Economic Forecasting
Lecture 9: Smoothing Methods

Richard G. Pierse

1 Introduction

Smoothing methods are rather different from the model-based methods that we have been looking at up to now in this module. With smoothing methods there is no search for a 'best-fitting' model and the forecasts produced will not generally be optimal forecasts (although in some circumstances they may be). Instead, smoothing methods are a simple way to put a smooth line through the data and project that line into the future. Smoothing methods are useful in situations where model-based methods cannot or should not be used. One such situation is when there are too few observations to estimate a model. For example, with only four observations, there are just too few degrees of freedom to be able to estimate any model, however simple. On the other hand, smoothing methods can still be used, even with such a small number of observations. Another situation in which smoothing methods are useful is where model-based forecasts would be too expensive, because of the size of the forecasting task or because of lack of time. Smoothing methods are quick and dirty and require little attention. They are an example of what are called 'automatic forecasts', which can be produced very cheaply and with little effort.

Although some academics dislike smoothing methods because of their quick and dirty nature, they do have a good track record in practice and are therefore worthy of consideration as a forecasting tool.

2 Moving average smoothing

One simple way of smoothing a time series is the moving average. For example, the *two-sided moving average* of a time series y_t is defined by

$$\overline{y}_t = \frac{1}{2m+1} \sum_{j=-m}^{m} y_{t-j} \tag{1}$$

where m is the smoothing parameter. The larger the value of m, the smoother the resulting series will be. The two-sided moving average takes a symmetric average of past and future values of y_t to smooth the current value. This means that it is not suitable for forecasting the series forwards since it requires future values to be known. There are also problems near the end-points of the series, y_1 and y$_T$ since the smoother can only be applied from the time observations y_{m+1} up to the observation y_{T-m}. Despite this, two-sided moving averages form an important part of seasonal adjustment algorithms such as the U.S. Census *X-12 ARIMA* package, as we have seen in a previous lecture.

A *one-sided moving average* of y_t is defined by

$$\overline{y}_t = \frac{1}{m+1}\sum_{j=0}^{m} y_{t-j} \tag{2}$$

where m is the smoothing parameter and, as before, the larger the value of m, the smoother the resulting series. Since the one-sided moving average only uses current and past values of y_t in order to smooth the current observation, the smoothed series can be calculated in the current period and can then be used to generate forecasts of future values of y_t.

A *weighted one-sided moving average* of y_t is defined by

$$\overline{y}_t = \sum_{j=0}^{m} w_j y_{t-j} \tag{3}$$

where w_0, w_1, $\ldots$, w_m are weights. Note that the simple one-sided moving average (2) is a special case of the weighted one-sided moving average where the weights are all equal with $w_j = 1/(m+1)$. More generally, the weighting parameter w_j allows flexibility in the way that the past is discounted. For example, often we want to use declining weights so as to discount the distant past more heavily than the recent past. Exponential smoothing methods use a particular form of declining weights.

3 Simple exponential smoothing

Simple exponential smoothing is a form of one-sided weighted moving average smoothing defined by

$$\overline{y}_t = \sum_{j=0}^{t-1} \alpha(1-\alpha)^j y_{t-j}$$

where α is the smoothing parameter with $0 < \alpha < 1$ and the smaller the value of α, the smoother the resulting series. It can be seen that the weights decline exponentially as

$$\overline{y}_t = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots + \alpha(1-\alpha)^{t-1} y_1.$$

For example, with $\alpha = 0.5$, the weights are $w_0 = 0.5$, $w_1 = 0.25$, $w_2 = 0.125$ etc. As the smoothing parameter α approaches 0, the smoothed series approaches a constant value of 0 and as α approaches 1, the smoothed series tends to the original unsmoothed series with $\overline{y}_t \to y_t$. In practice α will generally be chosen *a priori* although, if there are enough observations, it can be estimated by minimising the mean square error

$$\sum_{t=1}^{T} (y_t - \overline{y}_t)^2. \tag{4}$$

The exponential smoother can be generated recursively by initalising $\overline{y}_1$ from the initial observation

$$\overline{y}_1 = y_1 \tag{5}$$

and then using the recursion

$$\overline{y}_t = \alpha y_t + (1-\alpha)\overline{y}_{t-1} \tag{6}$$

for $t = 2, \ldots, T$. This makes exponential smoothing extremely easy to implement. Moreover, the smoothed series can then be used to forecast the series forwards using the formula

$$\widehat{y}_{t+h,t} = \overline{y}_t$$

which implies a flat forecast of the latest smoothed estimator.

Figure 1 illustrates two exponential smoothing models for the Canadian employment data series produced by *EViews* using values of the smoothing parameter α of 0.1 and 0.2 respectively. Note that *EViews* initialises the recursion (6) for the smoothed series using the sample mean

$$\overline{y}_1 = \overline{y} = \frac{1}{T} \sum_{t=1}^{T} y_t$$

instead of (5). Estimating the value of α by minimising the mean square error (4) gives the value 0.999 which suggests that the 'best' exponential smoothed estimator is essentially the original unsmoothed series itself with implied forecasts

$$\widehat{y}_{t+h,t} = 0.999 y_t$$

which simply projects the latest value. This is clearly not a very satisfactory forecast. Examining the smoothed series in Figure 1 shows that turning points

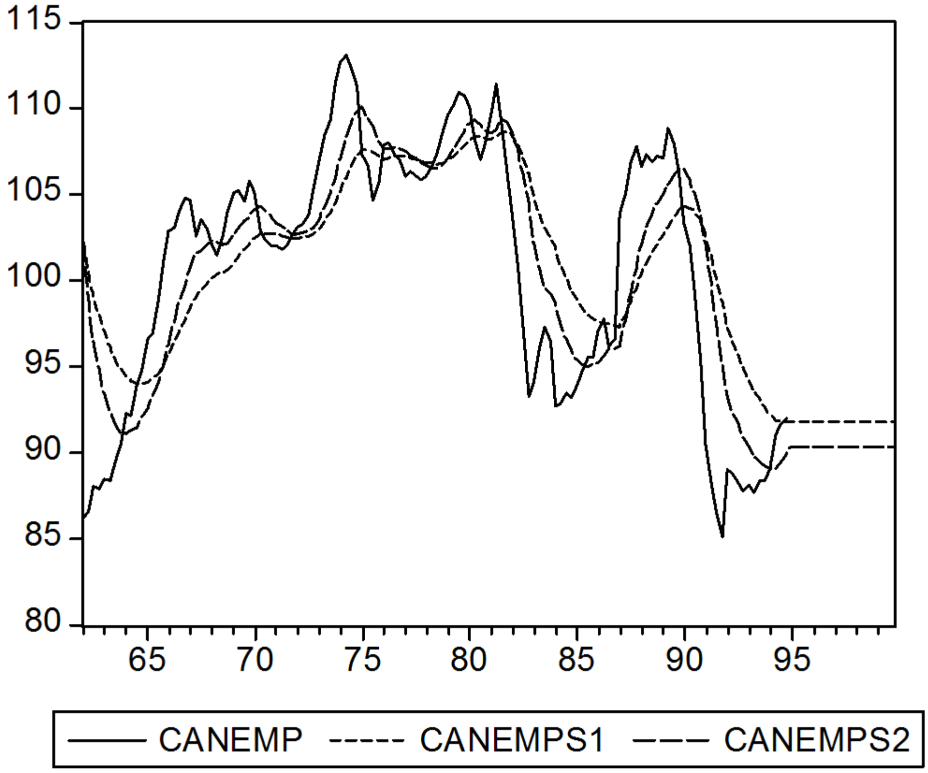

Figure 1: Exponential smoothing on Canadian employment data: $\alpha = 0.1$ and $\alpha = 0.2$

in the series are predicted with a lag of around one to two years. This is a general feature of exponential smoothing models, which are not good at forecasting turning points because they are based on weighted averages of past values.

Simple exponential smoothing is suitable for forecasting a time series that is non-seasonal and has no trend. It can be shown that the simple exponential smoothing model is equivalent to an optimal forecast in the *ARIMA(0,1,1)* model

$$\Delta y_t = \varepsilon_t - \theta \varepsilon_{t-1}. \tag{7}$$

In this model the optimal one-step ahead forecast is given by

$$\Delta \widehat{y}_{t+1,t} = -\theta \varepsilon_t$$

or

$$\widehat{y}_{t+1,t} = y_t - \theta \varepsilon_t. \tag{8}$$

Rewriting (7) and substituting backwards to eliminate lagged values of ε_t we get

$$\begin{aligned}
\varepsilon_t &= \Delta y_t + \theta\varepsilon_{t-1} \\
&= \Delta y_t + \theta\Delta y_{t-1} + \theta^2\Delta y_{t-2} + \cdots \\
&= y_t - (1-\theta)y_{t-1} - \theta(1-\theta)y_{t-2} - \theta^2(1-\theta)y_{t-3} - \cdots .
\end{aligned}$$

Substituting this last expression into (8) for the optimal forecast, we get

$$\widehat{y}_{t+1,t} = (1-\theta)y_t + \theta(1-\theta)y_{t-1} + \theta^2(1-\theta)y_{t-2} + \theta^3(1-\theta)y_{t-3} + \cdots$$

or

$$\widehat{y}_{t+1,t} = (1-\theta)\sum_{j=0}^{t-1}\theta^j y_{t-j}$$

which is an exponential smoothing model with parameter $\alpha = 1-\theta$. Thus the simple exponential smoothing model gives optimal forecasts for the integrated *MA(1)* process (7). Note that this process, although it has a unit root, is not trended since $E\Delta y_t = 0$ so that $Ey_t = y_{t-1}$. Note also that since α must lie between 0 and 1, the equivalence between exponential smoothing and the optimal forecast in the *ARIMA(0,0,1)* model (7) *only holds* for positive θ which means *negative autocorrelation.*

4 Double exponential smoothing

Double exponential smoothing generalises simple exponential smoothing to deal with time series with a linear trend. It uses the simple smoothing algorithm twice, with the same smoothing parameter α, to define a single smoothed series $\overline{y}_t$ and a double smoothed series $\overline{\overline{y}}_t$. The recursions are defined by

$$\overline{y}_t = \alpha y_t + (1-\alpha)\overline{y}_{t-1}$$

and

$$\overline{\overline{y}}_t = \alpha\overline{y}_t + (1-\alpha)\overline{\overline{y}}_{t-1}$$

where the two smoothed series are initialised by

$$\overline{\overline{y}}_t = \overline{y}_t = y_1.$$

Forecasts from double smoothing are then defined by

$$\begin{aligned}
\widehat{y}_{t+h,t} &= (2 + \frac{\alpha h}{1-\alpha})\overline{y}_t - (1 + \frac{\alpha h}{1-\alpha})\overline{\overline{y}}_t \\
&= \left[2\overline{y}_t - \overline{\overline{y}}_t\right] + [\frac{\alpha}{1-\alpha}(\overline{y}_t - \overline{\overline{y}}_t)]h
\end{aligned}$$

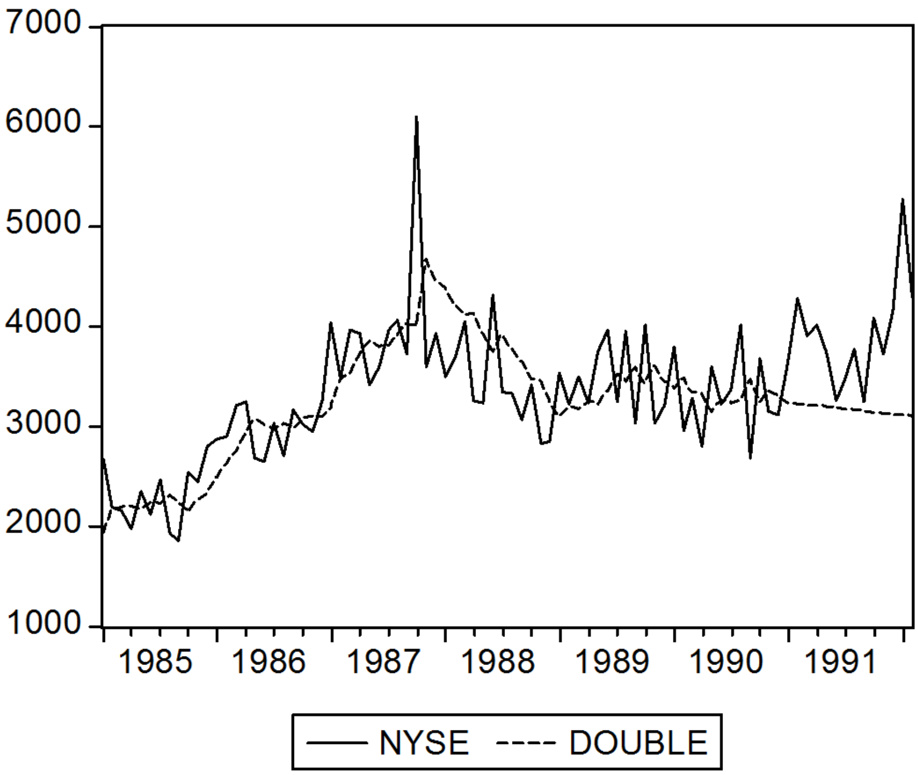

Figure 2: Double exponential smoothing on NYSE data

which is a straight line forecast with the first term in square brackets being the intercept and the second term being the slope of the trend.

Figure 2 illustrates double exponential smoothing applied to the New York Stock Exchange data series. The parameter $\alpha = 0.144$ was estimated over the period 1947 to 1990 and then the smoothed series forecast from 1991-1992. The forecast is a downward sloping trend which is rather surprising given the strong upward trend in the series as a whole.

5 Holt-Winters smoothing

Holt (1957) and Winters (1960) developed an alternative smoothing procedure to deal with non-seasonal but *trended* time series. It is defined by the initialisations

$$\overline{y}_2 = y_2$$

and

$$T_2 = y_2 - y_1$$

and the recursions

$$\overline{y}_t = \alpha y_t + (1-\alpha)(\overline{y}_{t-1} + T_{t-1})$$

and

$$T_t = \beta \Delta \overline{y}_t + (1-\beta) T_{t-1}$$

for $t = 3, \ldots T$, where $0 < \alpha < 1$ and $0 < \beta < 1$. The additional component here, T_t is the smoothed trend and its smoothness is controlled by the parameter β while α controls the smoothness of the level.

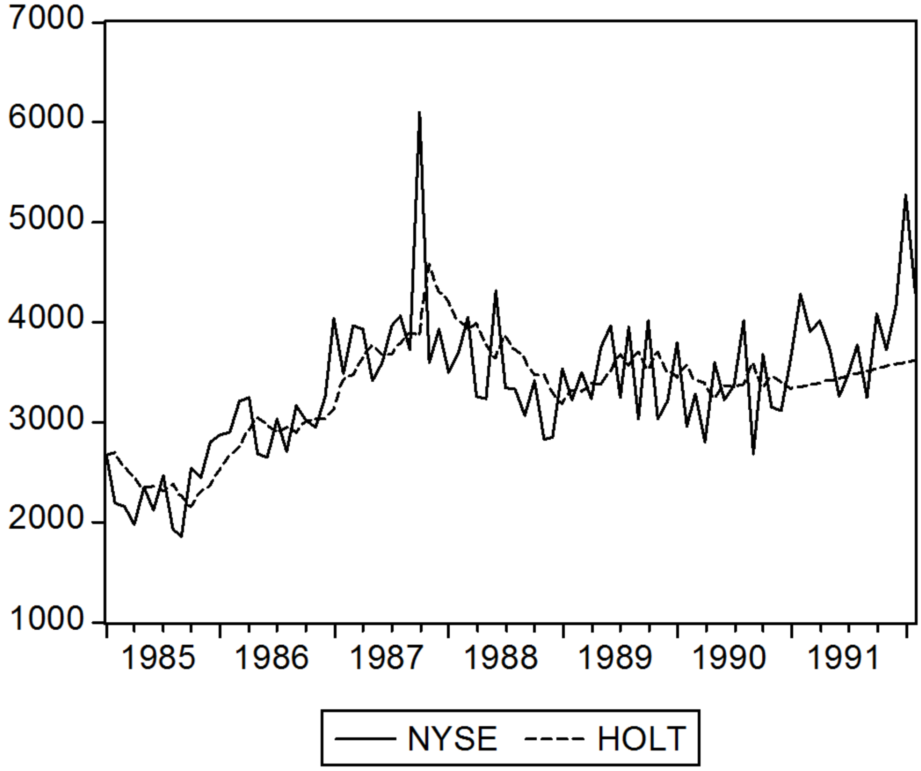

Figure 3: Holt-Winters smoothing on NYSE data

Forecasts are generated by the formula

$$\widehat{y}_{t+h,t} = \overline{y}_t + hT_t$$

and will lie on a straight line with an intercept of $\overline{y}_t$ and a slope of T_t.

Figure 3 illustrates Holt-Winters exponential smoothing applied to the New York Stock Exchange data series. The parameters $\alpha = 0.31$ and $\beta = 0$ were estimated over the period 1947 to 1990 and then the smoothed series forecast from 1991-1992. The forecast is an upward sloping straight line trend which is an improvement on the trend from the double exponential smoothing model.

6 Holt-Winters smoothing with seasonality

Holt and Winters also developed a procedure for dealing with seasonal trended time series with periodicity s. This is defined by the initialisations

$$\overline{y}_s = \frac{1}{s}\sum_{t=1}^{s} y_t$$

$$T_s = 0$$

$$S_j = \frac{y_j}{\overline{y}_s}, \quad j = 1, \cdots, s$$

and the recursions

$$\overline{y}_t = \alpha(y_t - S_{t-s}) + (1-\alpha)(\overline{y}_{t-1} + T_{t-1})$$

$$T_t = \beta\Delta\overline{y}_t + (1-\beta)T_{t-1}$$

$$S_t = \gamma(y_t - \overline{y}_t) + (1-\gamma)S_{t-s}$$

for $t = s+1, \ldots T$, where $0 < \alpha, \beta, \gamma < 1$. The additional component here, S_t is the smoothed additive seasonal component and its smoothness is controlled by the parameter γ.

Forecasts are generated by the formula

$$\widehat{y}_{t+h,t} = \overline{y}_t + hT_t + S_{t+[h/s]+1-s}$$

where $[h/s]$ is the remainder when h is divided by s so that the forecast uses the last s smoothed seasonal components S_t, S_{t-1}, …, S_{t-s+1}.

Figure 4 shows the Holt-Winters seasonal smoothing model applied the the US liquor series where the parameters α, β and γ have been estimated over the period 1960-1990 and the smoothed series forecast over the period 1991-1994. The estimated parameter values were $\alpha = 0.32$, $\beta = 0$ and $\gamma = 0.95$. As can be seen, the forecast is a straight line trend plus a constant seasonal factor.

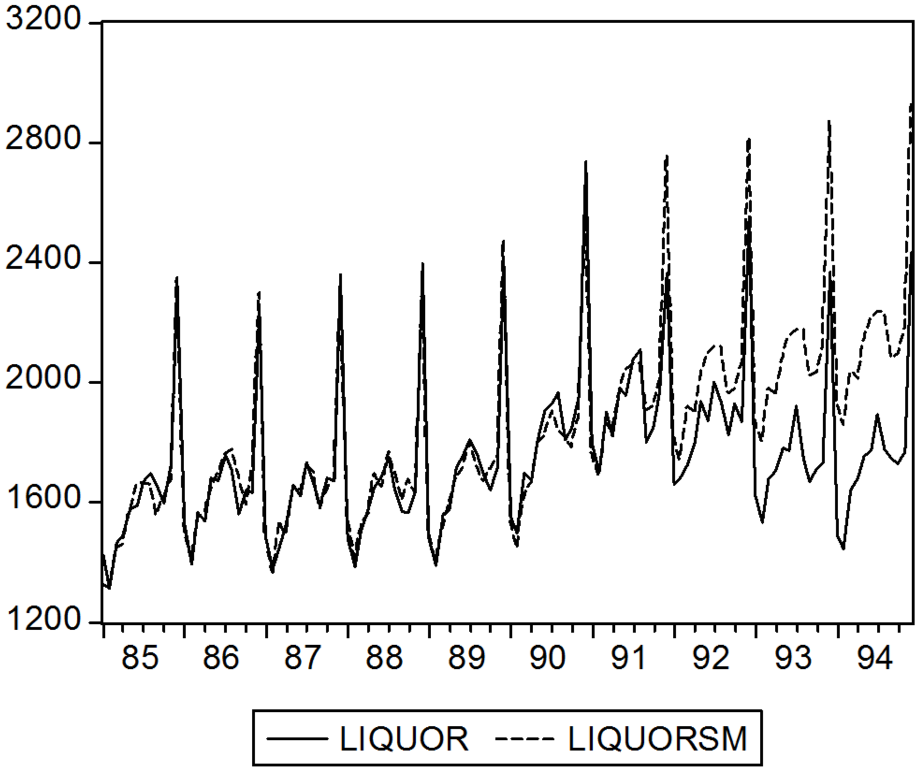

Figure 4: Holt-Winters seasonal smoothing on US Liquor series

References

[1] Holt, C.C. (1957), 'Forecasting seasonals and trends by exponentially weighted moving averages', *Office of Naval Research Memorandum 52*, Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

[2] Winters, P.R. (1960), 'Forecasting sales by exponentially weighted moving averages', *Management Science*, 6, 324–342.