QBUS6840 Lecture 6

Exponential Smoothing (Seasonal)

Discipline of Business Analytics

The University of Sydney Business School

Recap and an example

Last lecture, we discussed

- simple exponential smoothing
- the underlying statistical model and equivalent forms of the updates
- the error correction form for getting the prediction intervals
- extending the model to handle trend

Examples: Covid-19 forecasting [Google Cloud and Havard Global Health Institute]: https://cloud.google.com/blog/products/ai-machine-learning/google-cloud-is-releasing-the-covid-19-public-forecasts, Covid-19 mobility report [Google] https://www.google.com/covid19/mobility/

Last week: Simple exponential smoothing (SES)

- SES for forecasting in the 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 $\widehat{y}_{t|t-1}$.

Last week: Two Alternative Forms of SES

- 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.

Last week: Trend corrected exponential smoothing (TCES)

Recall the component form of SES:

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

SES doesn't take trend into account.

Trend corrected exponential smoothing:

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

Table of contents

Readings

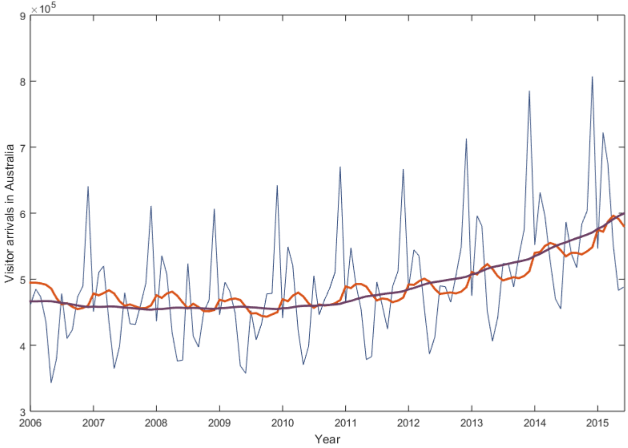
Online textbook Sections 7.4-7.5, or BOK Sec 8.4-8.5

Outline

Visitor arrivals in Australia

Original series (2006-2015), SES and TCES

Smoothing by SES (black) and TCES (red). They are not suitable for seasonal data



Holt-Winters smoothing

- is an Exponential smoothing method for data with seasonality.
- used for both Additive seasonality and Multiplicative seasonality.
  - Additive seasonality: the seasonal variation is constant along the trend
  - Multiplicative seasonality: the seasonal variation is changing along the trend

Additive Holt-Winters smoothing

Suppose the time series $\{Y_1, Y_2, ..., Y_T\}$ has an additive seasonality with seasonal frequency M.

The basic idea of the Holt-Winters method is to use exponential smoothing for all the level, trend and seasonal component.

$$
\begin{aligned}
l_t &= \alpha(y_t - S_{t-M}) + (1-\alpha)(l_{t-1} + b_{t-1}), \qquad 0 \leq \alpha \leq 1 \\
b_t &= \gamma(l_t - l_{t-1}) + (1-\gamma)b_{t-1}, \qquad 0 \leq \gamma \leq 1 \\
S_t &= \delta(y_t - l_t) + (1-\delta)S_{t-M}, \qquad 0 \leq \delta \leq 1.
\end{aligned}
$$

The forecast of Y_{t+1} is

$$\widehat{y}_{t+1} = l_t + b_t + S_{t+1-M}.$$

Evolution of techniques

Simple exponential smoothing:

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

Trend corrected exponential smoothing:

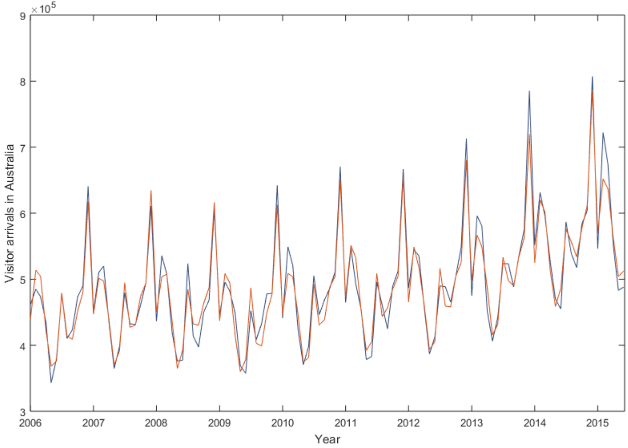
$$\begin{aligned} l_t &= \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}), \qquad 0 \le \alpha \le 1 \\ b_t &= \gamma(l_t - l_{t-1}) + (1-\gamma) b_{t-1}, \qquad 0 \le \gamma \le 1 \\ \widehat{y}_{t+1} &= l_t + b_t. \end{aligned}$$

Holt-Winters method

$$\begin{aligned} l_t &= \alpha(y_t - S_{t-M}) + (1-\alpha)(l_{t-1} + b_{t-1}), \qquad 0 \le \alpha \le 1 \\ b_t &= \gamma(l_t - l_{t-1}) + (1-\gamma) b_{t-1}, \qquad 0 \le \gamma \le 1 \\ S_t &= \delta(y_t - l_t) + (1-\delta) S_{t-M}, \qquad 0 \le \delta \le 1 \\ \widehat{y}_{t+1} &= l_t + b_t + S_{t+1-M}. \end{aligned}$$

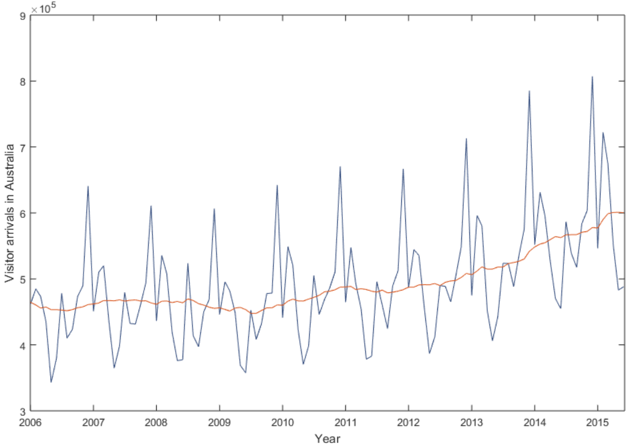
Visitor arrivals in Australia: Additive Holt-Winters method

See `Lecture06_Example01.py`



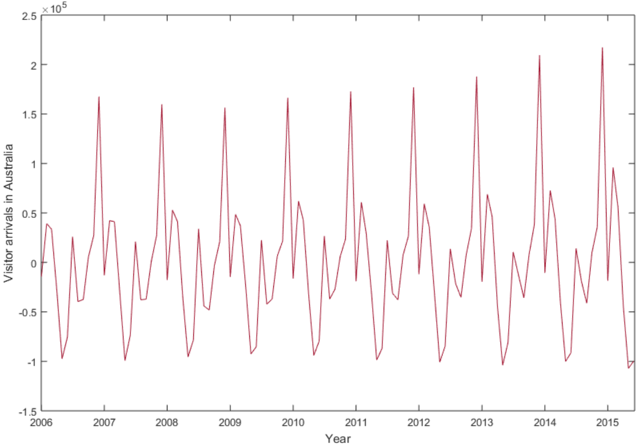
Visitor arrivals in Australia

Additive Holt-Winters level component estimate



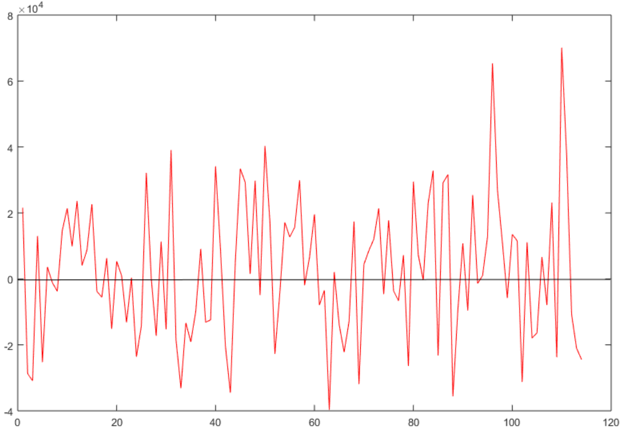
Visitor arrivals in Australia

Additive Holt-Winters seasonal factors



Visitor arrivals in Australia

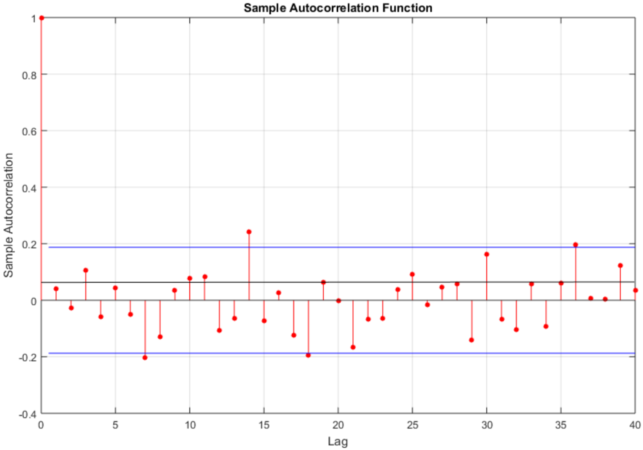
Additive Holt-Winters residuals



Visitor arrivals in Australia

Additive Holt-Winters residual autocorrelations

You will learn about Sample Autocorrelation in Week 7, basically it measures the serial correlation in a time series



Additive Holt-Winters smoothing

Choice of initial values

How should we set the initial values l_0, b_0, S_0, S_{-1}, ..., S_{2-M}, S_{1-M}?
Suggested Method

1. Do a linear least square regression over the data $y_1, \ldots, y_T$ to find out
$$\widehat{y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 t$$
2. Take $l_0 = \widehat{\beta}_0$ and $b_0 = \widehat{\beta}_1$
3. Find out $\widehat{s}_t = y_t - \widehat{y}_t$, then take the average of $\widehat{s}_t$ as one of S_0, S_{-1}, ..., S_{2-M}, S_{1-M} according to each season.

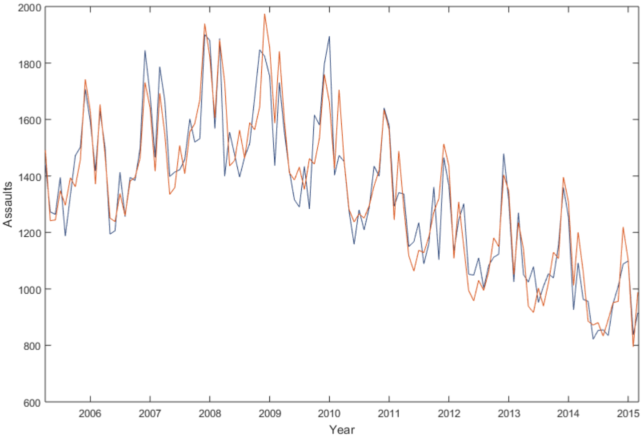
Additive Holt-Winters smoothing

Some notes

- Useful when seasonal variation is not changing much along the trend
- Choice of initial seasonal indices can be important.

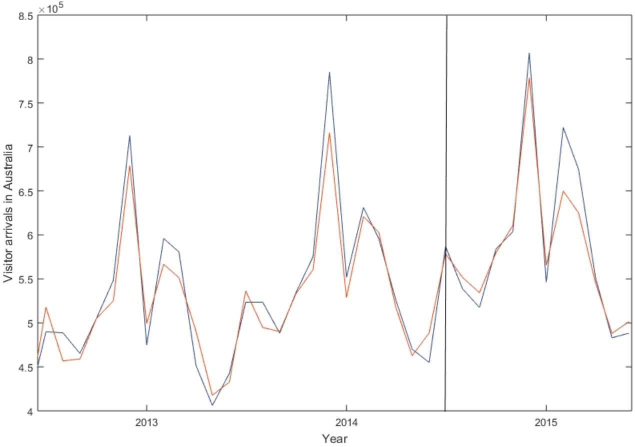
Alcohol related assaults in NSW

Additive Holt-Winters fit



Visitor arrivals in Australia

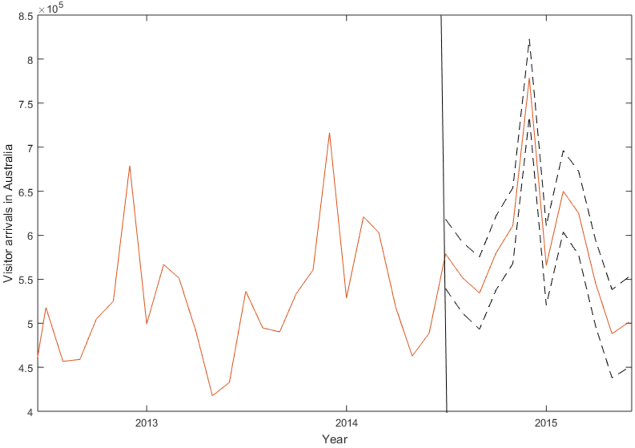
Additive Holt-Winters point forecast



Visitor arrivals in Australia

Additive Holt-Winters forecast intervals

We need a statistical model in order to construct forecast intervals



Additive Holt-Winters smoothing

Model

$$\begin{aligned} l_t &= \alpha(y_t - S_{t-M}) + (1-\alpha)(l_{t-1} + b_{t-1}), \\ b_t &= \gamma(l_t - l_{t-1}) + (1-\gamma)b_{t-1}, \\ S_t &= \delta(y_t - l_t) + (1-\delta)S_{t-M}, \end{aligned}$$

$$y_{t+1} = l_t + b_t + S_{t+1-M} + \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim N(0, \sigma^2).$$

We can chose the parameters α, γ and δ by minimising

$$SSE = \sum_{t=1}^{T} (y_t - l_{t-1} - b_{t-1} - S_{t-M})^2$$

Additive Holt-Winters smoothing

Error correction formulation

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

Additive Holt-Winters smoothing

Error correction formulation

From

$$l_t = l_{t-1} + b_{t-1} + \alpha(y_t - S_{t-M} - l_{t-1} - b_{t-1}),$$

we have that

$$l_t - l_{t-1} - b_{t-1} = \alpha(y_t - S_{t-M} - l_{t-1} - b_{t-1}).$$

Hence,

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

Additive Holt-Winters smoothing

Error correction formulation

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

Hence

$$
\begin{aligned}
S_t &= \delta(y_t - l_t) + (1-\delta)S_{t-M} \\
&= S_{t-M} + \delta(y_t - l_t - S_{t-M}) \\
&= S_{t-M} + \delta(1-\alpha)\varepsilon_t
\end{aligned}
$$

Additive Holt-Winters smoothing

Error correction formulation

$$
\begin{aligned}
l_t &= l_{t-1} + b_{t-1} + \alpha\varepsilon_t \\
b_t &= b_{t-1} + \alpha\gamma\varepsilon_t \\
S_t &= S_{t-M} + \delta(1-\alpha)\varepsilon_t \\
y_t &= l_{t-1} + b_{t-1} + S_{t-M} + \varepsilon_t \\
\epsilon_t &\sim N(0, \sigma^2).
\end{aligned}
$$

Additive Holt-Winters smoothing

Forecasting equations

$$\begin{aligned}
\widehat{y}_{t+1} &= \mathbb{E}(l_t + b_t + S_{t-M+1} + \varepsilon_{t+1} | y_{1:t}) \\
&= l_t + b_t + S_{t-M+1}
\end{aligned}$$

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

$$\vdots$$

$$\widehat{y}_{t+h} = l_t + hb_t + S_{t-M+(h \bmod M)}$$

What is $h \bmod M$?

Additive Holt-Winters smoothing

Variance for interval forecasts

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

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

Additive Holt-Winters smoothing

Variance for interval forecasts

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

$$
\begin{aligned}
\mathrm{Var}(y_{t+h}|y_{1:t}) &= \mathrm{Var}\left(l_t + hb_t + S_{t-M+h} + \alpha \sum_{i=1}^{h-1} (1+i\gamma)\varepsilon_{t+i} + \varepsilon_{t+h}|y_{1:t} \right) \\
&= \sigma^2 \left(1 + \alpha^2 \sum_{i=1}^{h-1} (1+i\gamma)^2 \right), \quad \text{for } h \leq M \text{ only.}
\end{aligned}
$$

Additive Holt-Winters smoothing*

Variance for interval forecasts

For $h > M$,

$$
\begin{aligned}
\mathrm{Var}(y_{t+h}|y_{1:t}) &= \mathrm{Var}\left(l_t + hb_t + S_{t-M+h} + \alpha \sum_{i=1}^{h-1} (1 + i\gamma)\varepsilon_{t+i} + \varepsilon_{t+h} | y_{1:t} \right) \\
&= \mathrm{Var}\Bigg(l_t + hb_t + S_{t-2M+h} + \delta(1-\alpha)\varepsilon_{t-M+h} \\
&\quad + \alpha \sum_{i=1}^{h-1} (1 + i\gamma)\varepsilon_{t+i} + \varepsilon_{t+h} | y_{1:t} \Bigg) \\
&= \sigma^2 \left(1 + \sum_{i=1}^{h-1} \left[\alpha(1 + i\gamma) + I_{i,M}\delta(1-\alpha) \right]^2 \right),
\end{aligned}
$$

where $I_{i,M} = 1$ if $h - i$ is an integer multiple of M, and 0 otherwise.

Additive Holt-Winters smoothing

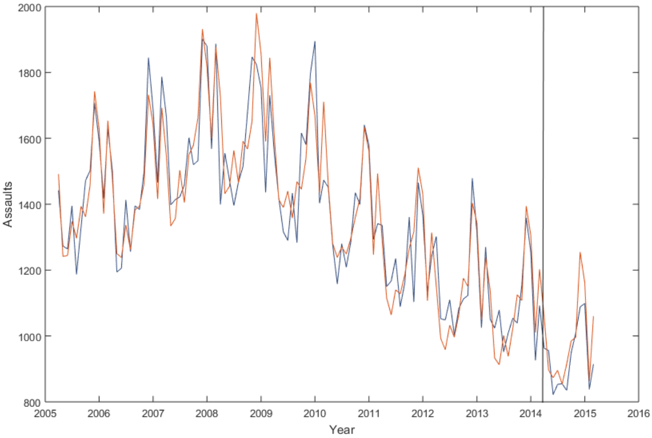
Forecasting: collecting the results

$$\widehat{y}_{t+h} = l_t + hb_t + S_{t-M+(h \bmod M)}.$$

$$\mathrm{Var}(y_{t+h}|y_{1:t}) = \sigma^2 \left(1 + \sum_{i=1}^{h-1} \left[\alpha(1+i\gamma) + I_{i,M}\delta(1-\alpha) \right]^2 \right).$$

Alcohol related assaults in NSW

Additive Holt-Winters forecast



Alcohol related assaults in NSW

Additive Holt-Winters forecast

Multiplicative Holt-Winters smoothing

Most useful when the seasonal pattern changes in a strong pattern and is proportional to the level of the series.

Multiplicative Holt-Winters smoothing

Model

$$
\begin{aligned}
l_t &= \alpha(y_t/S_{t-M}) + (1-\alpha)(l_{t-1} + b_{t-1}),\\
b_t &= \gamma(l_t - l_{t-1}) + (1-\gamma)b_{t-1},\\
S_t &= \delta(y_t/l_t) + (1-\delta)S_{t-M},
\end{aligned}
$$

$$
y_{t+1} = (l_t + b_t) \times S_{t+1-M} + \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim N(0, \sigma^2).
$$

We can chose the parameters α, γ and δ by minimising

$$
SSE = \sum_{t=1}^{n} (y_t - (l_{t-1} + b_{t-1})S_{t-M})^2
$$

Multiplicative Holt-Winters smoothing

Error correction formulation

$$
\begin{aligned}
l_t &= \alpha(y_t/S_{t-M}) + (1-\alpha)(l_{t-1} + b_{t-1}) \\
&= l_{t-1} + b_{t-1} + \alpha(y_t/S_{t-M} - l_{t-1} - b_{t-1}) \\
&= l_{t-1} + b_{t-1} + \alpha\left(\frac{y_t - (l_{t-1} + b_{t-1})S_{t-M}}{S_{t-M}}\right) \\
&= l_{t-1} + b_{t-1} + \alpha\frac{\varepsilon_t}{S_{t-M}}
\end{aligned}
$$

Multiplicative Holt-Winters smoothing

Error correction formulation

$$\begin{aligned} b_t &= \gamma(l_t - l_{t-1}) + (1-\gamma)b_{t-1} \\ &= b_{t-1} + \gamma\alpha\left(\frac{y_t - (l_{t-1} + b_{t-1})S_{t-M}}{S_{t-M}}\right) \quad \text{see previous slide} \\ &= b_{t-1} + \alpha\gamma\frac{\varepsilon_t}{S_{t-M}} \end{aligned}$$

Multiplicative Holt-Winters smoothing

Error correction formulation

$$S_t = \delta(y_t/l_t) + (1-\delta)S_{t-M} = S_{t-M} + \delta\frac{y_t - l_t S_{t-M}}{l_t}$$

From the derivation for l_t we have

$$l_t S_{t-M} = (l_{t-1} + b_{t-1})S_{t-M} + \alpha(y_t - (l_{t-1} + b_{t-1})S_{t-M})$$

Hence

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

Hence

$$S_t = S_{t-M} + \delta(1-\alpha)\frac{\varepsilon_t}{l_t}$$

Multiplicative Holt-Winters smoothing

Error correction formulation

$$
\begin{aligned}
l_t &= l_{t-1} + b_{t-1} + \alpha \frac{\varepsilon_t}{S_{t-M}} \\
b_t &= b_{t-1} + \alpha\gamma \frac{\varepsilon_t}{S_{t-M}} \\
S_t &= S_{t-M} + \delta(1-\alpha)\frac{\varepsilon_t}{l_t} \\
y_t &= (l_{t-1} + b_{t-1}) \times S_{t-M} + \varepsilon_t
\end{aligned}
$$

Multiplicative Holt-Winters smoothing

Forecasting equations

$$
\begin{aligned}
\widehat{y}_{t+1} &= \mathbb{E}((l_t + b_t)S_{t-M+1} + \varepsilon_{t+1} | y_{1:t}) \\
&= (l_t + b_t)S_{t-M+1}
\end{aligned}
$$

$$
\begin{aligned}
\widehat{y}_{t+2} &= \mathbb{E}((l_{t+1} + b_{t+1})S_{t-M+2} + \varepsilon_{t+2} | y_{1:t}) \\
&= \mathbb{E}\left(\left[l_t + 2b_t + \alpha(1+\gamma)\frac{\varepsilon_{t+1}}{S_{t-M+1}} \right] S_{t-M+2} + \varepsilon_{t+2} \;\middle|\; y_{1:t} \right) \\
&= (l_t + 2b_t)S_{t-M+2} \\
&\vdots \\
\widehat{y}_{t+h} &= (l_t + hb_t)S_{t-M+(h \bmod M)}
\end{aligned}
$$

Multiplicative Holt-Winters smoothing

Variance for interval forecasts

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

$$
\begin{aligned}
\mathrm{Var}(y_{t+2}|y_{1:t}) &= \mathrm{Var}((l_{t+1} + b_{t+1})S_{t-M+2} + \varepsilon_{t+2}|y_{1:t}) \\
&= \mathrm{Var}\left(\left[l_t + 2b_t + \alpha(1+\gamma)\frac{\varepsilon_{t+1}}{S_{t-M+1}}\right] S_{t-M+2} + \varepsilon_{t+2}|y_{1:t}\right) \\
&= \sigma^2(1 + \alpha^2(1+\gamma)^2(S_{t-M+2}^2/S_{t-M+1}^2))
\end{aligned}
$$

Outline

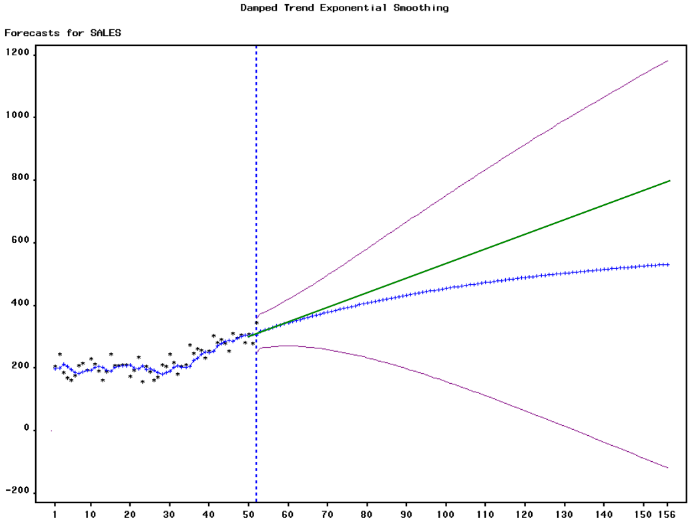
Dampened trend ES

Extrapolating trends indefinitely into the future can be problematic.

Dampened trend exponential smoothing aims to deal with this problem.

Dampened trend ES

Illustration



Dampened trend ES

Model

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

$$
y_{t+1} = l_t + \phi b_t + \varepsilon_{t+1},
$$

where ϕ is the dampening factor, with $0 \leq \phi \leq 1$.

Homework: put the model into the error correction form.

Dampened trend ES

Forecasting and variance equations

$$y_{t+1} = l_t + \phi b_t + \varepsilon_{t+1}$$

$$\widehat{y}_{t+1} = l_t + \phi b_t$$
$$\mathrm{Var}(y_{t+1} | y_{1:t}) = \sigma^2$$

Dampened trend ES

Forecasting and variance equations

$$\begin{aligned} y_{t+2} &= l_{t+1} + \phi b_{t+1} + \varepsilon_{t+2} \\ &= l_t + \phi b_t + \phi^2 b_t + \alpha(1+\phi\gamma)\varepsilon_{t+1} + \varepsilon_{t+2} \end{aligned}$$

$$\begin{aligned} \widehat{y}_{t+2} &= l_t + b_t(\phi + \phi^2) \\ \mathrm{Var}(y_{t+1}|y_{1:t}) &= \sigma^2(1 + \alpha^2(1+\phi\gamma)^2) \end{aligned}$$

Dampened trend ES

Forecasting formula

$$\widehat{y}_{t+h} = l_t + \phi b_t + \phi^2 b_t + \phi^3 b_t + \ldots + \phi^h b_t$$

Compared with the forecast of the trend correct exponential method

$$\widehat{y}_{t+h} = l_t + h \times b_t$$

What happens as h gets larger?
For the dampened forecast $\widehat{y}_{t+h} \rightarrow l_t + \frac{\phi}{1-\phi} b_t$
For the trend corrected forecast

$$\widehat{y}_{t+h} \rightarrow \infty$$

Dampened trend seasonal

Model

$$
\begin{aligned}
l_t &= \alpha(y_t - S_{t-M}) + (1-\alpha)(l_{t-1} + \phi b_{t-1}),\\
b_t &= \gamma(l_t - l_{t-1}) + (1-\gamma)\phi b_{t-1},\\
S_t &= \delta(y_t - l_t) + (1-\delta)S_{t-M},
\end{aligned}
$$

$$
y_{t+1} = l_t + \phi b_t + S_{t-M+1} + \varepsilon_{t+1},
$$

where ϕ is the dampening factor, with $0 \leq \phi \leq 1$.

Dampened trend seasonal

Forecasting formula

$$\widehat{y}_{t+h} = l_t + \phi b_t + \phi^2 b_t + \phi^3 b_t + \ldots + \phi^h b_t + S_{t+h-M}$$

Recap

We have looked at

- extension of SES and TCES to handle seasonality
- the resulting technique, Holt-Winters smoothing, for additive and multiplicative seasonalities
- how to dampen the trend forecasts

Next lecture: Autoregressive integrated moving average (ARIMA)