QBUS 6840 Lecture 5

Exponential Smoothing

Professor Junbin Gao

The University of Sydney Business School

Outline

Exponential smoothing

- Simple Exponential Smoothing
- Trend Corrected Exponential Smoothing (Holt's Linear Trend Method)

Reading

- Online Textbook Sections 7.1-7.2 (https://otexts.org/fpp2/expsmooth.html); and/or
- BOK Sec 8.1-8.3

Exponential smoothing methods

- In simple terms, exponential smoothing forecasts are weighted averages of previous observations. The weights decay exponentially as we go further into the past.
- Useful when parameters or components are changing with time.

Simple exponential smoothing 指数平滑

- Naïve Method

$$\widehat{y}_{T+1|1:T} = y_T$$

- Overall Average Method

$$\widehat{y}_{T+1|1:T} = \frac{1}{T}\sum_{t=1}^{T} y_t = \frac{1}{T}(y_T + y_{T-1} + \cdots + y_1)$$

- Something between two extremes

$$\begin{aligned}
\widehat{y}_{T+1|1:T} &= \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots + \alpha(1-\alpha)^{T-1} y_1 \\
&= \alpha y_T + (1-\alpha)[\alpha y_{T-1} + \alpha(1-\alpha)y_{T-2} + \cdots + \alpha(1-\alpha)^{T-2} y_1] \\
&= \alpha y_T + (1-\alpha)\widehat{y}_{T|1:T-1}
\end{aligned}$$

- Weighted average form

$$\widehat{y}_{t+1|1:t} = \alpha y_t + (1-\alpha)\widehat{y}_{t|1:t-1}$$

The forecast at time $t+1$ is equal to a weighted average between the most recent observation y_t and the most recent forecast $\widehat{y}_{t|1:t-1}$.

Two Alternative Forms

- The Component Form

$$\begin{aligned} l_t &= \alpha y_t + (1-\alpha) l_{t-1}, \qquad 0 \le \alpha \le 1. \\ \widehat{y}_{t+1|1:t} &= l_t. \end{aligned}$$

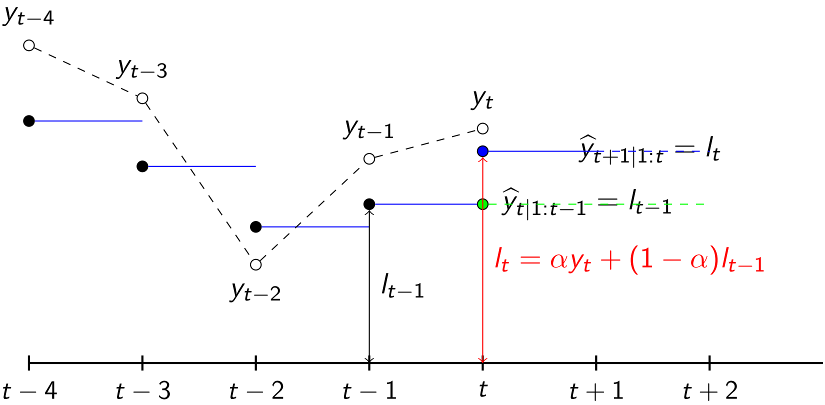
  l_t is called the level (or the smoothed value) of the series at time t. We first calculate the level l_t, then use it as the forecast $\widehat{y}_{t+1|1:t}$.

- The Error Correction Form

$$\begin{aligned} \widehat{y}_{t+1|1:t} &= l_t. \\ l_t &= \alpha y_t + (1-\alpha) l_{t-1} = l_{t-1} + \alpha (y_t - l_{t-1}) \\ &= l_{t-1} + \alpha \varepsilon_t \end{aligned}$$

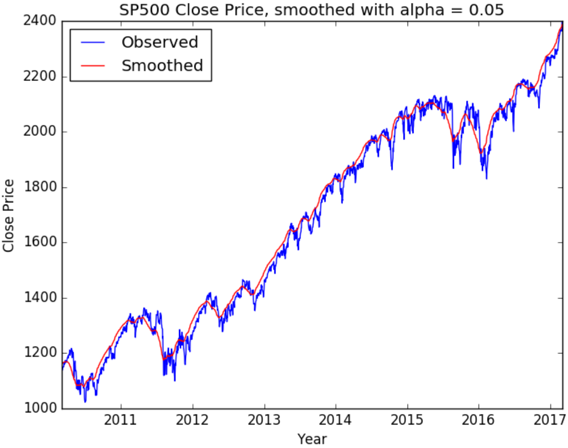
  where $\varepsilon_t = y_t - l_{t-1} = y_t - \widehat{y}_{t|1:t-1}$ is the forecast error at time t.

Explanation: Simple exponential smoothing



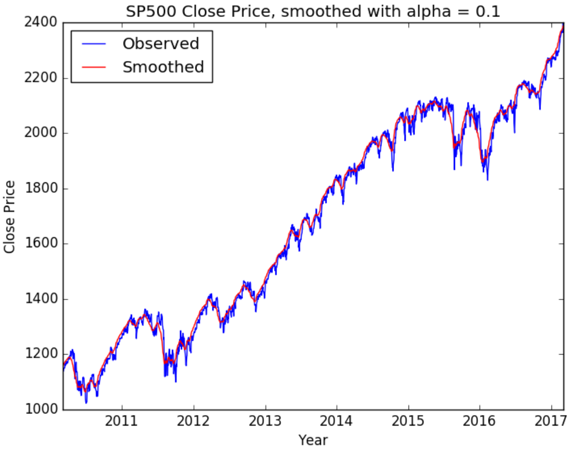
S&P 500 Closing Price (`Lecture05_Example01.py`)

Exponential smoothing with $\alpha = 0.05$



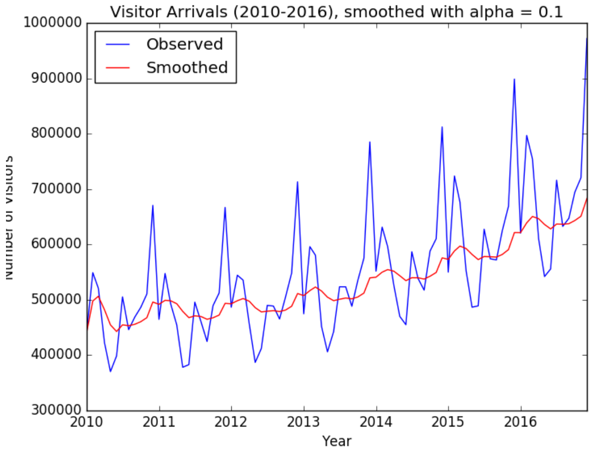
S&P 500 Closing Price (Lecture05_Example01.py)

Exponential smoothing with $\alpha = 0.1$



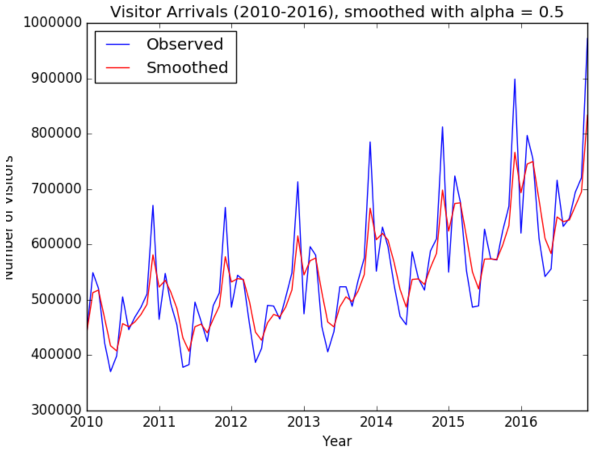
Visitor arrivals in Australia

Exponential smoothing with $\alpha = 0.1$



Visitor arrivals in Australia

Exponential smoothing with $\alpha = 0.5$



Simple exponential smoothing
Weights

Specify an initial value l_0 (an estimate or a guess, e.g., the average of y_1, y_2, y_3).

$$l_1 = \alpha y_1 + (1-\alpha) l_0$$

$$l_2 = \alpha y_2 + (1-\alpha) l_1 = \alpha y_2 + (1-\alpha)\alpha y_1 + (1-\alpha)^2 l_0$$

$$l_3 = \alpha y_3 + (1-\alpha) l_2 = \alpha y_3 + (1-\alpha)\alpha y_2 + (1-\alpha)^2 \alpha y_1 + (1-\alpha)^3 l_0$$

Simple exponential smoothing

Weights

$$
\begin{aligned}
l_4 &= \alpha y_4 + (1-\alpha) l_3 \\
&= \alpha y_4 + (1-\alpha)\alpha y_3 + (1-\alpha)^2 \alpha y_2 + (1-\alpha)^3 \alpha y_1 + (1-\alpha)^4 l_0
\end{aligned}
$$

$$\vdots$$

$$
\begin{aligned}
l_t =& \alpha y_t + (1-\alpha) l_{t-1} \\
=& \alpha y_t + (1-\alpha)\alpha y_{t-1} + (1-\alpha)^2 \alpha y_{t-2} + \ldots + (1-\alpha)^{t-1} \alpha y_1 \\
& + (1-\alpha)^t l_0
\end{aligned}
$$

A WMA smoother (is moving?)

Choice of initial level

- We left l_0 unspecified above.
- How should we set it?
  - Use the average of very initial observations, i.e., y_1, y_2, y_3 etc., or even simply y_1
  - Take l_0 as a parameters, and use an algorithm to estimate it.

Simple exponential smoothing

Some notes

- Useful when level is changing, but not too much.
- Weights all previous observations in smoothing.
- Weights decrease exponentially.
- Weights add to 1 (always, check it!).
- Low α reveals trend-cycle; Higher α reveals seasonality.
- Also called EWMA (exponentially weighted moving average)

Simple exponential forecasting: Theoretical Model

- The basic model:

$$l_t = \alpha y_t + (1-\alpha) l_{t-1}, \qquad 0 \le \alpha \le 1.$$

  and

$$y_{t+1} = l_t + \varepsilon_{t+1}, \text{ with } \varepsilon_{t+1} \sim N(0, \sigma^2).$$

- We assume all ε_t's are independent of each other
- The level is the underlying mechanism, where the next observation is "noised" current level.
- What is about when $\alpha = 1$?

- Recall the basic model:

$$\begin{aligned} y_{t+1} &= l_t + \varepsilon_{t+1} \\ l_t &= \alpha y_t + (1-\alpha) l_{t-1}. \end{aligned}$$

- We can chose α (and l_0) by minimising

$$\begin{aligned} SSE &= \sum_{t=1}^{n} (y_t - l_{t-1})^2 \\ &= \sum_{t=1}^{n} (y_t - \alpha y_{t-1} - (1-\alpha) l_{t-2})^2. \end{aligned}$$

Muth (1960) showed that EWMA forecasts are minimum MSE for the following statistical model

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t; \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma_e^2) \\
\mu_{t+1} &= \mu_t + \xi_t \qquad \xi_t \sim \mathcal{N}(0, \sigma_\xi^2)
\end{aligned}
$$

where $\{\varepsilon_t\}$ and $\{\xi_t\}$ are two independent Gaussian white noise series. This is an example of state-space models. We re-visit this in later lectures.

The initial value μ_1 is either given or follows a known distribution, and is independent of $\{\varepsilon_t\}$ and $\{\xi_t\}$.

Simple exponential smoothing: Analysis

Error correction formulation

- The basic model for the level can be rewritten in terms of errors

$$\begin{aligned} l_t &= \alpha y_t + (1-\alpha) l_{t-1} \\ &= l_{t-1} + \alpha (y_t - l_{t-1}) \\ &= l_{t-1} + \alpha \varepsilon_t . \end{aligned}$$

- The next observation is

$$\begin{aligned} y_{t+1} = l_t + \varepsilon_{t+1} &= l_{t-1} + \alpha \varepsilon_t + \varepsilon_{t+1} = \cdots \\ &= l_0 + \alpha \varepsilon_1 + \cdots + \alpha \varepsilon_t + \varepsilon_{t+1} \end{aligned}$$

- Similarly

$$\begin{aligned} y_{t+2} = l_{t+1} + \varepsilon_{t+2} &= l_t + \alpha \varepsilon_{t+1} + \varepsilon_{t+2} \\ &= l_{t-1} + \alpha \varepsilon_t + \alpha \varepsilon_{t+1} + \varepsilon_{t+2} = \cdots \\ &= l_0 + \alpha \varepsilon_1 + \cdots + \alpha \varepsilon_t + \alpha \varepsilon_{t+1} + \varepsilon_{t+2} \end{aligned}$$

Simple exponential smoothing: Analysis

Forecast equations

- The forecast is defined as the average over all possible uncertainty [Carefully understand the meaning of this!]

$$\widehat{y}_{t+h|1:t} := E(y_{t+h}|y_{1:t}), \qquad l_t = l_{t-1} + \alpha\varepsilon_t$$

- We have already observed up to time t, so l_t is certain. Uncertainty occurs after this time point, e.g., ε_{t+1}

$$\begin{aligned}\widehat{y}_{t+1|1:t} =& E(l_t + \varepsilon_{t+1}|y_{1:t}) \quad [\text{Note: } y_{t+1} = l_t + \varepsilon_{t+1}] \\ =& E(l_t) + E(\varepsilon_{t+1}|y_{1:t}) = l_t + E(\varepsilon_{t+1}) = l_t + 0 = l_t\end{aligned}$$

where we have used the assumption $\varepsilon_{t+1} \sim N(0, \sigma^2)$, i.e., $E(\varepsilon_{t+1}) = 0$.
- So, similarly

$$\widehat{y}_{t+2|1:t} = E(l_{t+1} + \varepsilon_{t+2}|y_{1:t}) = l_{t+1}$$

- Is this correct?

Simple exponential smoothing: Analysis

Forecast equations

- The forecast is defined as the average over all possible uncertainty [Carefully understand the meaning of this!]

$$\widehat{y}_{t+h|1:t} := E(y_{t+h}|y_{1:t}), \qquad l_t = l_{t-1} + \alpha\varepsilon_t$$

- We have already observed up to time t, so l_t is certain. Uncertainty occurs after this time point, e.g., ϵ_{t+1}

$$\begin{aligned}\widehat{y}_{t+1|1:t} =& E(l_t + \varepsilon_{t+1}|y_{1:t}) \quad [\text{Note: } y_{t+1} = l_t + \varepsilon_{t+1}] \\ =& E(l_t) + E(\varepsilon_{t+1}|y_{1:t}) = l_t + E(\varepsilon_{t+1}) = l_t + 0 = l_t\end{aligned}$$

where we have used the assumption $\varepsilon_{t+1} \sim N(0, \sigma^2)$, i.e., $E(\varepsilon_{t+1}) = 0$.

- Similarly

$$\begin{aligned}\widehat{y}_{t+2|1:t} &= E(l_{t+1} + \varepsilon_{t+2}|y_{1:t}) = E(l_t + \alpha\varepsilon_{t+1}|y_{1:t}) = l_t \\ &\vdots \\ \widehat{y}_{t+h|1:t} &= E(l_{t+h-1} + \varepsilon_{t+h}|y_{1:t}) = E(l_{t+h-2} + \alpha\varepsilon_{t+h-1}|y_{1:t}) = l_t\end{aligned}$$

Hence the forecast is always l_t after time t.

- Recall the model again:

$$y_{t+1} = l_t + \varepsilon_{t+1}, \qquad l_t = l_{t-1} + \alpha\varepsilon_t$$

- Consider the variance of the new observation

$$\begin{aligned}\mathrm{Var}(y_{t+1}|y_{1:t}) &= \mathrm{Var}(l_t + \varepsilon_{t+1}|y_{1:t}) \\ &= \mathrm{Var}(l_t) + \mathrm{Var}(\varepsilon_{t+1}|y_{1:t}) \\ &= 0 + \sigma^2 = \sigma^2. \quad \text{Why?}\end{aligned}$$

- Similarly

$$\begin{aligned}\mathrm{Var}(y_{t+2}|y_{1:t}) &= \mathrm{Var}(l_{t+1} + \varepsilon_{t+2}|y_{1:t}) = \mathrm{Var}(l_t + \alpha\varepsilon_{t+1} + \varepsilon_{t+2}|y_{1:t}) \\ &= \mathrm{Var}(l_t) + \mathrm{Var}(\alpha\varepsilon_{t+1}) + \mathrm{Var}(\varepsilon_{t+2}|y_{1:t}) \\ &= 0 + \alpha^2\mathrm{Var}(\varepsilon_{t+1}) + \mathrm{Var}(\varepsilon_{t+2}|y_{1:t}) \\ &= \alpha^2\sigma^2 + \sigma^2 = \sigma^2(1+\alpha^2)\end{aligned}$$

Simple exponential smoothing

Variance for interval forecasts

$$\begin{aligned}
\mathrm{Var}(y_{t+3}|y_{1:t}) &= \mathrm{Var}(l_{t+2} + \varepsilon_{t+3}|y_{1:t}) \\
&= \mathrm{Var}(l_{t+1} + \alpha\varepsilon_{t+2} + \varepsilon_{t+3}|y_{1:t}) \\
&= \mathrm{Var}(l_t + \alpha\varepsilon_{t+1} + \alpha\varepsilon_{t+2} + \varepsilon_{t+3}|y_{1:t}) \\
&= \sigma^2(1 + 2\alpha^2)
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(y_{t+h}|y_{1:t}) &= \mathrm{Var}(l_{t+h-1} + \varepsilon_{t+h}|y_{1:t}) \\
&= \mathrm{Var}(l_{t+h-2} + \alpha\varepsilon_{t+h-1} + \varepsilon_{t+h}|y_{1:t}) \\
&= \mathrm{Var}(l_t + \sum_{i=1}^{h-1} \alpha\varepsilon_{t+h-i} + \varepsilon_{t+h}|y_{1:t}) \\
&= \sigma^2(1 + (h-1)\alpha^2)
\end{aligned}$$

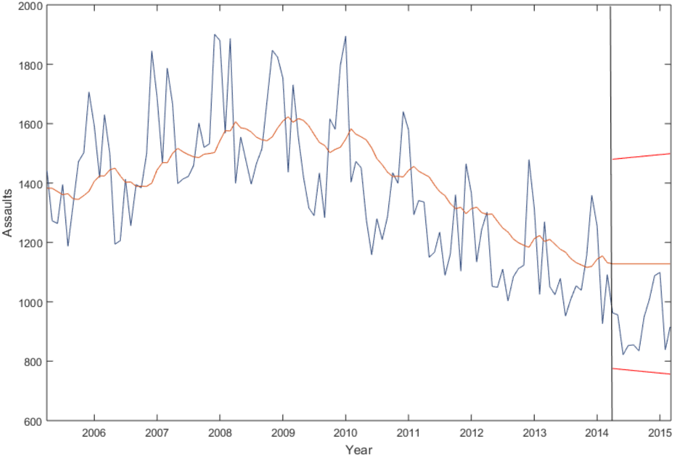
Simple exponential smoothing

Forecasting: collecting the results

$$\widehat{y}_{t+h|1:t} = \mathrm{E}(y_{t+h}|y_{1:t}) = l_t$$

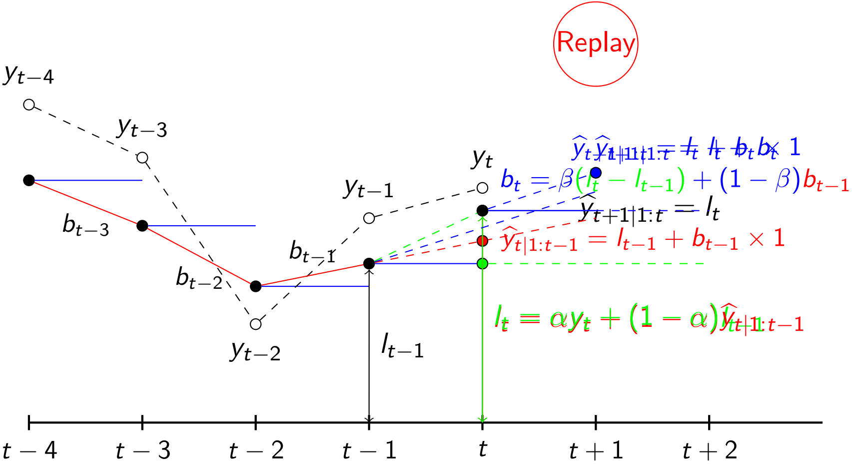
$$\mathrm{Var}(y_{t+h}|y_{1:t}) = \sigma^2(1 + (h-1)\alpha^2)$$

What happens as h increases?



Are the forecasts reasonable?

Explanation: Including Trend Information



Trend corrected exponential smoothing

$$
\begin{aligned}
l_t &= \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1} \times 1), \qquad 0 \le \alpha \le 1 \\
&\quad (= \alpha y_t + (1-\alpha)\widehat{y}_t | 1 : t-1)) \\
b_t &= \beta(l_t - l_{t-1}) + (1-\beta) b_{t-1}, \qquad 0 \le \beta \le 1
\end{aligned}
$$

$$
\widehat{y}_{t+1|1:t} = l_t + b_t \times 1.
$$

Trend corrected exponential smoothing

$$
\begin{aligned}
l_t &= \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}), \qquad 0 \le \alpha \le 1 \\
&\quad (= \alpha y_t + (1-\alpha)\widehat{y}_{t|1:t}) \\
b_t &= \beta(l_t - l_{t-1}) + (1-\beta) b_{t-1}, \qquad 0 \le \beta \le 1
\end{aligned}
$$

$$
\widehat{y}_{t+1|1:t} = l_t + b_t.
$$

Trend corrected exponential smoothing

Model

$$
\begin{aligned}
y_{t+1} &= l_t + b_t + \varepsilon_{t+1} \\
l_t &= \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) \\
b_t &= \beta(l_t - l_{t-1}) + (1-\beta) b_{t-1} \\
\varepsilon_{t+1} &\sim N(0, \sigma^2)
\end{aligned}
$$

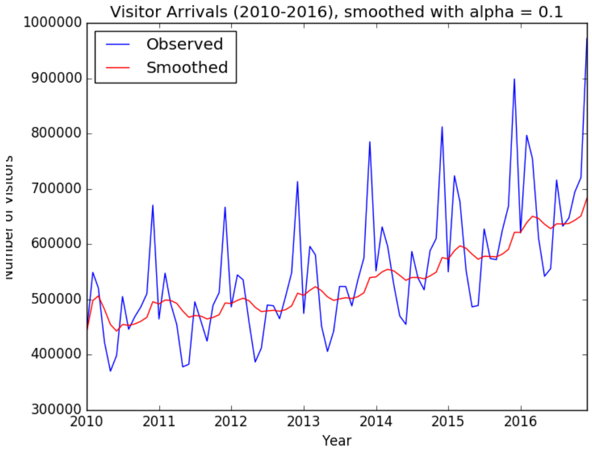
We can choose α and β by minimising

$$
\text{SSE}(\alpha, \beta) = \sum_{t=2}^{n} (y_t - l_{t-1} - b_{t-1})^2
$$

You have opportunity to try this in Assignment 1!

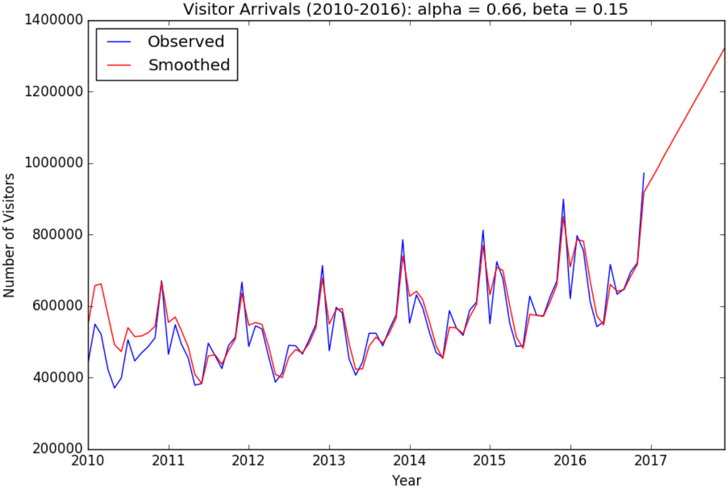
Visitor arrivals in Australia: `Lecture05_Example02.py`

Original series (2010-2016)



Visitor arrivals in Australia: `Lecture05_Example02.py`

Seasonally adjusted series (2010-2016)



Trend corrected exponential smoothing

Error correction formulation

The basic model for the trend corrected exponential smoothing can be written in many ways. We can express all the components in terms of errors:

$$
\begin{aligned}
l_t &= \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) \\
&= l_{t-1} + b_{t-1} + \alpha(y_t - l_{t-1} - b_{t-1}) \\
&= l_{t-1} + b_{t-1} + \alpha\varepsilon_t
\end{aligned}
$$

$$
\begin{aligned}
y_{t+1} &= l_{t-1} + b_{t-1} + b_t + \alpha\varepsilon_t + \varepsilon_{t+1} \\
&= l_t + b_t + \varepsilon_{t+1}
\end{aligned}
$$

Trend corrected exponential smoothing

Error correction formulation

$$
\begin{aligned}
b_t &= \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \\
&= b_{t-1} + \beta(l_t - l_{t-1} - b_{t-1}) \\
&= b_{t-1} + \beta\alpha\varepsilon_t \;\; \text{from } l_t = l_{t-1} + b_{t-1} + \alpha\varepsilon_t \\
(&= b_{t-1} + \beta\alpha(y_t - l_{t-1} - b_{t-1}))
\end{aligned}
$$

$$
\begin{aligned}
l_t &= l_{t-1} + b_{t-1} + \alpha\varepsilon_t \\
b_t &= b_{t-1} + \beta\alpha\varepsilon_t \\
y_{t+1} &= l_t + b_t + \varepsilon_{t+1} \\
&\quad (= l_{t-1} + 2b_{t-1} + \alpha(1+\beta)\varepsilon_t + \varepsilon_{t+1})
\end{aligned}
$$

Trend corrected exponential smoothing

Forecasting equations

$$
\begin{aligned}
\widehat{y}_{t+1|1:t} :&= E(y_{t+1}|y_{1:t}) \\
&= E(l_t + b_t + \varepsilon_{t+1}|y_{1:t}) \\
&= l_t + b_t \\
(&= \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1})) \\
(&= l_{t-1} + 2b_{t-1} + \alpha(1+\beta)\varepsilon_t)
\end{aligned}
$$

$$
\begin{aligned}
\widehat{y}_{t+2|1:t} &= E(l_{t+1} + b_{t+1} + \varepsilon_{t+2}|y_{1:t}) \\
&= E(l_t + 2b_t + \alpha(1+\beta)\varepsilon_{t+1} + \varepsilon_{t+2}|y_{1:t}) \\
&= l_t + 2b_t
\end{aligned}
$$

Trick: We iteratively expand the formula until we arrive the time point all are known.

Trend corrected exponential smoothing

Forecasting equations

$$
\begin{aligned}
\widehat{y}_{t+3|1:t} &= E(l_{t+2} + b_{t+2} + \varepsilon_{t+3} | y_{1:t}) \\
&= E((l_{t+1} + b_{t+1} + \alpha\varepsilon_{t+2}) + (b_{t+1} + \beta\alpha\varepsilon_{t+2}) + \varepsilon_{t+3}) \\
&= E(l_{t+1} + 2b_{t+1} + \alpha(1+\beta)\varepsilon_{t+2} + \varepsilon_{t+3}) \\
&= E((l_t + b_t + \alpha\varepsilon_{t+1}) + 2(b_t + \beta\alpha\varepsilon_{t+1}) + \alpha(1+\beta)\varepsilon_{t+2} + \varepsilon_{t+3}) \\
&= E(l_t + 3b_t + \alpha(1+2\beta)\varepsilon_{t+1} + \alpha(1+\beta)\varepsilon_{t+2} + \varepsilon_{t+3}) \\
&= l_t + 3b_t \\
&\vdots \\
\widehat{y}_{t+h|1:t} &= l_t + hb_t
\end{aligned}
$$

Trend corrected exponential smoothing

Variance for interval forecasts

$$
\begin{aligned}
\mathrm{Var}(y_{t+1}|y_{1:t}) &= \mathrm{Var}(l_t + b_t + \varepsilon_{t+1}|y_{1:t}) \\
&= \sigma^2
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}(y_{t+2}|y_{1:t}) &= \mathrm{Var}(l_{t+1} + b_{t+1} + \varepsilon_{t+2}|y_{1:t}) \\
&= \mathrm{Var}(l_t + 2b_t + \alpha(1+\beta)\varepsilon_{t+1} + \varepsilon_{t+2}|y_{1:t}) \\
&= \sigma^2(1 + \alpha^2(1+\beta)^2)
\end{aligned}
$$

Trend corrected exponential smoothing

Variance for interval forecasts

$$\begin{aligned}
\mathrm{Var}(y_{t+3}|y_{1:t}) &= \mathrm{Var}(l_{t+2} + b_{t+2} + \varepsilon_{t+3}|y_{1:t}) \\
&= \mathrm{Var}(l_{t+1} + 2b_{t+1} + \alpha(1+\beta)\varepsilon_{t+2} + \varepsilon_{t+3}|y_{1:t}) \\
&= \mathrm{Var}(l_t + 3b_t + \alpha(1+2\beta)\varepsilon_{t+1} + \alpha(1+\beta)\varepsilon_{t+2} + \varepsilon_{t+3}|y_{1:t}) \\
&= \mathrm{Var}(\alpha(1+2\beta)\varepsilon_{t+1} + \alpha(1+\beta)\varepsilon_{t+2} + \varepsilon_{t+3}|y_{1:t}) \\
&= \alpha^2(1+2\beta)^2\sigma^2 + \alpha^2(1+\beta)^2\sigma^2 + \sigma^2 \\
&= \sigma^2(1 + \alpha^2(1+\beta)^2 + \alpha^2(1+2\beta)^2)
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(y_{t+h}|y_{1:t}) &= \mathrm{Var}\left(l_t + hb_t + \alpha\sum_{i=1}^{h-1}(1+i\beta)\varepsilon_{t+i} + \varepsilon_{t+h}|y_{1:t}\right) \\
&= \sigma^2\left(1 + \alpha^2\sum_{i=1}^{h-1}(1+i\beta)^2\right) \\
&= \sigma^2\left(1 + \alpha^2\left(\frac{\beta^2}{6}h(h-1)(2h-1) + (\beta h + 1)(h-1)\right)\right)
\end{aligned}$$

(I used the formula for the sum of an arithmetic progression to get to the last step)

Trend corrected exponential smoothing

Forecasting: collecting the results

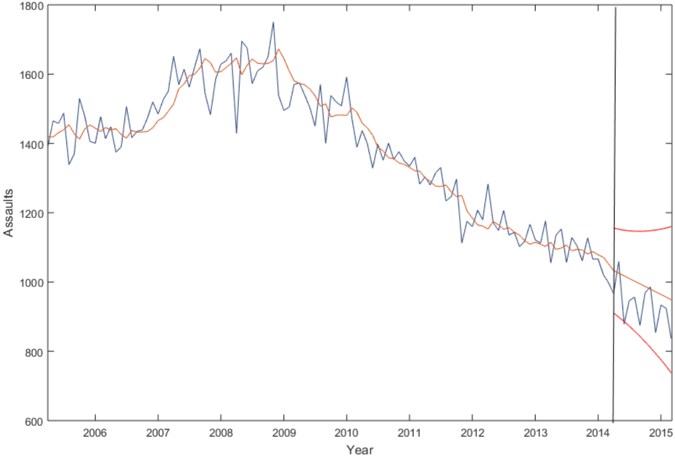
$$\widehat{y}_{t+h|1:t} = \widehat{l_t} + h\widehat{b}_t$$

$$\text{Var}(y_{t+h}|y_{1:t}) = \sigma^2 \left(1 + \alpha^2 \left(\frac{\beta^2}{3}h(h-1)(h-2) + (\beta+1)h - 1\right)\right)$$

What happens as h increases?

Alcohol related assaults in NSW

Forecasting the seasonally adjusted series (last 12 months)

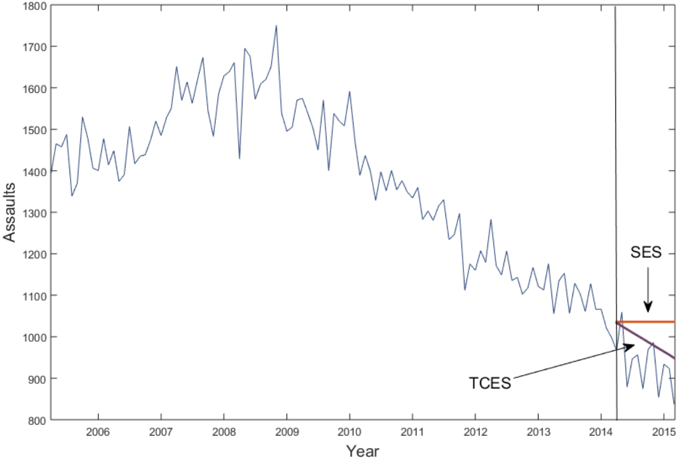


Comparison

What criteria should we use to compare SES and TCES?

Alcohol related assaults in NSW

Forecasting the seasonally adjusted series (last 12 months)



Alcohol related assaults in NSW

Forecasting the seasonally adjusted series (last 12 months)

One month ahead forecasts

	SES	TCES
RMSE	70.9	63.5
MAE	56.5	55.8
MAPE	6.2	6.0