Exponential Smoothing

Introduction

A simple method for forecasting. Does not require long series. Enables to decompose the series into a trend and seasonal effects. Particularly useful method when there is a need to forecast many series in “real time”.

Simple exponential smoothing

Suppose that we have a **stationary** series, $X_1, X_2, ..., X_N$, and we want to forecast X_{N+1} as a linear combination of previous observations: $\hat{X}(N,1) = C_0 X_N + C_1 X_{N-1} +, ...$

It makes sense to require $C_0 > C_1 > C_2 > ...$ since we are dealing with a time series.

Let $C_i = \alpha(1-\alpha)^i \Rightarrow \sum_{i=0}^{\infty} C_i = 1$, $0 \le \alpha \le 1$.

Simple exponential smoothing (cont.)

$$\hat{X}(N,1) = \alpha X_N + \alpha(1-\alpha)X_{N-1} + \alpha(1-\alpha)^2 X_{N-2} + \ldots$$

$$= \alpha X_N + (1-\alpha)[\alpha X_{N-1} + \alpha(1-\alpha)X_{N-2} + \ldots]$$

$$\Downarrow$$

$$\hat{X}(N,1) = \alpha X_N + (1-\alpha)\hat{X}(N-1,1) = \hat{X}(N-1,1) + \alpha e_N;$$

$$e_N = [X_N - \hat{X}_{N-1}(1)].$$

Notice: $\hat{X}(N,2) = \alpha\hat{X}(N,1) + (1-\alpha)\hat{X}(N,1) = \hat{X}(N,1)$
and so forth.

Optimality of simple exponential smoothing.

Let $X_t = \mu + \varepsilon_t + \beta(\varepsilon_{t-1} + \varepsilon_{t-2} + \ldots)$

$$W_t = X_t - X_{t-1} = \varepsilon_t - (1-\beta)\varepsilon_{t-1} = \varepsilon_t - \theta\varepsilon_{t-1} \quad \textbf{MA(1)}$$

Under the MA(1) model:

$$\hat{X}_t(1) = X_t - \theta[X_t - \hat{X}_{t-1}(1)] = \hat{X}_{t-1}(1) + (1-\theta)[X_t - \hat{X}_{t-1}(1)]$$

$$= \hat{X}_{t-1}(1) + \alpha[X_t - \hat{X}_{t-1}(1)]; \quad \alpha = (1-\theta).$$

- Forecasting equation of exponential smooth.

Choise of α for simple smoothing

A- Subjective considerations

B- By minimization of sum of squares of forecasting errors:

$$\hat{X}(1,1) = X_1 \Rightarrow e_2 = (X_2 - X_1)$$

$$\hat{X}(2,1) = \hat{X}(1,1) + \alpha e_2 = X_1 + \alpha(X_2 - X_1) \Rightarrow$$

$$\Rightarrow e_3 = X_3 - [\alpha X_2 + (1-\alpha) X_1], \ldots$$

Estimation of α by minimization of $\sum_{t=2}^{N} e_t^2$, or a **weighted** sum. Minimization can be carried out by a grid search in the range $0 < \alpha < 1$.

Important: simple exponential smoothing can only be applied to a stationary series. For a nonstationary series it is no longer effective.

Example:

$$X_t = \gamma_0 + \gamma_1 t; \; \gamma_1 > 0 \Rightarrow \; \alpha X_t + (1-\alpha)\hat{X}_{t-1}(1) < X_{t+1}.$$

Exponential smoothing: Holt & Winters method

A. Additive decomposition: $X_t = L_t + S_t + u_t$

L_t - trend level at time t, S_t - seasonal effect, u_t - irregular (noise) term.

Iterative process. When a new observation X_{t+1} becomes available, **update** L_t, R_T (**slope**, see below) and S_t as follows:

$$L_{t+1} = \alpha\left(X_{t+1} - S_t^{t+1}\right) + (1-\alpha)\left(L_t + R_t\right);$$

$$R_{t+1} = \gamma\left(L_{t+1} - L_t\right) + (1-\gamma)R_t;$$

$$S*_{t+1}^{t+1} = \delta\left(X_{t+1} - L_{t+1}\right) + (1-\delta)S_t^{t+1}; \quad \sum_{i=0}^{s-1} S_{t+1}^{t+1+i} = 0.$$

Forecast *m* steps ahead:

$$\hat{X}_t(m) = L_t + mR_t + S_t^{t+m}$$

$0 \le \delta, \gamma, \alpha \le 1$ are **smoothing coefficients.**

Note: each updating equation is a weighted average of two estimates of the corresponding quantity. One from past observations and a new one.

Alternative equations for additive smoothing

Let $e_{t+1} = X_{t+1} - \hat{X}_t(1)$ denote the one step ahead forecasting error. Then,

$$L_{t+1} = L_t + R_t + \alpha e_{t+1} \; ; \; R_{t+1} = R_t + \alpha\gamma e_{t+1}$$

$$S_{t+1}^{t+1} = S_t^{t+1} + \frac{s-1}{s}\delta(1-\alpha)e_{t+1};$$

$$S_{t+1}^{t+1+i} = S_t^{t+1+i} - \frac{1}{s}\delta(1-\alpha)e_{t+1}, \quad i = 1......(s-1).$$

s is the length of the seasonal cycle (4, 12, …).

Implication: if the sum of the seasonal effects is nullified at the start of the smoothing ($t = 1$), it will remain like this for every time point.

Special case

Consider a non-seasonal series;

$$\hat{X}_t(1) = L_t + R_t \Rightarrow e_t = X_t - (L_{t-1} + R_{t-1}) \Rightarrow X_t = e_t + (L_{t-1} + R_{t-1}).$$

From previous equations,

$L_t - L_{t-1} = R_{t-1} + \alpha e_t$; $R_t - R_{t-1} = \alpha\gamma e_t$ and hence,

$$(1-B)^2 L_t = \alpha\gamma e_{t-1} + \alpha(e_t - e_{t-1}) = \alpha[1-(1-\gamma)B]e_t.$$

Similarly,

$$(1-B)^2(L_{t-1} + R_{t-1}) = \{\alpha B[1-(1-\gamma)B] + \alpha\gamma B(1-B)\}e_t,$$

$$(1-B)^2 X_t = \{(1-B)^2 + \alpha B[1-(1-\gamma)B] + \alpha\gamma B(1-B)\}e_t.$$

Or, $X_t = 2X_{t-1} - X_{t-2} + [\alpha(1+\gamma) - 2]e_{t-1} + (1-\alpha)e_{t-2}$;

optimal prediction for **ARIMA(0,2,2)**.

Exponential smoothing: Holt & Winters method (cont.)

B. Multiplicative decomposition:

$X_t = L_t \times \tilde{S}_t \times I_t$

$\tilde{S}_t$, I_t are now percentages ($\tilde{S}_t$ is a **seasonal factor**, I_t is the **irregular term**.)

Rationale of multiplicative decomposition:

the **seasonal effect** (**not** seasonal factor) is

proportional to the **trend level**.

Suppose that in months $t, t+12, t+24$,$\tilde{S} = 1.1$.

Month	L	$L\tilde{S}$	$L\tilde{S} - L$
t	100	110	10
$t+12$	200	220	20
$t+24$	300	330	30

$\tilde{S}$ is the **seasonal factor**,

$L\tilde{S} - L$ is the **seasonal effect**.

Exponential smoothing for multiplicative decomposition

Possibility I: use the log transformation, apply additive smoothing:

$$Y_t = \log(X_t) = \log(\tilde{L}_t) + \log(\tilde{S}_t) + \log(I_t)$$

Transforming back the smoothed values yields

$$R_{t+1} = \left(\frac{L_{t+1}}{L_t}\right)^{\gamma} \times (R_t)^{(1-\gamma)} \;\; ; \;\; L_{t+1} = \left(\frac{X_{t+1}}{\tilde{S}_t^{t+1}}\right)^{\alpha} (L_t \times R_t)^{(1-\alpha)}$$

$$\tilde{S}_{t+1}^{*t+1} = \left(\frac{X_{t+1}}{L_{t+1}}\right)^{\delta} (\tilde{S}_t^{t+1})^{(1-\delta)}; \qquad \prod_{i=0}^{s-1} \tilde{S}_{t+1}^{t+1+i} = 1.$$

- Geometric mean of seasonal factors **=1**.

Forecast: $\hat{X}_t(m) = L_t \times (R_t)^m \times \tilde{S}_t^{t+m}$.

Prediction error: $e_t = (X_t / \hat{X}_{t/t-1})$.

Problem: assumes implicitly that the trend evolves in constant rates. Not very realistic in practice.

Exponential smoothing for multiplicative decomposition (cont.)

Possibility II: (original procedure of Winter**)**.

$$L_{t+1} = \alpha\left(\frac{X_{t+1}}{\tilde{S}_t^{t+1}}\right) + (1-\alpha)\left(L_t + R_t\right) \text{ ;}$$

$$R_{t+1} = \gamma\left(L_{t+1} - L_t\right) + (1-\gamma)R_t \text{ ;}$$

$$\tilde{S}_{t+1}^{*t+1} = \delta\left(\frac{X_{t+1}}{L_{t+1}}\right) + (1-\delta)\tilde{S}_t^{t+1} \text{ ;} \qquad \sum_{i=0}^{s-1} \tilde{S}_{t+1}^{t+1+i} = s$$

Implications:

The trend evolves in **constant increments;**

Arithmetic mean of seasonal factors **=1**.

Why require that the arithmetic mean of the **seasonal factors =1? Reasonable** to impose that the **sum** of the **seasonal effects** over **s** successive time points is null. Let **s=12**.

If $\sum_{t=1}^{12} \tilde{S}_t / 12 = 1,$ $\Rightarrow$

$$\sum_{t=1}^{12} (L_t \tilde{S}_t - L_t) = \sum_{t=1}^{12} L_t (\tilde{S}_t - 1) = 12 Cov(L_t, \tilde{S}_t) = 0.$$

- **Trend level** and **seasonal factors** "**uncorrelated**".

Starting values for smoothing

Suppose that the series is long enough such that we can use the first **3** years for starting values.

Trend levels: $\frac{1}{24}[2][12]X_t$**,** centered moving average. First a simple moving average of 12 successive points and then a simple moving average of every two successive averages. This way the first trend value is for **t=7**.

Seasonal effects (factors): average the two differences $(X_t - \hat{L}_t) \cong \hat{S}_t$ (additive decomposition) or the two ratios $(X_t / \hat{L}_t) \cong \tilde{S}_t$ (multiplicative decomposition), for each calendar month. **Subtract** (**divide by**) the mean of the resulting estimates from each estimate such that their mean equals **0** for the additive decomposition (equals **1** for the multiplicative decomposition).

Increment: $\frac{1}{12}(\hat{L}_{25} + \ldots + \hat{L}_{36}) - \frac{1}{12}(\hat{L}_{13} + \ldots + \hat{L}_{24})$, or just, $(\hat{L}_{36} - \hat{L}_{35})$.

Graphical test for choosing between additive and multiplicative decomposition

Compute $\hat{L}_t$ for the whole series in the same way as for the starting values, and $\hat{S}_t = (X_t - \hat{L}_t)$. Plot $\hat{S}_t$ against $\hat{L}_t$ **for each calendar month separately**.

The monthly estimates $\hat{S}_t$ may be subjected to large noise, so an alternative procedure is to plot the annual geometric means $(\prod_{j=1}^{12} | X_{ij} - \hat{L}_{ij} |)^{1/12}$ against $(\prod_{j=1}^{12} \hat{L}_{ij})^{1/12}$.

Rationale: Denote $\hat{S}_{ij} = (X_{ij} - \hat{L}_{ij})$. If $S_{ij} = K_j L_{ij}$

$$\Rightarrow (\prod_{j=1}^{12} | S_{ij} |) = \prod_{j=1}^{12} | K_j | \prod_{j=1}^{12} L_{ij} = K \prod_{j=1}^{12} L_{ij} .$$

Graphical test of smoothing performance (CUSUM)

Let $m_1 = e_1$, $m_2 = m_1 + e_2 = e_1 + e_2$ **...** $m_t = m_{t-1} + e_t$,

m_t **=** the sum of forecasting errors until time t.

Plot m_t against t. $-1 \le [m_t / \sum_{k=1}^{t} | e_k |] \le 1$ and we expect it to be close to **0**.

Example

The table below shows the sales of a business in 4 years.

Year\quarter	I	II	III	IV
1	221	303	358	288
2	221	325	398	326
3	257	358	421	364
4	275	380	464	421

It is desired to forecast the amount of sales in the third and fourth quarters of Year 4, using the data until the second quarter of Year 4.

We shall use the first 3 years for starting values and start the smoothing from the third quarter of Year 3 ($t = 11$). Let $\alpha = \gamma = \delta = 0.1$.

Computation of starting values

Starting values for trend levels: $\frac{1}{8}[2][4]X_t$.

$$\hat{L}_t = 0.125X_{t-2} + 0.25X_{t-1} + 0.25X_t + 0.25X_{t+1} + 0.125X_{t+2}$$

$$\hat{L}_3 = 0.125 \times 221 + 0.25(303 + 358 + 288) + 0.125 \times 221 = 292.5$$

Similar computations yield,

Year\quarter	I	II	III	IV
1	287*	289.75*	292.5	295.25
2	303	312.75	322	330.63
3	337.63	345.25	352.25	357.25
4	365.38	377.88	390.38*	402.88*

(*) The first two values are computed by subtracting the estimated increment (295.25-292.5). The last two values are computed by adding the estimated increment (377.88-365.38). These values are used for the graphical test but are not needed for the smoothing process.

Computation of starting values (cont.)

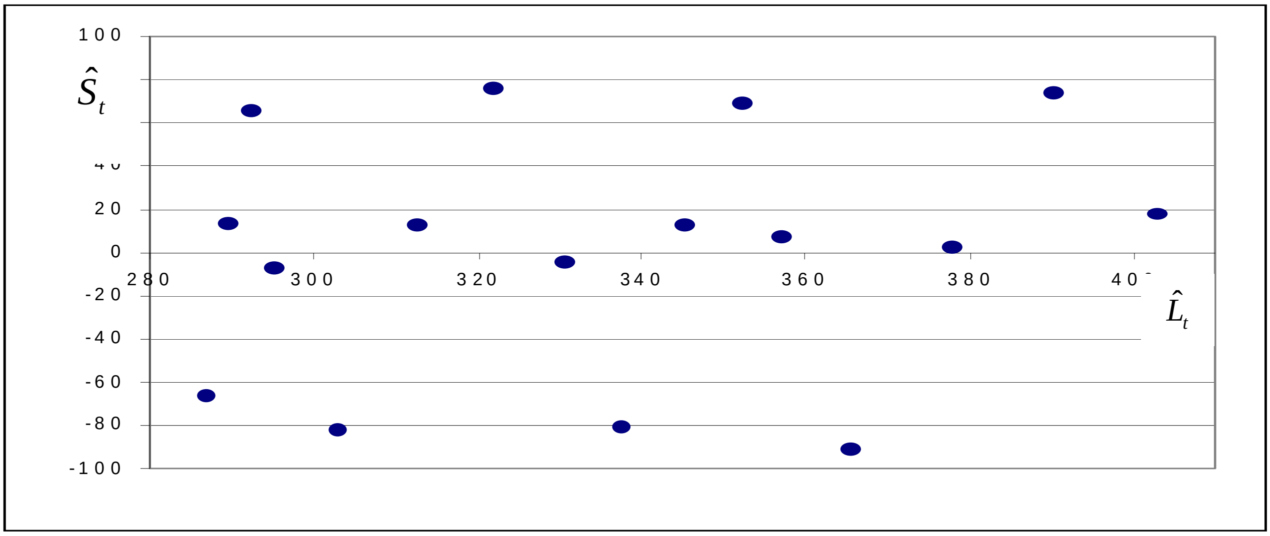
Starting values for seasonal effects: $\hat{S}_t = (X_t - \hat{L}_t)$

Year\quarter	I	II	III	IV	Mean
1	-66.0	13.25	65.5	-7.25	1.375
2	-82.0	12.25	76	-4.63	0.405
3	-80.6	12.75	68.75	6.75	1.905
4	-90.4	2.12	73.62	18.12	0.87

- The table shows that the series is highly seasonal. Annual average of seasonal effects close to zero despite of very rough estimation of trend levels.

- The figures in the table don't indicate that the seasonal effects increase when the trend increases, so an additive decomposition seems right. (One can apply both procedures and compare the forecasts.)

Graphical test



The Figure shows very clearly that there is no apparent relationship between $\hat{S}_t$ and $\hat{L}_t$, suggesting that the additive decomposition is more appropriate.

Computation of the annual geometric means of $|\hat{S}_t| = |X_t - \hat{L}_t|$yields:

$$\bar{G}_{(1)} = 25.39,\ \bar{G}_{(2)} = 24.38,\ \bar{G}_{(3)} = 26.28,\ \bar{G}_{(4)} = 22.49.$$

The annual geometric means of $\hat{L}_t$ are:

$$\bar{L}_{(1)} = 291.11,\ \bar{L}_{(2)} = 316.93,\ \bar{L}_{(3)} = 348.02,\ \bar{L}_{(4)} = 383.88$$

- The annual geometric means of the trend levels increase, but the annual geometric means of the seasonal effects are more or less constant.

Computation of starting values for smoothing

$$S_1^* = [-82 + (-80.63)]/2 = -81.32$$
$$S_2^* = (12.25 + 12.75)/2 = 12.5$$
$$S_3^* = (65.5 + 76)/2 = 70.75 \qquad \Rightarrow \qquad \underline{\bar{S}* = -1}$$
$$S_4^* = -(7.25 + 4.63)/2 = -5.94$$

Starting values for seasonal effects by quarter:

$\underline{S_1 = S_1^* - \bar{S}* = -80.32}$, $\underline{S_2 = 13.5}$, $\underline{S_3 = 71.75}$,
$\underline{S_4 = -4.94}$

We start the smoothing at time $t = 10$.

$$\underline{\hat{L}_{10} = 345.25,\ \hat{R}_{10} = \hat{L}_{10} - \hat{L}_9 = 7.62}$$

Smoothing under the additive decomposition

$$e_{11} = X_{11} - \hat{X}_{10}(1) = 421 - (345.25 + 7.62 + 71.75) = -3.62$$

$\hat{S}_{10}^{11} = 71.75 \rightarrow$ seasonal effect computed at time $t = 10$ for time $t = 11$.

Smoothed values for time $t = 11$

$$\hat{L}_{11} = \hat{L}_{10} + \hat{R}_{10} + \alpha e_{11} = 345.25 + 7.62 + 0.1(-3.62) = 352.83$$

$$\hat{R}_{11} = \hat{R}_{10} + \alpha\gamma e_{11} = 7.62 - (0.1)^2(-3.62) = 7.58$$

$$S_3 = S_{10}^3 + 0.75\delta(1-\alpha)e_{11}$$
$$= 71.75 + 0.75 \times 0.1 \times 0.9(-3.62) = 71.51$$

$$S_4 = S_{10}^4 - 0.25\delta(1-\alpha)e_{11}$$
$$= -4.94 - 0.25 \times 0.1 \times 0.9(-3.62) = -5.02$$

$$S_1 = S_{10}^1 - 0.25\delta(1-\alpha)e_{11}$$
$$= -80.32 - 0.25 \times 0.1 \times 0.9(-3.62) = -80.40$$

$$S_2 = S_{10}^2 - 0.25\delta(1-\alpha)e_{11}$$
$$= 13.5 - 0.25 \times 0.1 \times 0.9(-3.62) = -5.02$$

$$\sum_{j=1}^{4} S_j = 0$$

New forecast:

$$X_{11}(1) = 352.83 + 7.58 - 5.02 = 355.4$$

New forecast error: $e_{12} = 364 - 355.4 = 8.61$

Smoothing (cont.)

Continuing the same way for times $t = 12$, $t = 13$ and $t = 14$ yields the following smoothed values:

time	$\hat{L}_t$	$\hat{R}_t$	$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_4$	$\hat{X}_t(1)$	e_{t+1}
***t*=12**	**361.27**	**7.66**	**-80.21**	**13.61**	**71.70**	**-4.44**	**288.72**	**-13.72**
***t*=13**	**368.8**	**7.52**	**-81.14**	**13.30**	**71.40**	**-4.75**	**389.62**	**-9.62**
***t*=14**	**376.22**	**7.42**	**-81.36**	**12.62**	**71.20**	**-4.97**	**454.84**	**9.16**

Prediction for time *t*=15:

$\hat{X}_{14}(1) = 376.22 + 7.42 + 71.2 =$ 454.84

$e_{15} = 464 - 454.84 =$ 9.16 **(less than 2%)**.

Prediction for time *t*=15:

$\hat{X}_{14}(2) = 376.22 + 2 \times 7.42 - 4.97 =$ 386.1

$e_{15} =$ 34.9 **(8% error)**