vec),2), "\n lambda
  = ", opt.lambda))
```

To show the value of optimal λ, you need to type **opt.lambda** at the end of the code.

4.6.1 Constant Process: Example XI



4.6.2 Linear Trend Process I

The τ-step-ahead forecast for the linear trend model is given by

$$\hat{y}_{T+\tau}(T) = \hat{\beta}_{0,T} + \hat{\beta}_{1,T}(T+\tau) \tag{52}$$
$$= \hat{\beta}_{0,T} + \hat{\beta}_{1,T}T + \hat{\beta}_{1,T}\tau \tag{53}$$
$$= \hat{y}_T + \hat{\beta}_{1,T}\tau. \tag{54}$$

where $\hat{\beta}_{1,T}$ is calculated by equation (17).
By using the exponential smoothers, we can rewrite the (54) as

$$\hat{y}_{T+\tau}(T) = \left(2\tilde{y}_T^{(1)} - \tilde{y}_T^{(2)}\right) + \frac{\lambda}{1-\lambda}\left(\tilde{y}_T^{(1)} - \tilde{y}_T^{(2)}\right)\tau \tag{55}$$
$$= \left(2 + \frac{\lambda}{1-\lambda}\tau\right)\tilde{y}_T^{(1)} - \left(1 + \frac{\lambda}{1-\lambda}\tau\right)\tilde{y}_T^{(2)} \tag{56}$$

The predictions for the trend model depend on the lead time, as opposed to the constant model, will be different for different lead times.
(Go to Fig 4.23)

4.6.2 Linear Trend Process II

The $100(1-\frac{\alpha}{2})$ percent prediction Intervals for any lead time τ is

$$\left[\left(2+\frac{\lambda}{1-\lambda}\tau\right)\tilde{y}_T^{(1)}-\left(1+\frac{\lambda}{1-\lambda}\tau\right)\tilde{y}_T^{(2)}\right]\pm Z_{\frac{\alpha}{2}}\frac{c_\tau}{c_1}\hat{\sigma}_e, \qquad (57)$$

where

$$c_i^2=1+\frac{\lambda}{(2-\lambda)^3}[(10-14\lambda+5\lambda^2)+2i\lambda(4-3\lambda)+2i^2\lambda^2].$$

Example 4.5

Reconsider the CPI data in Example 4.2. Assume that we are currently in December 2003 and would like to make predictions of the CPI for the following year. Although the data from January 1995 to December 2003 clearly exhibit a linear trend, **We "pretend" that this is a constant process and use first order smoother to select λ.** We will then calculate the best value that minimizes the sum of the squared one-step-ahead prediction errors. The predictions and prediction errors for various values are given in Table 4.7.

TABLE 4.7 The Predictions and Prediction Errors for Various λ Values for CPI Data

		$\lambda = 0.1$		$\lambda = 0.2$		$\lambda = 0.3$			$\lambda = 0.9$		$\lambda = 0.99$	
Month-Year	CPI	Prediction	Error	Prediction	Error	Prediction	Error		Prediction	Error	Prediction	Error
Jan-1995	150.3	150.30	0.00	150.30	0.00	150.30	0.00	…	150.30	0.00	150.30	0.00
Feb-1995	150.9	150.30	0.60	150.30	0.60	150.30	0.60	…	150.30	0.60	150.30	0.60
Mar-1995	151.4	150.36	1.04	150.42	0.98	150.48	0.92	…	150.84	0.56	150.89	0.51
Apr-1995	151.9	150.46	1.44	150.62	1.28	150.76	1.14	…	151.34	0.56	151.39	0.51
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮		⋮	⋮	⋮	⋮
Nov-2003	184.5	182.29	2.21	183.92	0.58	184.45	0.05	…	185.01	−0.51	185.00	−0.50
Dec-2003	184.3	182.51	1.79	184.03	0.27	184.46	−0.16	…	184.55	−0.25	184.51	−0.21
SS_E			1061.50		309.14		153.71			31.90		28.62

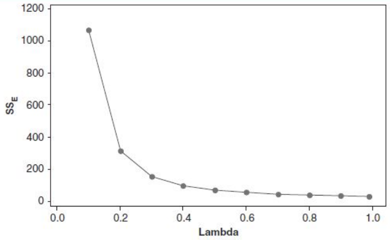


FIGURE 4.21 Scatter plot of the sum of the squared one-step-ahead prediction errors versus λ.

As λ gets closer to 1, the SS_E keeps getting smaller. This indicates a higher order smoothing is needed.

Then in theory the λ may be selected by using second order smoother. We skip this selection (Minimize $\sum_{t=1}^{T} e_t(1)^2 = \sum_{t=1}^{T}(y_t - \hat{y}_t(t-1))^2$ for a 2nd order smoother.) and let $\lambda = 0.3$.

4.6.2 Linear Trend Process: Example III

Assuming that we were at December 2003, and wanted to forecast year 2004, then we can either

1. In Dec 2003, make forecast for the entire 2004 year, that is, 1-step-ahead, 2-step-ahead, . . . , 12-step-ahead forecasts. For that, we can use (54) or (56).The forecasts given in Fig 4.23.

   Note that the forecasts further in the future (for the later part of 2004) are quite a bit off. To remedy this we may instead use the following strategy.

2. In Dec 2003, make the one-step-ahead forecast for next period(Jan 2004). When the data for Jan 2004 becomes available, then make one-step-ahead forecast for Feb 2004, and so on. The forecasts given in Fig 2.24.

4.6.2 Linear Trend Process: Example IV

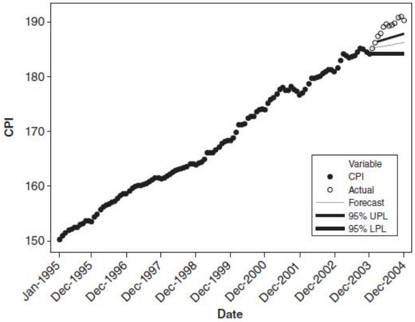


FIGURE 4.23 The 1- to 12-step-ahead forecasts of the CPI data for 2004.

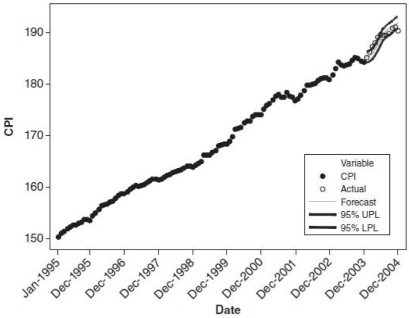


FIGURE 4.24 The one-step-ahead forecasts of the CPI data for 2004.

4.6.3 Estimation of σ_e^2 I

1. The estimated variance of forecast errors, $\hat{\sigma}_e^2$ is calculated as the mean of all squared one-step-ahead forecast errors.

$$\hat{\sigma}_e^2 = \frac{1}{T}\sum_{t=1}^{T} e_t^2(1)$$

Notice this is the estimate made at time period T, so we can also denote it by $\hat{\sigma}^2_{e,T}$.

When data at T+1 is collected, the estimate can be updated as

$$\hat{\sigma}^2_{e,T+1} = \frac{1}{T+1}(T\hat{\sigma}^2_{e,T} + e^2_{T+1}(1))$$

Similarly, when estimating the variance of the τ-step-ahead forecast errors, we just find the mean of the all squared τ-step-ahead forecast errors.

4.6.3 Estimation of σ_e^2 II

2. Define the **mean absolute deviation** Δ as

$$\Delta = E\left(|e - E(e)|\right) \tag{58}$$

and assuming the model is correct, calculate its estimate by

$$\hat{\Delta}_T = \delta|e_T(1)| + (1-\delta)\hat{\Delta}_{T-1}. \tag{59}$$

Then the estimate of the σ_e^2 is given by

$$\hat{\sigma}_e^2 = 1.25\hat{\Delta}_T \tag{60}$$

For further details, see Montgomery et al. (1990).

4.6.4 Updating the Discount Factor λ I

For some data with changing patterns, it is difficult to use exponential smoother with fixed discount factor to follow the changes. Thus we can modify the factor if necessary. There are many methods of updating the factor, one of which was originally described by Trigg and Leach.

As an example, we will consider the first-order exponential smoother and modify it as

$$\tilde{y}_T = \lambda_T y_T + (1 - \lambda_T)\tilde{y}_{T-1} \tag{61}$$

The discount factor λ_T depends on time T.

4.6.4 Updating the Discount Factor λ II

We now introduce some quantities needed for this process.

$$\hat{\Delta}_T = \delta|e_T(1)| + (1-\delta)\hat{\Delta}_{T-1}, \tag{62}$$

where δ is a smoothing parameter and is also used in defining the smoothed error Q_T:

$$Q_T = \delta e_T(1) + (1-\delta)Q_{T-1}. \tag{63}$$

Trigg and Leach (1967) defined the discount factor

$$\lambda_T = |\frac{Q_T}{\hat{\Delta}_T}| \tag{64}$$

as the adaptive discount factor for time T.
The forecasting system performs well then λ_T is closed to 0;
The forecasting system model fails, then λ_T is closed to 1.

4.6.4 An Example I

The Dow Jones data shows no constant process or linear trend process pattern, hence we can use the adaptive discount factor. A comparison is made between this method and a simple exponential smoother with fixed factor $\lambda = 0.3$.

4.6.4 An Example II

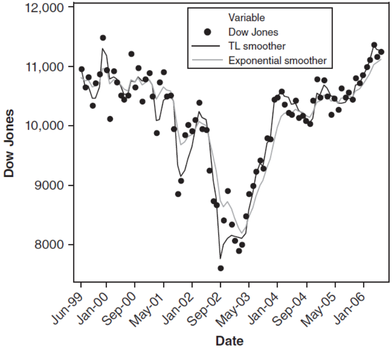


FIGURE 4.25 Time series plot of the Dow Jones Index from June 1999 to June 2006, the simple exponential smoother with $\lambda = 0.3$, and the Trigg–Leach (TL) smoother with $\delta = 0.3$.

4.6.4 An Example III

Date	Dow Jones	Smoothed	λ	Error	Qt	Dt
Jun-99	10,970.80					
Jul-99	10,655.20					
Aug-99	10,829.30					
Sep-99	10,337					
Oct-99	10,729.90					
⋮	⋮	⋮	⋮	⋮	⋮	⋮

$$\hat{y}_{T+1}(T) = \tilde{y}_T = \lambda_T y_T + (1-\lambda_T)\tilde{y}_{T-1}. \tag{65}$$

$$e_T(1) = y_T - \hat{y}_T(T-1). \tag{66}$$

$$\hat{\Delta}_T = \delta|e_T(1)| + (1-\delta)\hat{\Delta}_{T-1}. \tag{67}$$

$$Q_T = \delta e_T(1) + (1-\delta)Q_{T-1}. \tag{68}$$

$$\lambda_T = |\frac{Q_T}{\hat{\Delta}_T}| \tag{69}$$

$$\delta = 0.3, Q_0 = 0, \Delta_0 = 0, or, D_0 = 0. \tag{70}$$

4.6.4 An Example IV

$y_1 = 10970.8, y_2 = 10655.20, y_3 = 10829.30, y_4 = 10337, y_5 = 10729.90$

1. T=0, (Jun-99)

$$\hat{y}_1(0) = \tilde{y}_0 = y_1 = 10970.8 \tag{71}$$

$$e_1(1) = y_1 - \hat{y}_1(0) = 10970.8 - 10970.8 = 0 \tag{72}$$

$$\hat{\Delta}_1 = \delta|e_1(1)| + (1-\delta)\hat{\Delta}_0 = 0.3 * 0 + 0.7 * 0 = 0 \tag{73}$$

$$Q_1 = \delta e_1(1) + (1-\delta)Q_0 = 0.3 * 0 + 0.7 * 0 = 0. \tag{74}$$

$$\lambda_1 = 1 \tag{75}$$

4.6.4 An Example V

2. T=1,(Jul-99)

$$\begin{aligned}
\hat{y}_2(1) &= \tilde{y}_1 = \lambda_1 y_1 + (1-\lambda_1)\tilde{y}_0 \\
&= 1 * 10970.8 + 0 * 10970.8 = 10970.8 \qquad (76)
\end{aligned}$$

$$e_2(1) = y_2 - \hat{y}_2(1) = y_2 - \tilde{y}_1 = 10655.2 - 10970.8 = -315.6 \qquad (77)$$

$$Q_2 = \delta e_2(1) + (1-\delta)Q_1 = 0.3 * (-315.6) + 0.7 * 0 = -94.68 \qquad (78)$$

$$\hat{\Delta}_2 = \delta|e_2(1)| + (1-\delta)\hat{\Delta}_1 = 0.3 * (315.6) + 0.7 * 0 = 94.68 \qquad (79)$$

$$\lambda_2 = |\frac{Q_2}{\hat{\Delta}_2}| = |\frac{-94.68}{94.68}| = 1 \qquad (80)$$

3. T=2, (Aug-99)

$$\hat{y}_3(2) = \tilde{y}_2 = \lambda_2 y_2 + (1-\lambda_2)\tilde{y}_1 = 1 * 10655.20 = 10655.2 \tag{81}$$

$$e_3(1) = y_3 - \hat{y}_3(2) = y_3 - \tilde{y}_2 = 10829.3 - 10655.2 = 174.1 \tag{82}$$

$$Q_3 = \delta e_3(1) + (1-\delta)Q_2 = 0.3 * 174.1 + 0.7 * (-94.68) = -14.046 \tag{83}$$

$$\hat{\Delta}_3 = \delta|e_3(1)| + (1-\delta)\hat{\Delta}_2 = 0.3 * 174.1 + 0.7 * 94.68 = 118.506 \tag{84}$$

$$\lambda_3 = |\frac{Q_3}{\hat{\Delta}_3}| = |\frac{-14.046}{118.506}| = 0.11853 \tag{85}$$

4. T=3, (Sep-99)

$$\begin{aligned}
\hat{y}_4(3) &= \tilde{y}_3 = \lambda_3 y_3 + (1-\lambda_3)\tilde{y}_2 \\
&\quad = 0.11853 * 10829.3 + (1 - 0.11853) * 10655.2 = 10675.835 \\
e_4(1) &= y_4 - \hat{y}_4(3) = y_4 - \tilde{y}_3 = 10337 - 10675.84 = -338.84 \\
Q_4 &= \delta e_4(1) + (1-\delta)Q_3 = 0.3 * (-338.84) + 0.7 * (-14.046) \\
&= -111.4842 \\
\hat{\Delta}_4 &= \delta|e_4(1)| + (1-\delta)\hat{\Delta}_3 = 0.3 * (338.84) + 0.7 * (118.506) \\
&= 184.6062 \\
\lambda_4 &= |\frac{Q_4}{\hat{\Delta}_4}| = |\frac{-111.4842}{184.6062}| = 0.6039028
\end{aligned}$$

4.6.4 An Example VIII

5. T=4, (Oct-99),

$$\begin{aligned}\hat{y}_5(4) = \tilde{y}_4 &= \lambda_4 y_4 + (1-\lambda_4)\tilde{y}_3 \\ &= 0.6039028 * 10337 + (1-0.6039028) * 10675.835 \\ &= 10471.213 \end{aligned} \tag{86}$$

...

The calculations for this procedure are given below.

Date	Dow Jones	Smoothed	λ	Error	Q_t	D_t
Jun-99	10,970.8	10,970.8	1		0	0
Jul-99	10,655.2	10,655.2	1	−315.6	−94.68	94.68
Aug-99	10,829.3	10,675.835	0.11853	174.1	−14.046	118.506
Sep-99	10,337	10,471.213	0.6039	−338.835	−111.483	184.605
Oct-99	10,729.9	10,471.753	0.00209	258.687	−0.43178	206.83
⋮	⋮	⋮	⋮	⋮	⋮	⋮
May-06	11,168.3	11,283.962	0.36695	−182.705	68.0123	185.346
Jun-06	11,247.9	11,274.523	0.26174	−36.0619	36.79	140.561

4.7 Exponential Smoothing for Seasonal Data I

The exponential smoothing techniques can be adjusted to fit seasonal time series, which was proposed by **Holt and Winters**, and is called Winters' Method. We will discuss this method for

- Additive Seasonal Model,
- Multiplicative Seasonal Model.

4.7 Exponential Smoothing for Seasonal Data II

An additive Seasonal Model has the form:

$$y_t = L_t + S_t + \epsilon_t, \tag{87}$$

where L_t is the linear trend component and therefore $L_t = \beta_0 + \beta_1 t$; S_t represents the seasonal adjustment with $S_t = S_{t+s} = S_{t+2s} = \cdots$ for $t = 1, \cdots, s-1$, where s is the length of the season (period) of the cycles; the ϵ_t are assumed to be uncorrelated with mean 0 and constant variance σ_ϵ^2. Sometimes the level is called the permanent component. One usual restriction on this model is that the seasonal adjustments add to zero during one season,

$$\sum_{t=1}^{s} S_t = 0. \tag{88}$$

4.7 Exponential Smoothing for Seasonal Data III

In the model given in (87) for forecasting the future observations, we will employ first-order exponential smoothers with different discount factors. The Winters' method is an iteration process that updates the estimates for L_T, $\beta_{1,T}$ and S_T. The updating process once the current observation y_T is obtained goes as follows:

1. step 1

$$\hat{L}_T = \lambda_1(y_T - \hat{S}_{T-s}) + (1-\lambda_1)(\hat{L}_{T-1} + \hat{\beta}_{1,T-1}),$$

   where $0 < \lambda_1 < 1$.

2. step 2

$$\hat{\beta}_{1,T} = \lambda_2(\hat{L}_T - \hat{L}_{T-1}) + (1-\lambda_2)\hat{\beta}_{1,T-1},$$

   where $0 < \lambda_2 < 1$.

4.7 Exponential Smoothing for Seasonal Data IV

3. step 3
$$\hat{S}_T = \lambda_3(y_T - \hat{L}_T) + (1 - \lambda_3)\hat{S}_{T-s},$$
where $0 < \lambda_3 < 1$.
4. step 4
$$\hat{y}_{T+\tau} = \hat{L}_T + \hat{\beta}_{1,T}\tau + \hat{S}_T(\tau - s),$$

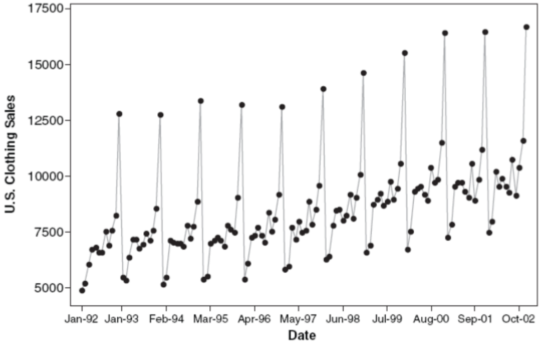
where s is the length of the season.

The choices of their initial values are given on page 279 in the textbook.

4.7 Exponential Smoothing for Seasonal Data V

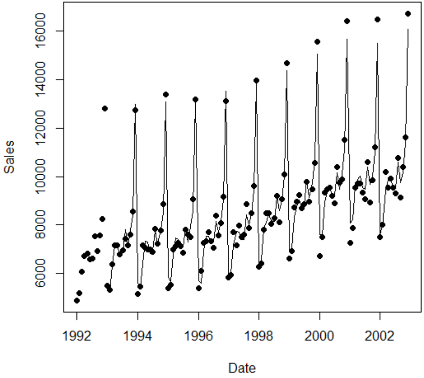
Example 4.7

The data set contains the clothing sales from Jan 1992 to Dec 2003. A plot of the data is as follows:

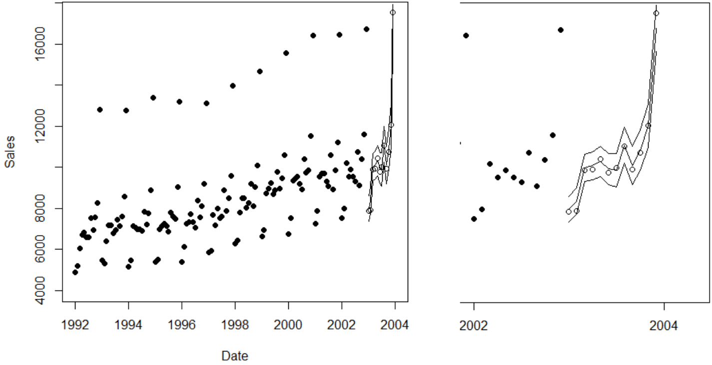


We see seasonal pattern with the same amplitude, so we use the additive model.

clo.hw1<- HoltWinters(clo.data[,2],alpha=0.2,beta=0.2, gamma=0.2, seasonal="additive"). The smoothed data captures the seasonality decently.



We assume that we were standing at 2002 and wanting to predict 2003. The plot below shows the prediction values with the actual data and the 95% predict limits.



4.7 Exponential Smoothing for Seasonal Data VIII

A multiplicative seasonal model has the form:

$$y_t = L_t S_t + \epsilon_t. \tag{89}$$

Again, we have $L_t = \beta_0 + \beta_1 t$ being the linear component, and S_t being the seasonal component with length s. The restriction for the seasonal adjustments is

$$\sum_{t=1}^{s} S_t = s. \tag{90}$$

4.7 Exponential Smoothing for Seasonal Data IX

The Winters' method for this case also updates three parameters and the smoothed value as in the additive model.
The updating process goes as follows:

1. step 1
$$\hat{L}_T = \lambda_1 \frac{y_T}{\hat{S}_{T-s}} + (1-\lambda_1)(\hat{L}_{T-1} + \hat{\beta}_{1,T-1}),$$
2. step 2
$$\hat{\beta}_{1,T} = \lambda_2(\hat{L}_T - \hat{L}_{T-1}) + (1-\lambda_2)\hat{\beta}_{1,T-1},$$
3. step 3
$$\hat{S}_T = \lambda_3 \frac{y_T}{\hat{L}_T} + (1-\lambda_3)\hat{S}_{T-s},$$
4. step 4
$$\hat{y}_{T+1} = (\hat{L}_T + \hat{\beta}_{1,T})\hat{S}_T(1-s),$$

where λ_1,λ_2, and λ_3 are parameters that are between 0 and 1.

The choices of their initial values are given on textbook page 285.

4.7 Exponential Smoothing for Seasonal Data XI

Example 4.8 Consider the data set of liquor store. A plot of the data is as follows:

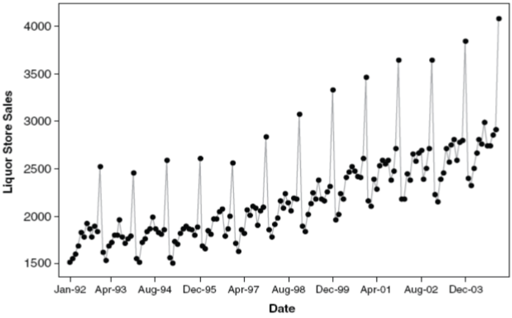
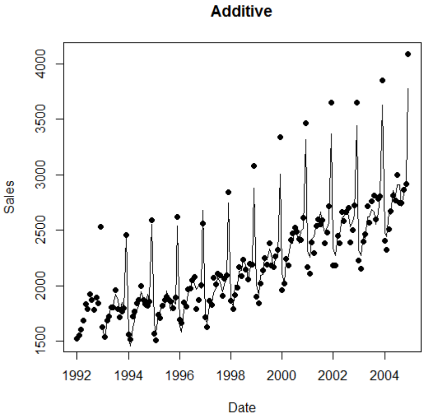


FIGURE 4.29 Time series plot of liquor store sales data from January 1992 to December 2004.

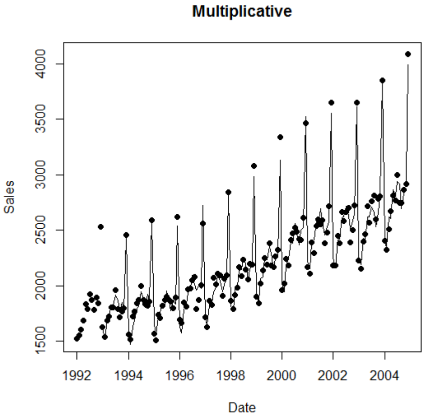
The amplitude of the periodic behavior gets larger, hence the multiplicative model is appropriate.

This is the plot of the smoothed data by using additive model. We see it does not capture the increasing magnitude very well in the end.



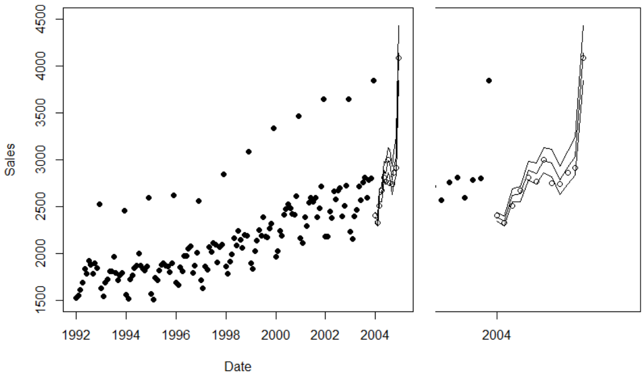
4.7 Exponential Smoothing for Seasonal Data XIII

On the other hand, we see the multiplicative model is doing a better job, so it should be preferred in this case.



4.7 Exponential Smoothing for Seasonal Data XIV

As for forecasting, we assume that we were in December 2003 and wanted to forecast 2004.



4.9 Exponential Smoothers and ARIMA Models I

The first-order exponential smoothers is particularly good in forecasting time series data with certain specific characteristics.

$$\tilde{y}_T = \lambda y_T + (1-\lambda)\tilde{y}_{T-1} \tag{91}$$

$$\hat{y}_{T+1}(T) = \tilde{y}_T \tag{92}$$

and the forecast error is defined as

$$e_T = y_T - \hat{y}_T(T-1) = y_T - \tilde{y}_{T-1} \tag{93}$$

Similarly, we have

$$e_{T-1} = y_{T-1} - \hat{y}_{T-1}(T-2) = y_{T-1} - \tilde{y}_{T-2} \tag{94}$$

Simplify $e_T - (1-\lambda)e_{T-1}$

$$
\begin{aligned}
& e_T - (1-\lambda)e_{T-1} \\
& = y_T - \tilde{y}_{T-1} - (1-\lambda)(y_{T-1} - \tilde{y}_{T-2}) \\
& = y_T - y_{T-1} - \tilde{y}_{T-1} + \lambda y_{T-1} + (1-\lambda)\tilde{y}_{T-2} \\
& = y_T - y_{T-1} - \tilde{y}_{T-1} + \tilde{y}_{T-1} \\
& = y_T - y_{T-1}
\end{aligned} \tag{95}
$$

Rewrite (95) as

$$
y_T - y_{T-1} = e_T - \theta e_{T-1} \tag{96}
$$

where $\theta = 1 - \lambda$.

4.9 Exponential Smoothers and ARIMA Models III

Recall from chapter 2 the **backshift** operator, **B**, defined as

$$\mathbf{B}y_t = y_{t-1} \tag{97}$$

Thus equation (96)

$$y_T - y_{T-1} = e_T - \theta e_{T-1}$$

becomes

$$y_T - \mathbf{B}y_T = e_T - \theta \mathbf{B} e_T \tag{98}$$

$$(1-\mathbf{B})y_T = (1-\theta\mathbf{B})e_T \tag{99}$$

The equation (99) is called the **integrated moving average** model denoted as **IMA(1,1)**, for the backshift operator is used only once on y_T and only once on the error e_T. It can be shown that if the process exhibits the dynamics defined in Eq.(99), the first-order exponential smoother provides minimum mean squared error (MMSE) forecasts.