Chapter 3 - Exponential smoothing

AMA4650
Fu Guanxing
guanxing.fu@polyu.edu.hk

Motivation

- Recall simple moving average in Chapter 1

$$M_t = \frac{x_{t-1} + \cdots x_{t-k}}{k}.$$

- The same weight of $x_{t-1} \cdots x_{t-k}$ does not reflect the fast change of data[1].
- More weight to the recent data and less to the historical data.
- **Exponential smoothing:** more emphasis is put on recent observations and the weights are decaying exponentially with time.
- Exponential smoothing techniques are suitable for non-stationary data, that is, data with a trend and/or seasonality.

[1] For example, around maturity, trading volume and volatility exhibit distinct and predictable behaviors. ES can help model this volatility by capturing recent fluctuations while dampening the effect of older, less relevant data. This allows risk managers to assess and hedge positions more effectively, especially during the critical period leading up to maturity.

Outline

- Simple exponential smoothing
- Double exponential smoothing
- Holt's trend method
- Damped Trend
- Holt-Winters' seasonal method

3.1 Simple Exponential Smoothing

Formulation.

- Let S_T be the one step forecast at time T, which is formulated by

$$S_T = \alpha x_T + (1-\alpha) S_{T-1},$$

where $0 < \alpha < 1$ is a given constant. $\{S_t\}$ is called *simple exponential smoothing (SES) forecast*.

- The SES forecast **at time** T = a weighted sum of the observation **at time** T and the forecast **for time** T.

Remark.

1. One step forecast at time T is the **forecast for** $T+1$!
2. Repeatedly use the SES formula,

$$\begin{aligned} S_T &= \alpha x_T + (1-\alpha) S_{T-1} \\ &= \alpha x_T + (1-\alpha)(\alpha x_{T-1} + (1-\alpha) S_{T-2}) \\ &= \alpha (x_T + (1-\alpha) x_{T-1}) + (1-\alpha)^2 S_{T-2} \\ &= \cdots \\ &= \alpha \sum_{k=0}^{T'-1} (1-\alpha)^k x_{T-k} + (1-\alpha)^{T'} S_{T-T'}, \end{aligned}$$

which implies ↷

Remark 2 *cont'd.*

- Let $T' = T$:

$$S_T = \alpha \sum_{k=0}^{T-1} (1-\alpha)^k x_{T-k} + (1-\alpha)^T S_0.$$

- Assume the forecast is bounded and let $T' \to \infty$:

$$S_T = \alpha \sum_{k=0}^{\infty} (1-\alpha)^k x_{T-k},$$

  which is a weighted average of all past observations.

3. The coefficient $(1-\alpha)^k = e^{k \log(1-\alpha)}$ decays exponentially.

Summary of SES

Let $\{x_t\}$ be a time series. The τ-step (ahead) forecast at time T is

$$\hat{x}_{T+\tau} = S_T,$$

where

1. Component form:
$$S_T = \alpha x_T + (1-\alpha) S_{T-1}.$$
2. Error correction form:
$$S_T = S_{T-1} + \alpha(x_T - S_{T-1}).$$
3. Weighted average form:
$$S_T = \alpha \sum_{k=0}^{T-1} (1-\alpha)^k x_{T-k} + (1-\alpha)^T S_0.$$
4. Infinite sum form:
$$S_T = \alpha \sum_{k=0}^{\infty} (1-\alpha)^k x_{T-k}.$$

Here, α is called the smoothing constant.

Determination of S_0 and α

Determination of S_0.

- When T gets larger, $(1-\alpha)^T$ becomes small, the contribution of S_0 to S_T can be ignored. Thus, for large data sets, the estimation of S_0 is not important.
- Method 1 for S_0 estimation: set $S_0 = x_1$.
- Method 2 for S_0 estimation: take average of the available data or a subset of the available data, and let S_0 be this average.

Determination of α (and S_0). Define the sum of the squared one-step-ahead forecast residuals as

$$\text{SSR}(\alpha, S_0) = \sum_{t=1}^{n} (x_t - S_{t-1})^2.$$

Calculate $\text{SSR}(\alpha, S_0)$ for various values of (α, S_0), and choose the (α, S_0) with minimal SSR.

Remark. S_0 and α are usually given in exercises.

Example 3.1 (1/2)

The monthly sales of a kind of chair in the past six months are shown in the following table. Using $S_0 = 393.13$, $\alpha = 0.1$, the smoothed statistics and the one-step ahead forecasts are shown in the following table.

Sequence no.	Actual sales, x_T	Smoothed stat, S_T	Forecast (one-step)
1	330	386.82	393.13
2	410	389.14	387
3	408	391.02	389
4	514	403.32	391
5	402	403.19	403
6	343	397.17	403

Remark: S_T is the forecast made at time T, so it is the forecast of X_{T+1}, but NOT the forecast of X_T!

Example 3.1 (2/2)

Solution. Recall $S_T = \alpha x_T + (1-\alpha)S_{T-1}$.

Observe

$$S_1 = 0.1x_1 + 0.9S_0.$$

That means

$$S_1 = 0.1(330) + 0.9(393.13) = 386.82,$$

and that is the forecast of X_2. Similarly,

$$S_2 = 0.1x_2 + 0.9S_1$$

So

$$S_2 = 0.1(410) + 0.9(386.82) = 389.14,$$

and that is the forecast of X_3.

$$\ldots$$

SES forecast for constant process

- Let $X_t = b + \epsilon_t$ be a constant process, where $\epsilon \sim NID(0, \sigma^2)$
- Let $\widehat{b}(T)$ be the one-step forecast/estimate of the level b at time T.
- $\widehat{b}(T) = S_T$ provides an **unbiased** forecast for b!

Relation with weighted least square

- Using (weighted) least square to estimate b, we minimize[2]

$$\text{SSR} = \sum_{t=1}^{T}(1-\alpha)^{T-t}\big(x_t - b\big)^2,$$

where recent errors are assigned more weight.

- Minimize SSR w.r.t. b gives

$$2\sum_{t=1}^{T}(1-\alpha)^{T-t}\big(x_t - b\big) = 0 \Rightarrow b\sum_{t=1}^{T}(1-\alpha)^{T-t} = \sum_{t=1}^{T}(1-\alpha)^{T-t}x_t$$

$$\Rightarrow \widehat{b}_T = \frac{\alpha}{1-(1-\alpha)^T}\sum_{t=1}^{T}(1-\alpha)^{T-t}x_t.$$

- Recall the SES estimate for b

$$S_T = \alpha\sum_{k=0}^{T-1}(1-\alpha)^k x_{T-k} + (1-\alpha)^T S_0 = \alpha\sum_{k=1}^{T}(1-\alpha)^{T-k}x_k + (1-\alpha)^T S_0$$

$$\approx \widehat{b}_T, \qquad \text{if } T \text{ is large enough.}$$

[2] SSR = sum of squared residuals

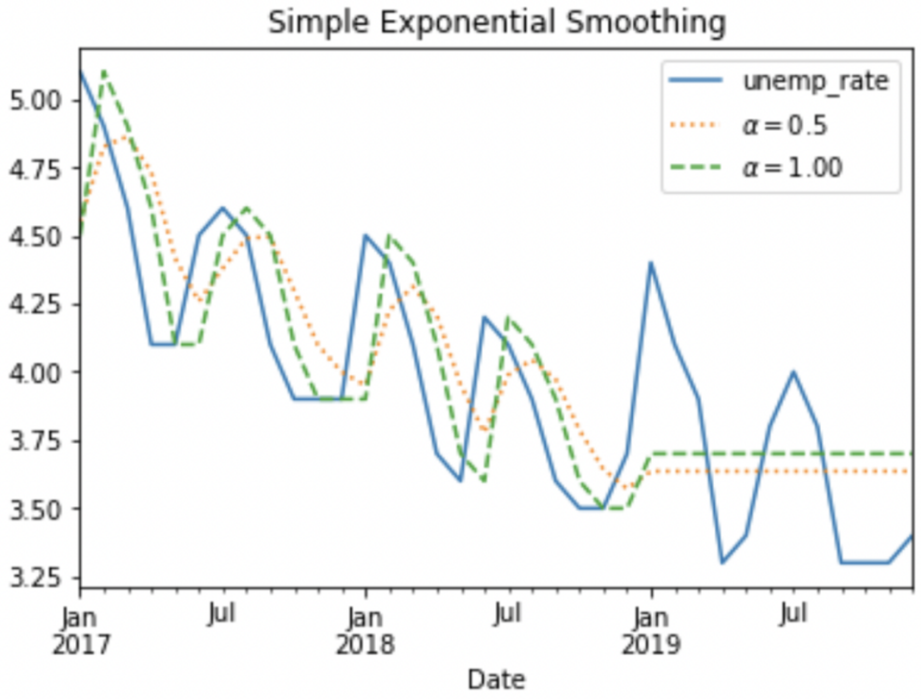


Figure 3.1: Forecast by SES.

Exponential smoothing for a linear trend process

Consider a linear trend process which has the form

$$X_t = b_1 + b_2 t + \epsilon_t, \quad \text{where } \epsilon_t \sim NID(0, \sigma^2).$$

If we use simple exponential smoothing, then recall

$$S_T = \alpha \sum_{k=0}^{\infty} (1-\alpha)^k X_{T-k},$$

Exponential smoothing for a linear trend process

We have

$$\begin{aligned}\mathbb{E}[S_T] &= b_1 + b_2 T - \left(\frac{1-\alpha}{\alpha}\right) b_2 \\ &= \mathbb{E}[X_T] - \left(\frac{1-\alpha}{\alpha}\right) b_2.\end{aligned} \tag{1}$$

Thus the smoothed statistic is a **biased** forecast for a linear trend process.

↷ Double exponential smoothing