Time-series Forecasting - Exponential Smoothing

Dr. Josif Grabocka

ISMLL, University of Hildesheim

Business Analytics

Motivation

Extreme *recentivism*: The **Naive** forecast considers only the last value:

$$\hat{y}_{T+h|T} = y_T$$

Extreme *nostalgia*: The **Average** method equally considers all previous values:

$$\hat{y}_{T+h|T} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

A middle way is needed: How to capture the influence of previous values while prioritizing recent measurements?

Exponential Smoothing

An exponentially-weighted average for $0 < \alpha \leq 1$:

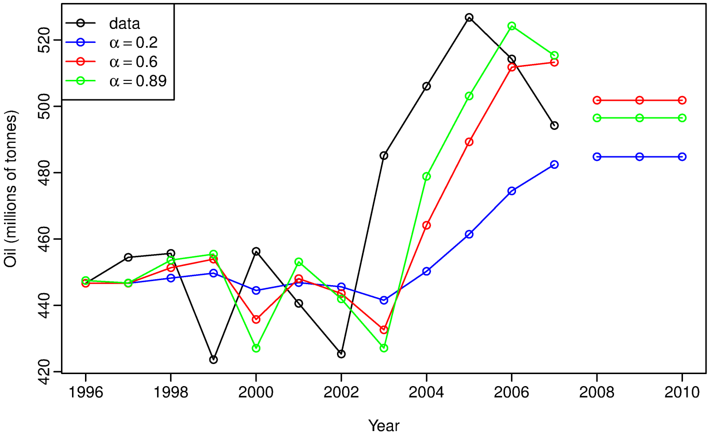
$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha) y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \dots$$

The weights decrease exponentially for past measurements:

Observation	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$
y_T	0.2	0.4	0.6	0.8
y_{T-1}	0.16	0.24	0.24	0.16
y_{T-2}	0.128	0.144	0.096	0.032
y_{T-3}	0.1024	0.0864	0.0384	0.0064
y_{T-4}	0.08192	0.05184	0.01536	0.00128

The recentivism is controlled by α, where $\sum_{t=1}^{T} \alpha(1-\alpha)^{T-t} \approx 1, \forall \alpha \in (0,1]$

Example - Oil Production



Oil production in Saudi Arabia (1996–2007). Hyndman et al. 2014

Weighted Average Form

Forecast of time $T+1$ is a weighted average of observation y_T and forecast $\hat{y}_{T|T-1}$:

$$
\begin{aligned}
\hat{y}_{T+1|T} &= \alpha y_T + (1-\alpha)\hat{y}_{T|T-1} \\
\hat{y}_{T|T-1} &= \alpha y_{T-1} + (1-\alpha)\hat{y}_{T-1|T-2} \\
&\dots \\
\hat{y}_{3|2} &= \alpha y_2 + (1-\alpha)\hat{y}_{2|1} \\
\hat{y}_{2|1} &= \alpha y_1 + (1-\alpha)\ell_0
\end{aligned}
$$

Component and Error Correction Forms

The only component influencing forecast $\hat{y}_{t|t-1}$ is the level ℓ_t:

$$\text{Forecast Equation} \qquad \hat{y}_{t+1|t} = \ell_t$$
$$\text{Smoothing Equation} \qquad \ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$$

Similarly we can derive the error (e_t) correction form:

$$\begin{aligned} \ell_t &= \ell_{t-1} + \alpha(y_t - \ell_{t-1}) \\ &= \ell_{t-1} + \alpha(y_t - \hat{y}_{t|t-1}) \\ &= \ell_{t-1} + \alpha e_t \end{aligned}$$

Holt's Linear Trend Method

Tackle trend in $\hat{y}_{t+1|h}$ by decomposing out a slope/trend term b_t:

$$
\begin{aligned}
&\text{Forecast Equation} \quad & \hat{y}_{t+h|t} &= \ell_t + hb_t \\
&\text{Level Equation} & \ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
&\text{Trend Equation} & b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}
\end{aligned}
$$

- $0 < \alpha \leq 1$, $0 < \beta^* \leq 1$
- ℓ_t as a waited average of current observation y_t and estimation $\hat{y}_{t|t-1} = \ell_{t-1} + b_{t-1}$
- b_t as a weighted average of prediction differences $\ell_t - \ell_{t-1}$ and previous trend b_{t-1}

Holt's Linear Trend Method - Error Correction

- Rewrite by taking ℓ_t and b_t on the left side:

$$\begin{aligned} \ell_t &= \ell_{t-1} + b_{t-1} + \alpha e_t \\ b_t &= b_{t-1} + \alpha\beta^* e_t \end{aligned}$$

- Where the isolated error is:

$$e_t = y_t - (\ell_{t-1} + b_{t-1}) = y_t - \hat{y}_{t|t-1}$$

- Set $\ell_0 = y_1$ and $b_0 = y_2 - y_1$ when making predictions

Exercise

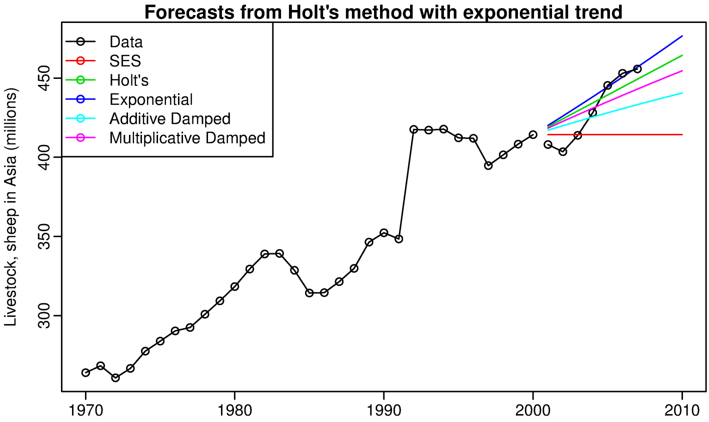
Given a series $y_1 = 2, y_2 = -3, y_3 = 3, y_4 = -2$, predict $\hat{y}_{5|4} =$? using

- Holt's Linear Trend with $\alpha = 0.5, \beta^* = 0.5$

In fact, it was later observed that $y_5 = 2$.

- $\ell_0 = 2, b_0 = -5$
- $\ell_1 = 0.5(2 + 2 - 5) = -0.5, b_1 = 0.5(-0.5 - 2 - 5) = -3.75$
- $\ell_2 = 0.5(-3 - 0.5 - 3.75) = -3.625, b_2 = 0.5(-3.625 + 0.5 - 3.75) = -3.4375$
- $\ell_3 = 0.5(3 - 3.625 - 3.4375) = -2.03125,$
  $b_3 = 0.5(-2.03125 + 3.625 - 3.4375) = -0.92188$
- $\ell_4 = 0.5(-2 - 2.03125 - 0.92188) = -2.47656,$
  $b_4 = 0.5(-2.47656 + 2.03125 - 0.92188) = -0.68359$
- $\hat{y}_{5|4} = -247656 - 0.68359 = -3.16016$

Example - Livestock, sheep in Asia



Comparison of different forecasting approaches using non-seasonal exponential smoothing. Hyndman et al. 2014

Exponential Trend Method

Replace an additive model (constant slope) with a multiplicative one (constant growth):

Forecast Equation	$\hat{y}_{t+h\|t} = \ell_t \, b_t^h$
Level Equation	$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} b_{t-1})$
Trend Equation	$b_t = \beta^* \frac{\ell_t}{\ell_{t-1}} + (1-\beta^*) b_{t-1}$

- b_t represents a multiplicative growth rate in relative terms $\frac{\ell_t}{\ell_{t-1}}$
- Forecasts project a constant growth rate rather than a constant slope (as in the additive model)

Exponential Trend Method - Error Correction

- Rewrite by taking ℓ_t and b_t on the left side:

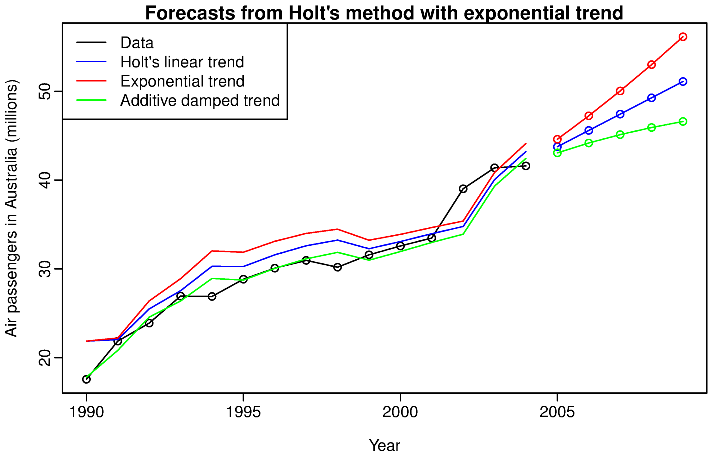
$$\begin{aligned} \ell_t &= \ell_{t-1} b_{t-1} + \alpha e_t \\ b_t &= b_{t-1} + \alpha \beta^* \frac{e_t}{\ell_{t-1}} \end{aligned}$$

- Where the isolated error is:

$$e_t = y_t - (\ell_{t-1}\, b_{t-1}) = y_t - \hat{y}_{t|t-1}$$

- Set $\ell_0 = y_1$ and $b_0 = y_2/y_1$ when making predictions

Example - Air Passengers (Hyndman et al. 2014)



Notice the over-forecast of the linear and exponential trend methods. The damped trend approach addresses the issue ...

Damped Additive Trend Model

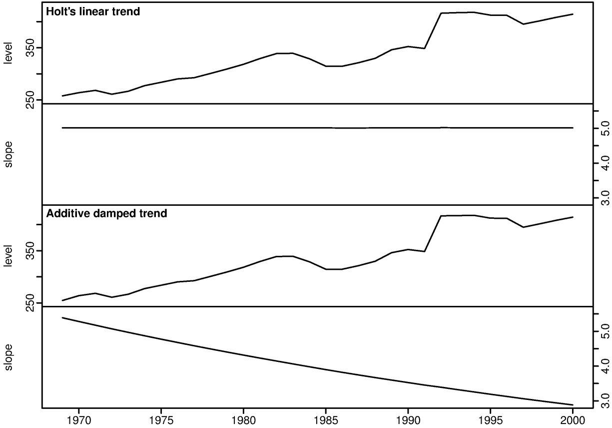
Include a damping/prohibitory parameter $0 < \phi < 1$:

Forecast Equation	$\hat{y}_{t+h\|t} = \ell_t + (\phi + \phi^2 + \cdots + \phi^h) b_t$
Level Equation	$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$
Trend Equation	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$

- $\phi = 1$ leads Holt's additive linear trend model
- Error correction form:

$$\begin{aligned} \ell_t &= \ell_{t-1} + \phi b_{t-1} + \alpha e_t \\ b_t &= \phi b_{t-1} + \alpha\beta^* e_t \end{aligned}$$

Example - Livestock, sheep in Asia



Level ℓ_t and trend b_t computed from a Livestock, sheep time series. Hyndman et al. 2014

Damped Multiplicative Trend Model

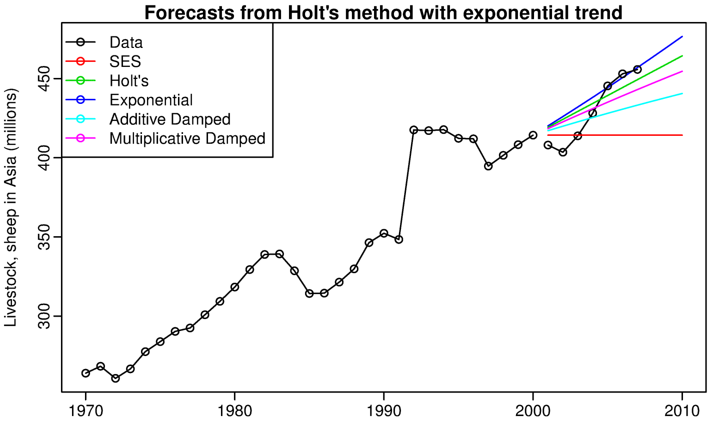
Include a damping/prohibitory parameter $0 < \phi < 1$ to the exponential trend:

Forecast Equation	$\hat{y}_{t+h\|t} = \ell_t \, b_t^{(\phi+\phi^2+\cdots+\phi^h)}$
Level Equation	$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} \, b_{t-1}^{\phi})$
Trend Equation	$b_t = \beta^* \dfrac{\ell_t}{\ell_{t-1}} + (1-\beta^*) b_{t-1}^{\phi}$

- $\phi = 1$ leads to exponential trend model
- Error correction form:

$$\begin{aligned} \ell_t &= \ell_{t-1} b_{t-1}^{\phi} + \alpha e_t \\ b_t &= b_{t-1}^{\phi} + \alpha\beta^* \frac{e_t}{\ell_{t-1}} \end{aligned}$$

Example - Livestock, sheep in Asia



Hyndman et al. 2014

Exercise (2)

Given a series $y_1 = 2, y_2 = -3, y_3 = 3, y_4 = -2$, predict $\hat{y}_{5|4} =$? using:

- Damped Additive Trend Model with $\alpha = 0.5, \beta^* = 0.5, \phi = 0.5$

In fact, it was later observed that $y_5 = 2$.

Solve it at home!

Setting parameters

- How to set $\alpha, \beta^*, b_0, \ell_0, \phi$?
- Search for combinations that minimize the Sum of Squared Errors:

$$SSE = \sum_{t=1}^{T} e_t^2 = \sum_{t=1}^{T} (y_t - \hat{y}_{t|t-1})^2$$

Holt-Winters Additive **Seasonal** Method

Forecast	$\hat{y}_{t+h\|t} = \ell_t + hb_t + s_{t-m+h_m^+}$
Level	$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$
Trend	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$
Seasonality	$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$

- $h_m^+ = \lfloor (h-1) \mod m \rfloor + 1$ ensures estimates come from final year
- ℓ_t is the weighted average between seasonally-adjusted observation and non-seasonal forecast
- b_t same as Holt's linear model
- s_t is the weighted average between current seasonal value and the last season's (m periods ago)

Holt-Winters Additive **Seasonal** Method (2)

With error correction forms:

$$\begin{aligned} \ell_t &= \ell_{t-1} + b_{t-1} + \alpha e_t \\ b_t &= b_{t-1} + \alpha\beta^* e_t \\ s_t &= s_{t-m} + \gamma e_t \end{aligned}$$

and residual errors:

$$e_t = y_t - (\ell_{t-1} + b_t + s_{t-m}) = y_t - \hat{y}_{t|t-1}$$

Holt-Winters Multiplicative **Seasonal** Method

Forecast	$\hat{y}_{t+h\|t} = (\ell_t + hb_t)s_{t-m+h_m^+}$
Level	$\ell_t = \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(\ell_{t-1} + b_{t-1})$
Trend	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$
Seasonality	$s_t = \gamma \frac{y_t}{\ell_{t-1} + b_{t-1}} + (1-\gamma)s_{t-m}$

Holt-Winters Multiplicative **Seasonal** Method (2)

With an error correction form:

$$
\begin{aligned}
\ell_t &= \ell_{t-1} + b_{t-1} + \alpha \frac{e_t}{s_{t-m}} \\
b_t &= b_{t-1} + \alpha\beta^* \frac{e_t}{s_{t-m}} \\
s_t &= s_{t-m} + \gamma \frac{e_t}{\ell_{t-1} + b_{t-1}}
\end{aligned}
$$

With residual errors:

$$
e_t = y_t - (\ell_{t-1} + b_t) s_{t-m}
$$

Exercise (3)

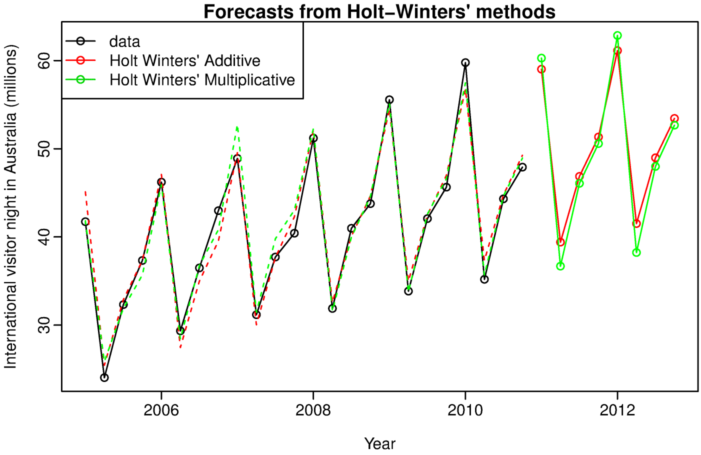
Given a series $y_1 = 2, y_2 = -3, y_3 = 3, y_4 = -2$, predict $\hat{y}_{5|4} =?$ using:

- Holt-Winters Additive Seasonal Model with
  $\alpha = 0.5, \beta^* = 0.5, \gamma = 0.5, m = 2$

In fact, it was later observed that $y_5 = 2$.

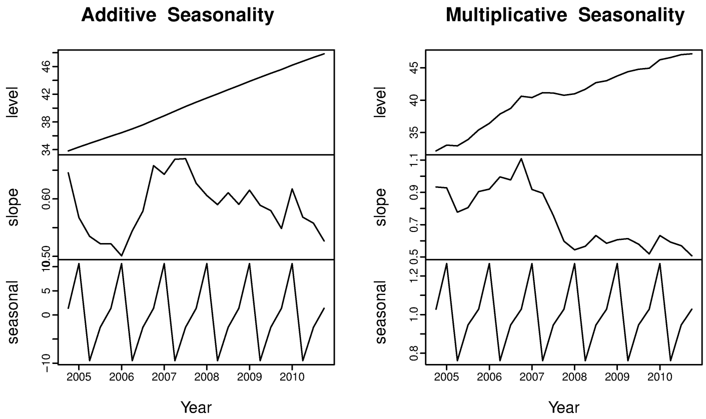
t	ℓ	b	s	$\hat{y}$
-1			2	
0	2	-5	2	
1	-1.5	-4.25	3.5	-1
2	-5.375	-4.0625	2.375	-3.75
3	-4.96875	-1.828125	7.96875	-5.9375
4	-5.5859375	-1.22265625	3.5859375	-4.421875
5				1.16015625

Example - International Visitors in Australia



Hyndman et al. 2014

Example - International Visitors in Australia (2)



Decomposing ℓ_t, b_t and s_t. Hyndman et al. 2014