Predictive Analytics

Week 12: Exponential Smoothing

Semester 2, 2018

Discipline of Business Analytics, The University of Sydney Business School

Week 12: Exponential Smoothing

1. Simple exponential smoothing

2. Trend corrected exponential smoothing

3. Holt winters smoothing

4. Damped trend exponential smoothing

Exponential smoothing forecasts are weighted averages of past observations, where the weights decay exponentially as we go further into the past.

Exponential smoothing can be useful when the time series components are changing over time.

Simple exponential smoothing

Simple exponential smoothing (keyboard)

The **simple exponential smoothing** method specifies the forecasting rule

$$\widehat{y}_{t+1} = \ell_t \qquad \text{(forecast equation)}$$

$$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1} \qquad \text{(smoothing equation)}$$

for an initial value ℓ_0 and $0 \leq \alpha \leq 1$.

Exponentially weighted moving average

$$\ell_1 = \alpha y_1 + (1-\alpha)\ell_0$$

$$\begin{aligned}\ell_2 &= \alpha y_2 + (1-\alpha)\ell_1 \\ &= \alpha y_2 + (1-\alpha)\alpha y_1 + (1-\alpha)^2\ell_0\end{aligned}$$

$$\begin{aligned}\ell_3 &= \alpha y_3 + (1-\alpha)\ell_2 \\ &= \alpha y_3 + (1-\alpha)\alpha y_2 + (1-\alpha)^2\alpha y_1 + (1-\alpha)^3\ell_0\end{aligned}$$

$$\begin{aligned}\ell_4 &= \alpha y_4 + (1-\alpha)\ell_3 \\ &= \alpha y_4 + (1-\alpha)\alpha y_3 + (1-\alpha)^2\alpha y_2 + (1-\alpha)^3\alpha y_1 + (1-\alpha)^4\ell_0 \\ &\vdots\end{aligned}$$

Exponentially weighted moving average

It follows that

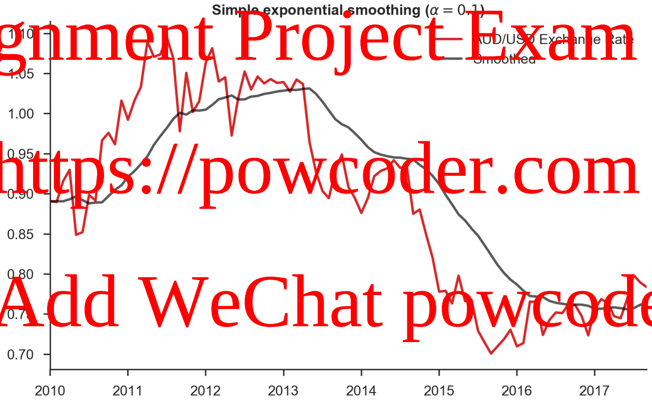
$$\begin{aligned}\ell_t =& \alpha y_t + (1-\alpha)\ell_{t-1} \\ =& \alpha y_t + (1-\alpha)\alpha y_{t-1} + (1-\alpha)^2 \alpha y_{t-2} + \ldots + (1-\alpha)^{t-1}\alpha y_1 \\ &+ (1-\alpha)^t \ell_0.\end{aligned}$$

Simple exponential smoothing is also known as the **exponentially weighted moving average** (EWMA) method.

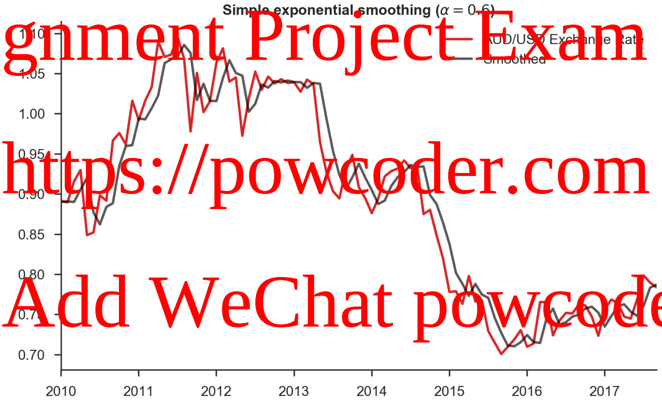
Simple exponential smoothing

- Useful for forecasting time series with changing levels.
- A higher α gives larger weight to recent observations, making the forecasts more adaptive to recent changes in the series.
- A lower α leads to a larger weights for past observations, making the forecasts smoother.
- Initialisation: we typically set $\ell_0 = y_1$ for simplicity. Alternatively, we can treat it as a parameter.

Example: AUD/USD exchange rate



Example: AUD/USD exchange rate



Estimation

We estimate α by least squares (empirical risk minimisation).

$$\widehat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \sum_{t=1}^{N} (y_t - \ell_{t-1})^2$$

Each ℓ_t is a nonlinear function of α, so that there is no formula for $\widehat{\alpha}$. We use numerical optimisation methods to obtain the solution.

Statistical model

In order to say more about the simple exponential smoothing method, we need to formulate it as a statistical model. We assume that

$$Y_t = \ell_{t-1} + \varepsilon_t,$$

$$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1},$$

where the errors ε_t are i.i.d with constant variance σ^2.

Statistical model

In forecasting, we want to:

1. To compute point forecasts for multiple forecasting horizons h.

2. To compute interval forecasts for multiple forecasting horizons h.

In order to this for the exponential smoothing method, we rewrite the model in **error correction form**.

Error correction form

We obtain the error correction form as

$$
\begin{aligned}
\ell_t &= \alpha Y_t + (1-\alpha)\ell_{t-1} \\
&= \ell_{t-1} + \alpha(Y_t - \ell_{t-1}) \\
&= \ell_{t-1} + \alpha\varepsilon_t.
\end{aligned}
$$

Hence, we can rewrite the model as:

$$
\begin{aligned}
Y_{t+1} &= \ell_t + \varepsilon_{t+1}, \\
\ell_t &= \ell_{t-1} + \alpha\varepsilon_t.
\end{aligned}
$$

Error correction form

Using $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$,

$$\ell_{t+1} = \ell_t + \alpha\varepsilon_{t+1}$$

$$\begin{aligned}\ell_{t+2} &= \ell_{t+1} + \alpha\varepsilon_{t+2} \\ &= \ell_t + \alpha\varepsilon_{t+1} + \alpha\varepsilon_{t+2}\end{aligned}$$

$$\begin{aligned}\ell_{t+3} &= \ell_{t+2} + \alpha\varepsilon_{t+3} \\ &= \ell_t + \alpha\varepsilon_{t+1} + \alpha\varepsilon_{t+2} + \alpha\varepsilon_{t+3}\end{aligned}$$

$$\vdots$$

$$\ell_{t+h} = \ell_t + \sum_{i=1}^{h} \alpha\varepsilon_{t+i}$$

Constant plus noise representation

Using $Y_t = \ell_{t-1} + \varepsilon_t$ and the previous slide,

$$
\begin{aligned}
Y_{t+1} &= \ell_t + \varepsilon_{t+1} \\
Y_{t+2} &= \ell_{t+1} + \varepsilon_{t+2} \\
&= \ell_t + \alpha\varepsilon_{t+1} + \varepsilon_{t+2} \\
Y_{t+3} &= \ell_{t+2} + \varepsilon_{t+3} \\
&= \ell_t + \alpha\varepsilon_{t+1} + \alpha\varepsilon_{t+2} + \varepsilon_{t+3} \\
&\vdots \\
Y_{t+h} &= \ell_{t+h-1} + \varepsilon_{t+h} \\
&= \ell_t + \sum_{i=1}^{h-1} \alpha\varepsilon_{t+i} + \varepsilon_{t+h}
\end{aligned}
$$

Point forecast

Constant plus noise representation of future observations:

$$Y_{t+h} = \ell_t + \sum_{i=1}^{h-1} \alpha\varepsilon_{t+i} + \varepsilon_{t+h}$$

From the linearity of expectations, the point forecast for any horizon h is

$$\begin{aligned}\widehat{y}_{t+h} &= E(Y_{t+h}|y_{1:t}) \\ &= E\left(\ell_t + \sum_{i=1}^{h-1} \alpha\varepsilon_{t+i} + \varepsilon_{t+h} \,\middle|\, y_{1:t}\right) \\ &= \ell_t\end{aligned}$$

Forecast variance

$$\begin{aligned}\mathsf{Var}(Y_{t+1}|y_{1:t}) &= \mathsf{Var}(\ell_t + \varepsilon_{t+1}|y_{1:t}) \\ &= \sigma^2\end{aligned}$$

$$\begin{aligned}\mathsf{Var}(Y_{t+2}|y_{1:t}) &= \mathsf{Var}(\ell_t + \alpha\varepsilon_{t+1} + \varepsilon_{t+2}|y_{1:t}) \\ &= \sigma^2(1+\alpha^2)\end{aligned}$$

$$\vdots$$

$$\begin{aligned}\mathsf{Var}(Y_{t+h}|y_{1:t}) &= \mathsf{Var}\left(\ell_t + \sum_{i=1}^{h-1}\alpha\varepsilon_{t+h-i} + \varepsilon_{t+h} \,\middle|\, y_{1:t}\right) \\ &= \sigma^2(1+(h-1)\alpha^2)\end{aligned}$$

Forecast equations for simple exponential smoothing

$$\hat{y}_{t+h|t} = \ell_t$$

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

Interval forecast

If we assume that $\varepsilon_t \sim N(0, \sigma^2)$,

$$Y_{t+h}|y_{1:t} \sim N\Big(\ell_t, \sigma^2\left[1 + (h-1)\alpha^2\right]\Big).$$

To compute an interval forecast, we use the estimated values of α and σ^2

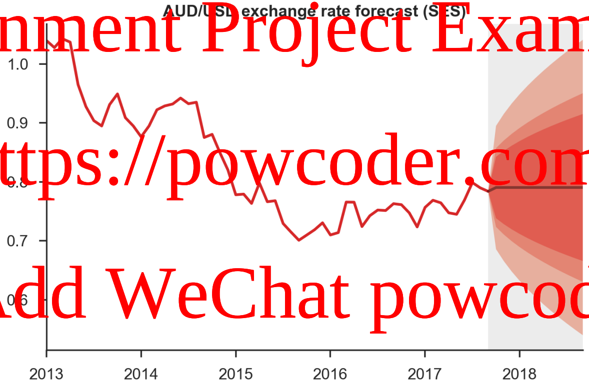
$$\widehat{\ell}_t \pm z_{\text{crit}} \times \sqrt{\widehat{\sigma}^2\left[1 + (h-1)\widehat{\alpha}^2\right]},$$

where

$$\widehat{\sigma}^2 = \frac{\sum_{t=1}^{n}(y_t - \ell_{t-1})^2}{N-1}.$$

If the errors are not normal, you should use the Bootstrap method or other distributional assumptions.

Example: AUD/USD exchange rate



Trend corrected exponential smoothing

The **trend corrected** or **Holt exponential smoothing** method allows for a time-varying trend:

$$\begin{aligned}
\widehat{y}_{t+1} &= \ell_t + b_t && \text{(forecast equation)}\\
\ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) && \text{(smoothing equation)}\\
b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta) b_{t-1} && \text{(trend equation)}
\end{aligned}$$

for an initial values ℓ_0 and b_0, $0 \leq \alpha \leq 1$, and $0 \leq \beta \leq 1$.

Trend corrected exponential smoothing

Consider the simple time series trend model

$$\ell_t = a + b \times t$$

$$Y_t = \ell_t + \varepsilon_t.$$

What is $\ell_t - \ell_{t-1}$ here?

Trend corrected exponential smoothing model

The statistical model is

$$Y_{t+1} = \ell_t + b_t + \varepsilon_{t+1},$$
$$\ell_t = \alpha Y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}),$$
$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta) b_{t-1},$$

where the errors ε_t are i.i.d with constant variance σ^2.

The least squares estimates of α and β are

$$\widehat{\alpha}, \widehat{\beta} = \underset{\alpha,\beta}{\text{argmin}} \sum_{t=1}^{N} (y_t - \ell_{t-1} - b_{t-1})^2$$

Error correction form

$$
\begin{aligned}
\ell_t &= \alpha Y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
&= \ell_{t-1} + b_{t-1} + \alpha(Y_t - \ell_{t-1} - b_{t-1}) \\
&= \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t
\end{aligned}
$$

$$
\begin{aligned}
b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \\
&= b_{t-1} + \beta(\ell_t - \ell_{t-1} - b_{t-1}) \\
&= b_{t-1} + \beta\alpha(\ell_{t-1} + b_{t-1} + \alpha\varepsilon_t - \ell_{t-1} - b_{t-1}) \\
&= b_{t-1} + \beta\alpha\varepsilon_t
\end{aligned}
$$

Error correction form

$$Y_{t+1} = \ell_t + b_t + \varepsilon_{t+1}$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$$

$$b_t = b_{t-1} + \beta\alpha\varepsilon_t$$

Constant plus noise representation

$$
\begin{aligned}
Y_{t+1} &= \ell_t + b_t + \varepsilon_{t+1} \\
Y_{t+2} &= \ell_{t+1} + b_{t+1} + \varepsilon_{t+2} \\
&= \ell_t + 2b_t + \alpha(1+\beta)\varepsilon_{t+1} + \varepsilon_{t+2} \\
Y_{t+3} &= \ell_{t+2} + b_{t+2} + +\varepsilon_{t+3} \\
&= \ell_{t+1} + 2b_{t+1} + \alpha(1+\beta)\varepsilon_{t+2} + \varepsilon_{t+3} \\
&= \ell_t + 3b_t + \alpha(1+2\beta)\varepsilon_{t+1} + \alpha(1+\beta)\varepsilon_{t+2} + \varepsilon_{t+3} \\
&\vdots \\
Y_{t+h} &= \ell_t + hb_t + \alpha\sum_{i=1}^{h-1}(1+i\beta)\varepsilon_{t+h-i} + \varepsilon_{t+h}
\end{aligned}
$$

Point forecast

Constant plus noise representation of future observations:

$$Y_{t+h} = \ell_t + hb_t + \alpha \sum_{i=1}^{h-1} (1 + i\beta)\varepsilon_{t+h-i} + \varepsilon_{t+h}$$

From the linearity of expectations, the point forecast for any horizon h is

$$\begin{aligned}\widehat{y}_{t+h} &= E(Y_{t+h}|y_{1:t}) \\ &= E\left(\ell_t + hb_t + \alpha \sum_{i=1}^{h-1} (1 + i\beta)\varepsilon_{t+h-i} + \varepsilon_{t+h} \,\middle|\, y_{1:t}\right) \\ &= \ell_t + hb_t.\end{aligned}$$

Forecast variance

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

$$
\begin{aligned}
\mathsf{Var}(Y_{t+2}|y_{1:t}) &= \mathsf{Var}(\ell_t + 2b_t + \alpha(1+\beta)\varepsilon_{t+1} + \varepsilon_{t+2}|y_{1:t}) \\
&= \sigma^2(1 + \alpha^2(1+\beta)^2) \\
&\vdots
\end{aligned}
$$

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

Forecast equations for the trend corrected smoothing method

Point forecast:

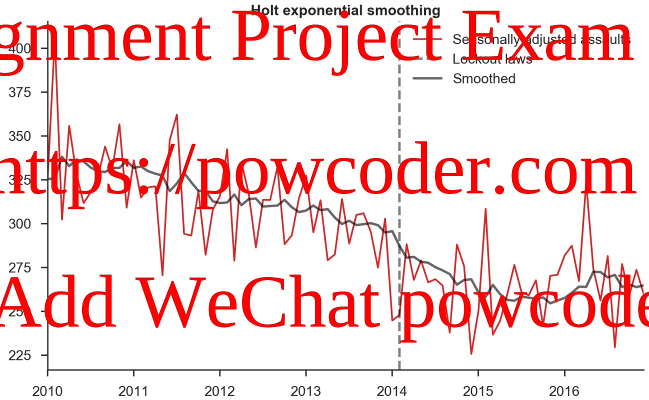
$$\widehat{y}_{t+h} = \widehat{\ell}_t + h\widehat{b}_t$$

Variance:

$$\mathsf{Var}(Y_{t+h}|y_{1:t}) = \sigma^2 \left(1 + \alpha^2 \sum_{i=1}^{h-1} (1 + i\beta)^2\right)$$

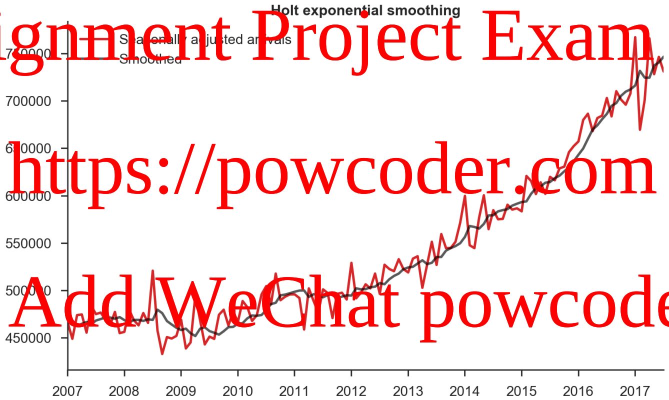
We compute interval forecasts as before.

Example: assaults in Sydney

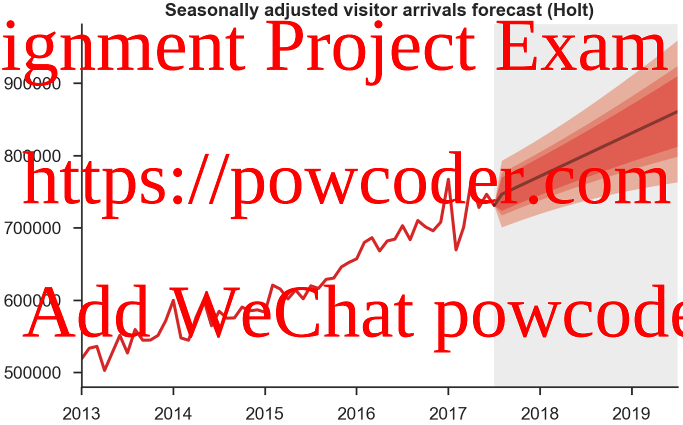


Example: assaults in Sydney

Example: visitor arrivals



Example: visitor arrivals



Holt winters smoothing

Holt Winters exponential smoothing

The **Holt-Winters** exponential smoothing method extend the trend corrected method to seasonal data. It allows for additive or multiplicative seasonality.

Additive Holt Winters Smoothing (key concept)

$$\widehat{y}_{t+1} = \ell_t + b_t + S_{t+1-L} \qquad \text{(forecast equation)}$$

$$\ell_t = \alpha(y_t - S_{t-L}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \qquad \text{(level)}$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1}, \qquad \text{(trend)}$$

$$S_t = \delta(y_t - \ell_t) + (1-\delta)S_{t-L}, \qquad \text{(seasonal indices)}$$

for a seasonal frequency L, initial values ℓ_0, b_0, and S_{i-L} for $i = 1, \ldots, L$, and parameters $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$, $0 \leq \delta \leq 1$.

Multiplicative Holt Winters Smoothing (key concept)

$$\widehat{y}_{t+1} = (\ell_t + b_t) \times S_{t+1-L} \qquad \text{(forecast equation)}$$

$$\ell_t = \alpha(y_t/S_{t-L}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \qquad \text{(level)}$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1}, \qquad \text{(trend)}$$

$$S_t = \delta(y_t/\ell_t) + (1-\delta)S_{t-L}, \qquad \text{(seasonal indices)}$$

for a seasonal frequency L, initial values ℓ_0, b_0, and S_{i-L} for $i = 1, \ldots, L$, and parameters $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$, $0 \leq \delta \leq 1$.

Statistical model

As before, we formulate a statistical model by specifying an observation equation.

Additive:

$$Y_{t+1} = \ell_t + b_t + S_{t+1-L} + \varepsilon_{t+1},$$

where ε_{t+1} is i.i.d with variance σ^2

Multiplicative:

$$y_{t+1} = (\ell_t + b_t) \times S_{t+1-L} + \varepsilon_{t+1},$$

where ε_{t+1} is i.i.d with variance σ^2.

We estimate α, β and δ by least squares.

Additive:

$$\widehat{\alpha}, \widehat{\beta}, \widehat{\delta} = \underset{\alpha,\beta,\delta}{\operatorname{argmin}} \sum_{t=1}^{N} (y_t - \ell_{t-1} - b_{t-1} - S_{t+1-L})^2$$

Multiplicative:

$$\widehat{\alpha}, \widehat{\beta}, \widehat{\delta} = \underset{\alpha,\beta,\delta}{\operatorname{argmin}} \sum_{t=1}^{N} (y_t - (\ell_t + b_t) \times S_{t+1-L})^2$$

Forecast equations

Additive:

$$\widehat{y}_{t+h} = \widehat{\ell}_t + h\widehat{b}_t + S_{t-L+(h \bmod L)}$$

$$\text{Var}(Y_{t+h}|y_{1:t}) = \sigma^2\left(1 + \sum_{i=1}^{h-1}\left[\alpha(1+i\beta) + I_{i,L}\delta(1-\alpha)\right]^2\right),$$

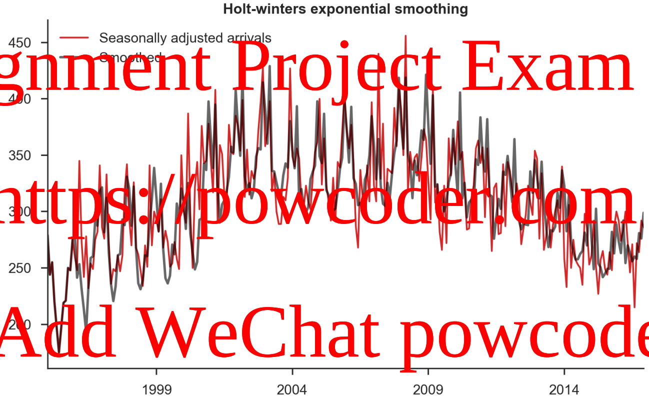
where mod is the modulo operator, $I_{i,L} = 0$ if $h \bmod L \neq i$ and $I_{i,L} = 1$ if $h \bmod L = i$.

Multiplicative:

$$\widehat{y}_{t+h} = (\widehat{\ell}_t + h\widehat{b}_t) \times S_{t-L+(h \bmod L)}$$

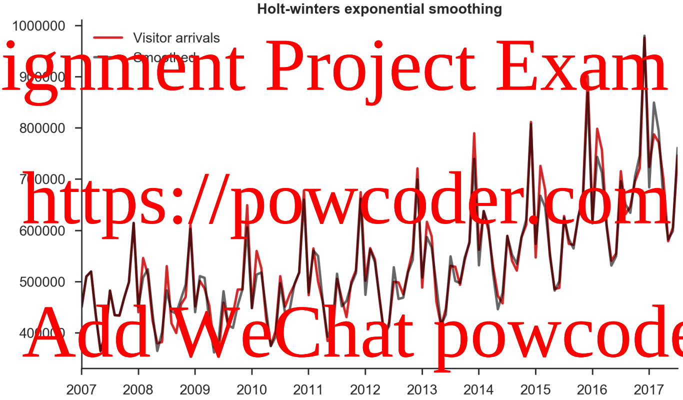
No simple expression exists for the variance in the multiplicative model.

Example: assaults in Sydney



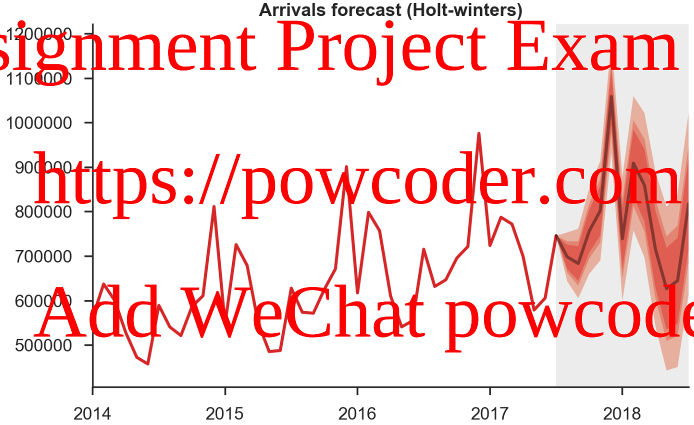
The estimated parameters are $\widehat{\alpha} = 0.117$, $\widehat{\beta} = 0.023$, and $\widehat{\delta} = 0.370$.

Example: visitor arrivals



The estimated parameters are $\widehat{\alpha} = 0.154$, $\widehat{\beta} = 0.088$, and $\widehat{\delta} = 0.271$.

Example: visitor arrivals



Damped trend exponential smoothing

Damped trend exponential smoothing

Damped trend exponential smoothing addresses the problem that extrapolating trends indefinitely into the future can lead to implausible forecasts.

Model and forecast

Model:

$$
\begin{aligned}
y_{t+1} &= \ell_t + \phi b_t + \varepsilon_{t+1},\\
\ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}),\\
b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)\phi b_{t-1},
\end{aligned}
$$

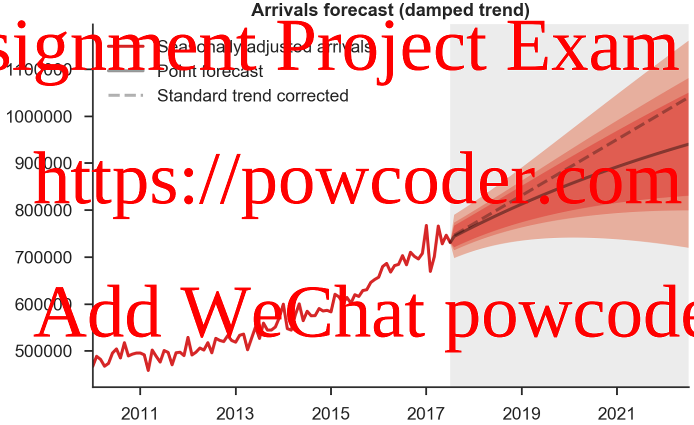
where ϕ is the damping parameter, with $0 \leq \phi \leq 1$.

Forecast equation:

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

We can extend it to allow for additive or multiplicative seasonality.

Illustration: visitor arrivals



Review questions

- What is exponential smoothing?
- What is the difference between simple, trend corrected, and Holt-Winters exponential smoothing methods?

- Derive the point forecasts and forecast variances for the SES and trend corrected methods, starting from the model equations.

- Explain how to compute forecast intervals based on the SES and trend corrected methods.