Methodology of the Monthly Index of Services

Annex B: The Holt-Winters Forecasting Method

Single Exponential Smoothing

This is the simplest form of exponential smoothing and can be used only for data without any systematic trend or seasonal components. Given such a time series, a sensible approach is to take a weighted average of past values. So for a series, y_1, y_2,…, y_n, the estimate of the value of y_{n+1}, given the information available up to time n, is:

$$\hat{y}_{n+1|n} = w_0 y_n + w_1 y_{n-1} + w_2 y_{n-2} + ...$$

or $$\hat{y}_{n+1|n} = \sum_{i=0}^{\infty} w_i y_{n-i}$$

where w_i are the weights given to the past values of the series and sum to one. Since the most recent observations of the series are also the most relevant, it is logical that they should be given more weight at the expense of observations further in the past. This is achieved by assigning geometrically declining weights to the series. These decrease by a constant ratio and are of the form:

$$w_i = a(1-a)^i$$

where i=0, 1, 2,… and a is the smoothing constant in the range 0 < a < 1. For example, if a is set to 0.5, the weights will be:

w_0 = 0.5
w_1 = 0.25
w_2 = 0.125 and so on.

It can be seen that weights taking this form will sum to ~1 for all values of a in the above range. The equation for the estimate of y_{n+1} then becomes:

$$\hat{y}_{n+1|n} = ay_n + a(1-a)y_{n-1} + a(1-a)^2 y_{n-2} + ...$$

Since:

$$\hat{y}_{n+1|n} = ay_n + (1-a)(ay_{n-1} + a(1-a)y_{n-2} + ...)$$

it can be seen that:

$$\hat{y}_{n+1|n} = ay_n + (1-a)\hat{y}_{n|n-1}$$

where $\hat{y}_{n|n-1}$ is simply the previous estimate.

This equation defines exponential smoothing and is extremely useful since we can now update each forecast using only two pieces of data: the latest observation; and the previous forecast. It can also be written in the alternative error correction form as follows:

$$\hat{y}_{n+1|n} = a e_n + \hat{y}_{n|n-1}$$

where $e_n = y_n - \hat{y}_{n|n-1}$ and is the one step ahead prediction error.

The final matter with which to deal is the value of a, the smoothing constant. This will depend on the characteristics of the particular time series, especially the irregularity of the data. A high value of a will lead to the majority of the weight being placed on the most recent observations whereas a low value of a will mean that observations further in the past will gain more importance. A value of a = 1 will obviously make the forecast equal to the final value of the series. The smoothing constant is commonly set between 0.05 and 0.3, although it is possible to estimate a by minimising the sum of squared prediction errors.

This can be done by calculating:

$$\sum e_t^2$$

for values of a between 0 and 1 (say in steps of 0.1), where e_t is the one step ahead prediction error as stated above. By doing this it is possible to obtain, to the nearest tenth, the value of a which minimises the sum of squared errors. Steps of 0.01 can then be taken around this value to minimise the sum of squares further. Accuracy at this point is not particularly important as the sum of squares function tends to be quite flat near the minimum.

Holt's Method

The next step involves introducing a term to take into account the possibility of a series exhibiting some form of trend, whether constant or non-constant. In single exponential smoothing, the forecast function is simply the latest estimate of the level. If a slope component is now added which itself is updated by exponential smoothing, the trend can be taken into account.

For a series $y_1, y_2, \ldots, y_n$, the forecast function, which gives an estimate of the series l steps ahead can be written as:

$$\hat{y}_{n+l|n} = m_n + l b_n \qquad l=1, 2, \ldots$$

where m_n is the current level and b_n is the current slope. Therefore, the one step ahead prediction is simply given by:

$$\hat{y}_{t/t-1} = m_{t-1} + b_{t-1}$$

Since there are now two terms to the exponential smoothing, two separate smoothing constants are required, a_0 for the level and a_1 for the slope. As in single exponential smoothing, the updated estimate of the level m_t is a linear combination of $\hat{y}_{t|t-1}$ and y_t:

$$m_t = \alpha_0 y_t + (1-\alpha_0)(m_{t-1} + b_{t-1}) \qquad 0 < \alpha < 1$$

This provides the level at time, t. Since the level at time t-1 is already known, it is possible to update the estimate of the slope:

$$b_t = \alpha_1 (m_t - m_{t-1}) + (1-\alpha_1) b_{t-1}$$

These equations can also be written in the appropriate error correction form:

$$m_t = m_{t-1} + b_{t-1} + \alpha_0 e_t$$

$$b_t = b_{t-1} + \alpha_0 \alpha_1 e_t$$

This method, known as Holt's method, requires starting values for m_t and b_t to be inputted, and estimates of the values for α_0 and α_1 to be made. It would be typical to find $0.02<\alpha_0,\alpha_1<0.2$, but they can be estimated by minimising the sum of squared errors as in single exponential smoothing. Also, it is often found that $m_1=y_1$, and $b_1=y_2-y_1$ are reasonable starting values.

Holt-Winters Forecasting

Holt's method can be extended to deal with time series which contain both trend and seasonal variations. The Holt-Winters method has two versions, additive and multiplicative, the use of which depends on the characteristics of the particular time series. The latter will be considered first.

The general forecast function for the multiplicative Holt-Winters method is:

$$\hat{y}_{n+l|n} = (m_n + l b_n) c_{n-s+l} \qquad l=1, 2, \ldots$$

where m_n is the component of level, b_n is the component of the slope, and c_{n-s+l} is the relevant seasonal component, with s signifying the seasonal period (e.g. 4 for quarterly data and 12 for monthly data.)

Therefore if a monthly time series is considered, the one step ahead forecast is given by:

$$\hat{y}_{n+1|n} = (m_n + b_n) c_{n-11}$$

The updating formulae for the three components will each require a smoothing constant. If once again α_0 is used as the parameter for the level and α_1 for the slope, and a third constant α_2, is added as the smoothing constant for the seasonal factor, the updating equations will be:

$$m_t = \alpha_0 \frac{y_t}{c_{t-s}} + (1-\alpha_0)(m_{t-1} + b_{t-1})$$

$$b_t = a_1(m_t - m_{t-1}) + (1-a_1)b_{t-1}$$

$$c_t = a_2 \frac{y_t}{m_t} + (1-a_2)c_{t-s}$$

Once again, a_0, a_1, and a_2 all lie between zero and one. If the aforementioned additive version of Holt-Winters was used, the seasonal factor is simply added as opposed to multiplied into the one step ahead forecast function, thus:

$$\hat{y}_{n+1|n} = m_n + b_n + c_{n-11}$$

and the level and seasonal updating equations involve differences as opposed to ratios:

$$m_t = a_0(y_t - c_{t-s}) + (1-a_0)(m_{t-1} + b_{t-1})$$

$$c_t = a_2(y_t - m_t) + (1-a_2)c_{t-s}$$

The slope component, b_t, remains unchanged.

The choice of starting values and smoothing parameters is of some importance and Chatfield and Yar (1988) discuss this in some depth. For starting values, it seems sensible to set the level component m_0, equal to the average observation in the first year, i.e.

$$m_0 = \sum_{t=1}^{s} y_t / s$$

where s is the number of seasons. The starting value for the slope component can be taken from the average difference per time period between the first and second year averages. That is:

$$b_0 = \frac{\left\{\sum_{t=1}^{s} y_t / s\right\} - \left\{\sum_{t=s+1}^{2s} y_t / s\right\}}{s}$$

Finally, the seasonal index starting value can be calculated after allowing for a trend adjustment, as follows:

$$c_0 = \frac{\{y_k - (k-1)b_0/2\}}{m_0} \qquad \text{(multiplicative)}$$

$$c_0 = y_k - \{m_0 + (k-1)b_0/2\} \qquad \text{(additive)}$$

where k=1, 2, …, s. Obviously this will lead to s separate values for c_0, which is what is required to gain the initial seasonal pattern.

The smoothing parameters are often selected between 0.02 and 0.2. It is again possible to estimate them by minimising the sum of the squared one-step-ahead errors, but there is no exclusive combination of a_0, a_1, and a_2 which will minimise the square errors for all t.

Correction for Autocorrelation of Residuals

When the correction for autocorrelation of residuals is included, the Holt-Winters forecast function then becomes:

$$\hat{y}_{n+l|n} = (m_n + lb_n)c_{n-s+l} + r_1^l e_n \qquad \text{(Multiplicative)}$$

$$\hat{y}_{n+l|n} = m_n + b_n + c_{n-s+l} + r_1^l e_n \qquad \text{(Additive)}$$

e_n is the one step ahead forecast error and r_1 is the first order autocorrelation coefficient of the forecast errors, given by:

$$r_1 = \frac{\sum e_n e_{n-1}}{\sum e_n^2}$$

References

"HOLT-WINTERS FORECASTING: SOME PRACTICAL ISSUES" Chatfield and Yar
The Statistician (1988)