Rob J Hyndman

Forecasting using R

5. Exponential smoothing methods

OTexts.com/fpp/7/

Outline

1 Simple exponential smoothing

2 Non-seasonal trend methods

Simple methods

Random walk forecasts

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

Average forecasts

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

- Want something in between that weights most recent data more highly.
- Simple exponential smoothing uses a weighted moving average with weights that decrease exponentially.

Simple methods

Random walk forecasts

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

Average forecasts

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

- Want something in between that weights most recent data more highly.
- Simple exponential smoothing uses a weighted moving average with weights that decrease exponentially.

Simple methods

Random walk forecasts

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

Average forecasts

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

- Want something in between that weights most recent data more highly.
- Simple exponential smoothing uses a weighted moving average with weights that decrease exponentially.

Simple methods

Random walk forecasts

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

Average forecasts

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

- Want something in between that weights most recent data more highly.
- Simple exponential smoothing uses a weighted moving average with weights that decrease exponentially.

Simple Exponential Smoothing

Forecast equation

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

where $0 \le \alpha \le 1$.

	Weights assigned to observations for:			
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}	$(0.2)(0.8)^4$	$(0.4)(0.6)^4$	$(0.6)(0.4)^4$	$(0.8)(0.2)^4$
y_{T-5}	$(0.2)(0.8)^5$	$(0.4)(0.6)^5$	$(0.6)(0.4)^5$	$(0.8)(0.2)^5$

Simple Exponential Smoothing

Forecast equation

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

where $0 \le \alpha \le 1$.

	Weights assigned to observations for:			
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}	$(0.2)(0.8)^4$	$(0.4)(0.6)^4$	$(0.6)(0.4)^4$	$(0.8)(0.2)^4$
y_{T-5}	$(0.2)(0.8)^5$	$(0.4)(0.6)^5$	$(0.6)(0.4)^5$	$(0.8)(0.2)^5$

Simple Exponential Smoothing

Weighted average form

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

for $t = 1, \ldots, T$, where $0 \le \alpha \le 1$ is the smoothing parameter.

The process has to start somewhere, so we let the first forecast of y_1 be denoted by ℓ_0. Then

$$\begin{aligned}
\hat{y}_{2|1} &= \alpha y_1 + (1-\alpha)\ell_0 \\
\hat{y}_{3|2} &= \alpha y_2 + (1-\alpha)\hat{y}_{2|1} \\
\hat{y}_{4|3} &= \alpha y_3 + (1-\alpha)\hat{y}_{3|2} \\
&\vdots
\end{aligned}$$

Simple Exponential Smoothing

Weighted average form

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

for $t = 1, \dots, T$, where $0 \le \alpha \le 1$ is the smoothing parameter.

The process has to start somewhere, so we let the first forecast of y_1 be denoted by ℓ_0. Then

$$\begin{aligned}
\hat{y}_{2|1} &= \alpha y_1 + (1-\alpha)\ell_0 \\
\hat{y}_{3|2} &= \alpha y_2 + (1-\alpha)\hat{y}_{2|1} \\
\hat{y}_{4|3} &= \alpha y_3 + (1-\alpha)\hat{y}_{3|2} \\
&\vdots
\end{aligned}$$

Simple Exponential Smoothing

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

Substituting each equation into the following equation:

$$\begin{aligned}
\hat{y}_{3|2} &= \alpha y_2 + (1-\alpha)\hat{y}_{2|1} \\
&= \alpha y_2 + (1-\alpha)\left[\alpha y_1 + (1-\alpha)\ell_0\right] \\
&= \alpha y_2 + \alpha(1-\alpha)y_1 + (1-\alpha)^2\ell_0 \\
\hat{y}_{4|3} &= \alpha y_3 + (1-\alpha)[\alpha y_2 + \alpha(1-\alpha)y_1 + (1-\alpha)^2\ell_0] \\
&= \alpha y_3 + \alpha(1-\alpha)y_2 + \alpha(1-\alpha)^2 y_1 + (1-\alpha)^3\ell_0 \\
&\vdots \\
\hat{y}_{T+1|T} &= \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots + (1-\alpha)^T\ell_0
\end{aligned}$$

Exponentially weighted average

$$\hat{y}_{T+1|T} = \sum_{j=0}^{T-1}\alpha(1-\alpha)^j y_{T-j} + (1-\alpha)^T\ell_0$$

Simple Exponential Smoothing

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

Substituting each equation into the following equation:

$$\begin{aligned}
\hat{y}_{3|2} &= \alpha y_2 + (1-\alpha)\hat{y}_{2|1} \\
&= \alpha y_2 + (1-\alpha)\left[\alpha y_1 + (1-\alpha)\ell_0\right] \\
&= \alpha y_2 + \alpha(1-\alpha) y_1 + (1-\alpha)^2 \ell_0 \\
\hat{y}_{4|3} &= \alpha y_3 + (1-\alpha)[\alpha y_2 + \alpha(1-\alpha) y_1 + (1-\alpha)^2 \ell_0] \\
&= \alpha y_3 + \alpha(1-\alpha) y_2 + \alpha(1-\alpha)^2 y_1 + (1-\alpha)^3 \ell_0 \\
&\vdots \\
\hat{y}_{T+1|T} &= \alpha y_T + \alpha(1-\alpha) y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots + (1-\alpha)^T \ell_0
\end{aligned}$$

Exponentially weighted average

$$\hat{y}_{T+1|T} = \sum_{j=0}^{T-1} \alpha(1-\alpha)^j y_{T-j} + (1-\alpha)^T \ell_0$$

Simple exponential smoothing

Initialization

- Last term in weighted moving average is $(1-\alpha)^T\hat{\ell}_0$.
- So value of ℓ_0 plays a role in *all* subsequent forecasts.
- Weight is small unless α close to zero or T small.
- Common to set $\ell_0 = y_1$. Better to treat it as a parameter, along with α.

Simple exponential smoothing

Initialization

- Last term in weighted moving average is $(1-\alpha)^T\hat{\ell}_0$.
- So value of ℓ_0 plays a role in *all* subsequent forecasts.
- Weight is small unless α close to zero or T small.
- Common to set $\ell_0 = y_1$. Better to treat it as a parameter, along with α.

Simple exponential smoothing

Initialization

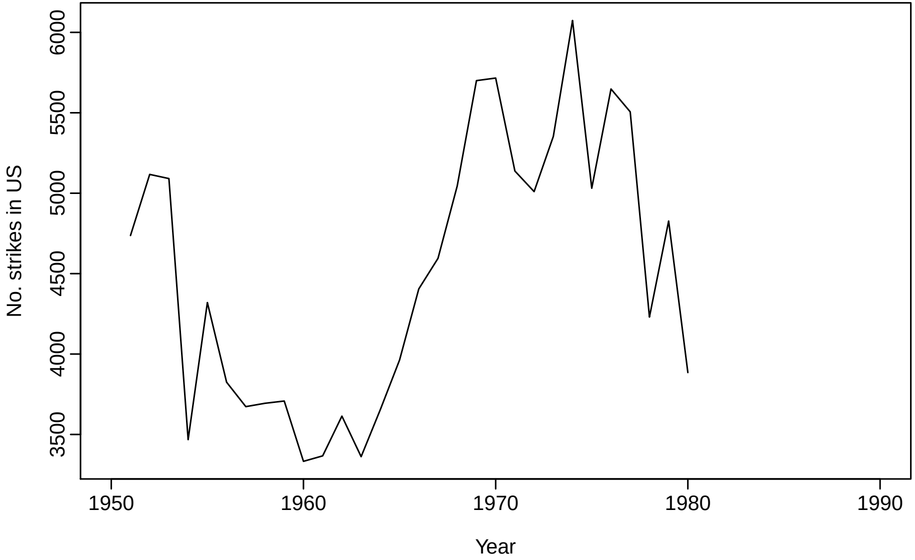
- Last term in weighted moving average is $(1-\alpha)^T\hat{\ell}_0$.
- So value of ℓ_0 plays a role in *all* subsequent forecasts.
- Weight is small unless α close to zero or T small.
- Common to set $\ell_0 = y_1$. Better to treat it as a parameter, along with α.

Simple exponential smoothing

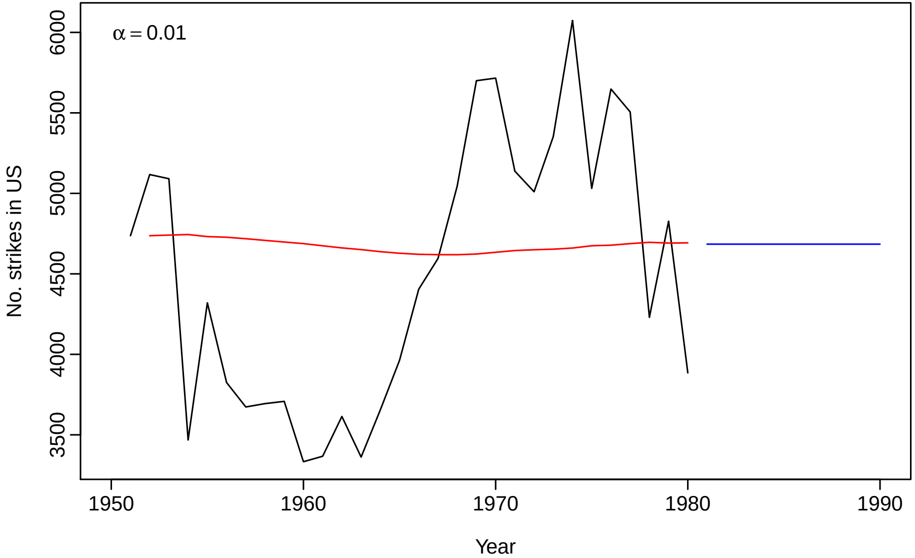
Initialization

- Last term in weighted moving average is $(1-\alpha)^T\hat{\ell}_0$.
- So value of ℓ_0 plays a role in *all* subsequent forecasts.
- Weight is small unless α close to zero or T small.
- Common to set $\ell_0 = y_1$. Better to treat it as a parameter, along with α.

Simple exponential smoothing



Simple exponential smoothing



Simple exponential smoothing

Optimization

- We can choose α and ℓ_0 by minimizing MSE:

$$\text{MSE} = \frac{1}{T-1}\sum_{t=2}^{T}(y_t - y_{t|t-1})^2$$

- Unlike regression there is no closed form solution — use numerical optimization.

Simple exponential smoothing

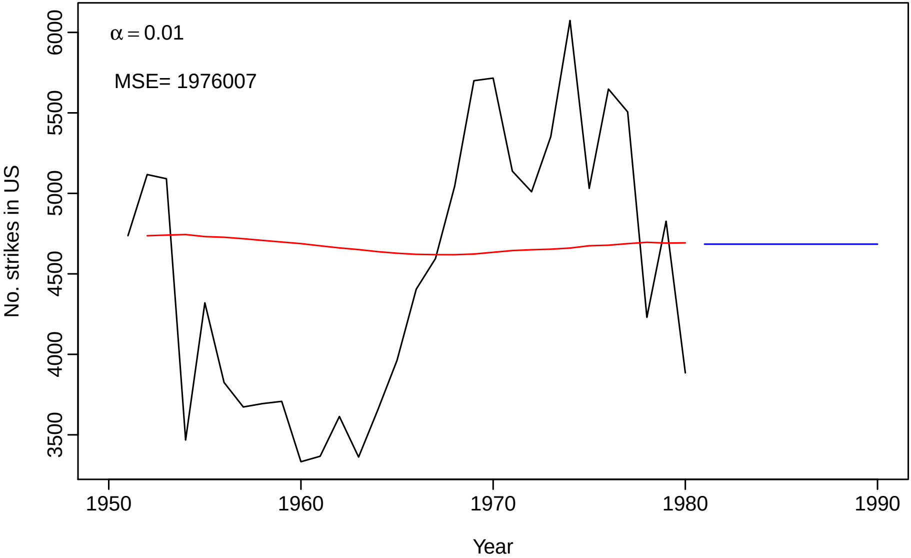
Optimization

- We can choose α and ℓ_0 by minimizing MSE:

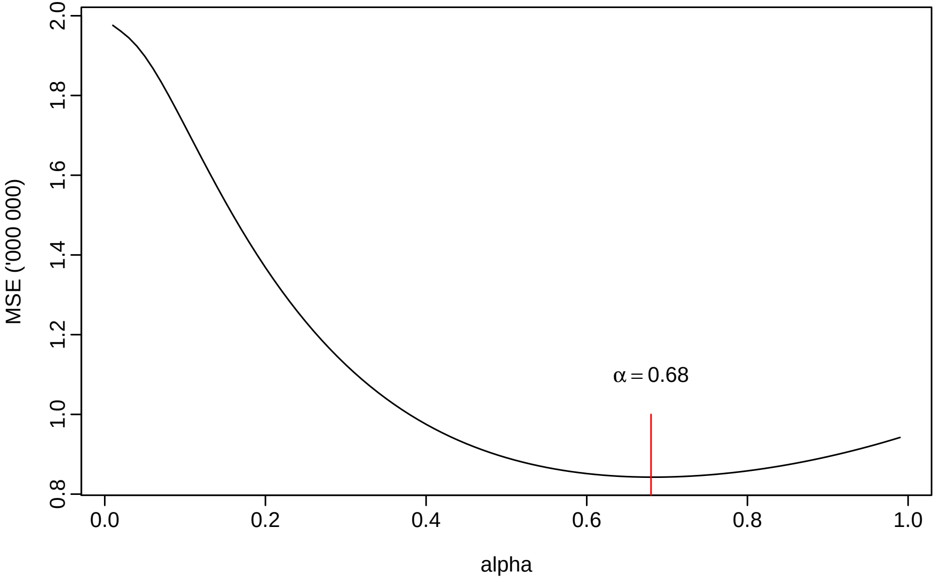
$$\text{MSE} = \frac{1}{T-1}\sum_{t=2}^{T}(y_t - y_{t|t-1})^2$$

- Unlike regression there is no closed form solution — use numerical optimization.

Simple exponential smoothing



Simple exponential smoothing



Simple exponential smoothing

Multi-step forecasts

$$\hat{y}_{T+h|T} = \hat{y}_{T+1|T}, \qquad h = 2, 3, \ldots$$

- A "flat" forecast function.
- Remember, a forecast is an estimated mean of a future value.
- So with no trend, no seasonality, and no other patterns, the forecasts are constant.

Simple exponential smoothing

Multi-step forecasts

$$\hat{y}_{T+h|T} = \hat{y}_{T+1|T}, \qquad h = 2, 3, \ldots$$

- A “flat” forecast function.
- Remember, a forecast is an estimated mean of a future value.
- So with no trend, no seasonality, and no other patterns, the forecasts are constant.

Simple exponential smoothing

Multi-step forecasts

$$\hat{y}_{T+h|T} = \hat{y}_{T+1|T}, \qquad h = 2, 3, \ldots$$

- A "flat" forecast function.
- Remember, a forecast is an estimated mean of a future value.
- So with no trend, no seasonality, and no other patterns, the forecasts are constant.

Simple exponential smoothing

Multi-step forecasts

$$\hat{y}_{T+h|T} = \hat{y}_{T+1|T}, \qquad h = 2, 3, \ldots$$

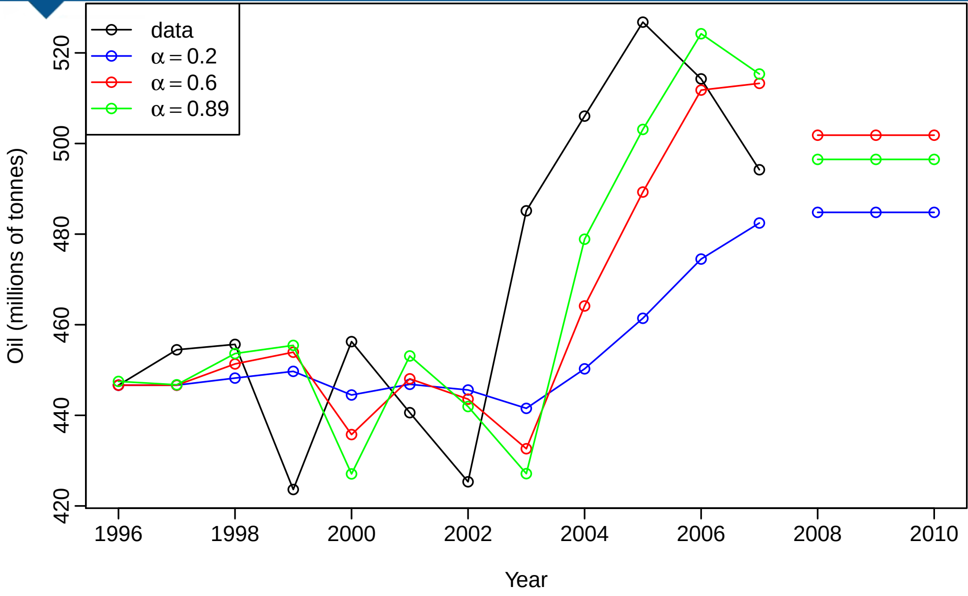
- A “flat” forecast function.
- Remember, a forecast is an estimated mean of a future value.
- So with no trend, no seasonality, and no other patterns, the forecasts are constant.

Example: Oil production

Year	Time period t	Observed values y_t	$\alpha = 0.2$	$\alpha = 0.6$	$\alpha = 0.89^*$
				Level ℓ_t	
–	0	–	446.7	446.7	447.5*
1996	1	446.7	446.7	446.7	446.7
1997	2	454.5	448.2	450.6	453.6
1998	3	455.7	449.7	453.1	455.4
1999	4	423.6	444.5	438.4	427.1
2000	5	456.3	446.8	447.3	453.1
2001	6	440.6	445.6	444.0	441.9
2002	7	425.3	441.5	434.6	427.1
2003	8	485.1	450.3	459.9	478.9
2004	9	506.0	461.4	483.0	503.1
2005	10	526.8	474.5	504.9	524.2
2006	11	514.3	482.5	509.6	515.3
2007	12	494.2	484.8	501.9	496.5
	h			Forecasts $\hat{y}_{T+h\|T}$	
2008	1	–	484.8	501.9	496.5
2009	2	–	484.8	501.9	496.5
2010	3	–	484.8	501.9	496.5

$^*\alpha = 0.89$ and $\ell_0 = 447.5$ are obtained by minimising SSE over periods $t = 1, 2, \dots, 12$.

Example: Oil production



SES in R

```
fit1 <- ses(oildata, alpha=0.2,
            initial="simple", h=3)
fit2 <- ses(oildata, alpha=0.6,
            initial="simple", h=3)
fit3 <- ses(oildata, h=3)

accuracy(fit1)
accuracy(fit2)
accuracy(fit3)
```

Equivalent forms

Weighted average form

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

Error correction form

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

Component form

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

- ℓ_t = estimate of level of series.

Equivalent forms

Weighted average form

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

Error correction form

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

Component form

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

- ℓ_t = estimate of level of series.

Equivalent forms

Weighted average form

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

Error correction form

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

Component form

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

- ℓ_t = estimate of level of series.

Outline

1 Simple exponential smoothing

2 Non-seasonal trend methods

Holt's local trend method

- Holt (1957) extended SES to allow forecasting of data with trends.
- Two smoothing parameters: α and β^* (with values between 0 and 1).

$$\hat{y}_{t+h|t} = \ell_t + hb_t$$
$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$$
$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$$

Holt's local trend method

- Holt (1957) extended SES to allow forecasting of data with trends.
- Two smoothing parameters: α and β^* (with values between 0 and 1).

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

- ℓ_t denotes an estimate of the level of the series at time t

Holt's local trend method

- Holt (1957) extended SES to allow forecasting of data with trends.
- Two smoothing parameters: α and β^* (with values between 0 and 1).

$$\hat{y}_{t+h|t} = \ell_t + hb_t$$
$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$$
$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$$

- ℓ_t denotes an estimate of the level of the series at time t
- b_t denotes an estimate of the slope of the series at time t.

Holt's local trend method

- Holt (1957) extended SES to allow forecasting of data with trends.
- Two smoothing parameters: α and β^* (with values between 0 and 1).

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

- ℓ_t denotes an estimate of the level of the series at time t
- b_t denotes an estimate of the slope of the series at time t.

Holt's local trend method

- Holt (1957) extended SES to allow forecasting of data with trends.
- Two smoothing parameters: α and β^* (with values between 0 and 1).

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

- ℓ_t denotes an estimate of the level of the series at time t
- b_t denotes an estimate of the slope of the series at time t.

Holt's local trend method

- Holt (1957) extended SES to allow forecasting of data with trends.
- Two smoothing parameters: α and β^* (with values between 0 and 1).

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

- ℓ_t denotes an estimate of the level of the series at time t
- b_t denotes an estimate of the slope of the series at time t.

Holt's local trend method

Optimization

- Need to find α and β^* by minimizing the value of MSE.
- We also optimize MSE for ℓ_0 and b_0.
- Optimizing in four dimensions is getting tricky!

Holt's method in R

```
fit1 <- holt(strikes)
plot(fit1$model)
plot(fit1, plot.conf=FALSE)
lines(fitted(fit1), col="red")
fit1$model

fit2 <- ses(strikes)
plot(fit2$model)
plot(fit2, plot.conf=FALSE)
lines(fit1$mean, col="red")

accuracy(fit1)
accuracy(fit2)
```

Comparing Holt and SES

- Holt’s method will almost always have better in-sample RMSE because it is optimized over one additional parameter.
- It may not be better on other measures.
- You need to compare out-of-sample RMSE (using a test set) for the comparison to be useful.
- But we don’t have enough data.
- A better method for comparison will be in the next session!

Comparing Holt and SES

- Holt's method will almost always have better in-sample RMSE because it is optimized over one additional parameter.
- It may not be better on other measures.
- You need to compare out-of-sample RMSE (using a test set) for the comparison to be useful.
- But we don't have enough data.
- A better method for comparison will be in the next session!

Comparing Holt and SES

- Holt's method will almost always have better in-sample RMSE because it is optimized over one additional parameter.
- It may not be better on other measures.
- You need to compare out-of-sample RMSE (using a test set) for the comparison to be useful.
- But we don't have enough data.
- A better method for comparison will be in the next session!

Comparing Holt and SES

- Holt's method will almost always have better in-sample RMSE because it is optimized over one additional parameter.
- It may not be better on other measures.
- You need to compare out-of-sample RMSE (using a test set) for the comparison to be useful.
- But we don't have enough data.
- A better method for comparison will be in the next session!

Comparing Holt and SES

- Holt's method will almost always have better in-sample RMSE because it is optimized over one additional parameter.
- It may not be better on other measures.
- You need to compare out-of-sample RMSE (using a test set) for the comparison to be useful.
- But we don't have enough data.
- A better method for comparison will be in the next session!

Multiplicative version of Holt's method

$$
\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t b_t^h \\
\ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} b_{t-1}) \\
b_t &= \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*) b_{t-1}
\end{aligned}
$$

- ℓ_t denotes an estimate of the level of the series at time t
- b_t denotes an estimate of the relative growth of the series at time t.
- In R: `holt(x, exponential=TRUE)`
- Comparing additive and multiplicative trend methods in-sample is ok because they have the same number of parameters to optimize.

Exponential trend method

Multiplicative version of Holt's method

$$
\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t b_t^h \\
\ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} b_{t-1}) \\
b_t &= \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*) b_{t-1}
\end{aligned}
$$

- ℓ_t denotes an estimate of the level of the series at time t
- b_t denotes an estimate of the relative growth of the series at time t.
- In R: `holt(x, exponential=TRUE)`
- Comparing additive and multiplicative trend methods in-sample is ok because they have the same number of parameters to optimize.

Exponential trend method

Multiplicative version of Holt's method

$$\begin{aligned}\hat{y}_{t+h|t} &= \ell_t b_t^h \\ \ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} b_{t-1}) \\ b_t &= \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*) b_{t-1}\end{aligned}$$

- ℓ_t denotes an estimate of the level of the series at time t
- b_t denotes an estimate of the relative growth of the series at time t.
- In R: `holt(x, exponential=TRUE)`
- Comparing additive and multiplicative trend methods in-sample is ok because they have the same number of parameters to optimize.

Exponential trend method

Multiplicative version of Holt's method

$$\hat{y}_{t+h|t} = \ell_t b_t^h$$
$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} b_{t-1})$$
$$b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*) b_{t-1}$$

- ℓ_t denotes an estimate of the level of the series at time t
- b_t denotes an estimate of the relative growth of the series at time t.
- In R: `holt(x, exponential=TRUE)`
- Comparing additive and multiplicative trend methods in-sample is ok because they have the same number of parameters to optimize.

Exponential trend method

Multiplicative version of Holt's method

$$
\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t b_t^h \\
\ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} b_{t-1}) \\
b_t &= \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*) b_{t-1}
\end{aligned}
$$

- ℓ_t denotes an estimate of the level of the series at time t
- b_t denotes an estimate of the relative growth of the series at time t.
- In R: `holt(x, exponential=TRUE)`
- Comparing additive and multiplicative trend methods in-sample is ok because they have the same number of parameters to optimize.

Damped trend method

- Gardner and McKenzie (1985) suggested that the trends should be “damped” to be more conservative for longer forecast horizons.
- Two smoothing parameters: α and β^* (with values between 0 and 1), and one damping parameter $0 < \phi < 1$.

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

Damped trend method

- Gardner and McKenzie (1985) suggested that the trends should be "damped" to be more conservative for longer forecast horizons.
- Two smoothing parameters: α and β^* (with values between 0 and 1), and one damping parameter $0 < \phi < 1$.

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

- ϕ dampens the trend so it approaches a constant.

Damped trend method

- Gardner and McKenzie (1985) suggested that the trends should be “damped” to be more conservative for longer forecast horizons.
- Two smoothing parameters: α and β^* (with values between 0 and 1), and one damping parameter $0 < \phi < 1$.

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

- ϕ dampens the trend so it approaches a constant.

Damped trend method

- Gardner and McKenzie (1985) suggested that the trends should be “damped” to be more conservative for longer forecast horizons.
- Two smoothing parameters: α and β^* (with values between 0 and 1), and one damping parameter $0 < \phi < 1$.

$$\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + (\phi + \phi^2 + \cdots + \phi^{h-1})b_t \\
\ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}
\end{aligned}$$

- ϕ dampens the trend so it approaches a constant.

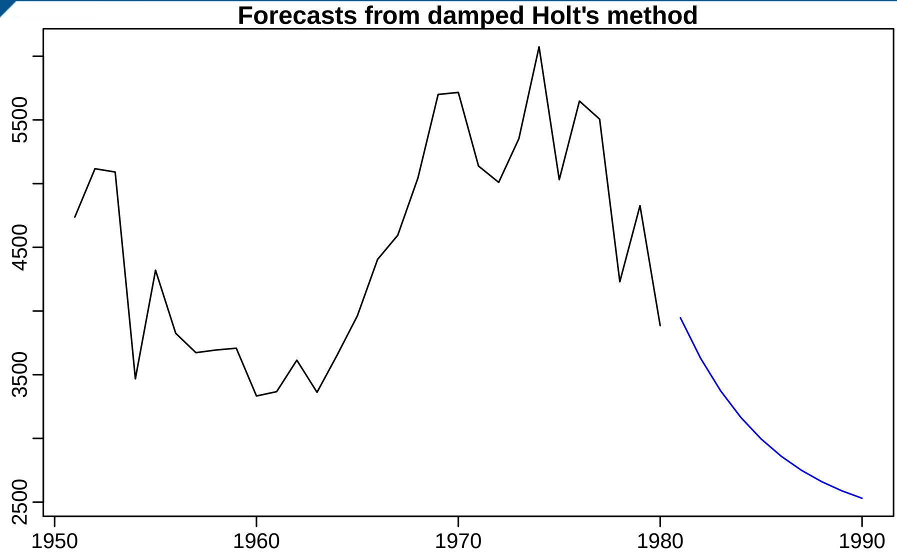
Damped trend method

- Gardner and McKenzie (1985) suggested that the trends should be "damped" to be more conservative for longer forecast horizons.
- Two smoothing parameters: α and β^* (with values between 0 and 1), and one damping parameter $0 < \phi < 1$.

$$\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + (\phi + \phi^2 + \cdots + \phi^{h-1})b_t \\
\ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}
\end{aligned}$$

- ϕ dampens the trend so it approaches a constant.

Damped trend method



Damped trend method

$$
\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + (\phi + \phi^2 + \cdots + \phi^{h-1})b_t \\
\ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}
\end{aligned}
$$

- If $\phi = 1$, this is the same as Holt's method.
- ϕ can be estimated along with α and β^* by minimizing the MSE.
- Damped trend method often gives better forecasts than linear trend.
- Forecasts converge to $\ell_T + \phi b_T/(1-\phi)$ as $h \to \infty$.

Damped trend method

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

- If $\phi = 1$, this is the same as Holt's method.
- ϕ can be estimated along with α and β^* by minimizing the MSE.
- Damped trend method often gives better forecasts than linear trend.
- Forecasts converge to $\ell_T + \phi b_T/(1-\phi)$ as $h \to \infty$.

Damped trend method

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

- If $\phi = 1$, this is the same as Holt's method.
- ϕ can be estimated along with α and β^* by minimizing the MSE.
- Damped trend method often gives better forecasts than linear trend.
- Forecasts converge to $\ell_T + \phi b_T/(1-\phi)$ as $h \to \infty$.

Damped trend method

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

- If $\phi = 1$, this is the same as Holt's method.
- ϕ can be estimated along with α and β^* by minimizing the MSE.
- Damped trend method often gives better forecasts than linear trend.
- Forecasts converge to $\ell_T + \phi b_T/(1-\phi)$ as $h \to \infty$.

Damped trend method

$$\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + (\phi + \phi^2 + \cdots + \phi^{h-1}) b_t \\
\ell_t &= \alpha y_t + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}
\end{aligned}$$

- If $\phi = 1$, this is the same as Holt's method.
- ϕ can be estimated along with α and β^* by minimizing the MSE.
- Damped trend method often gives better forecasts than linear trend.
- Forecasts converge to $\ell_T + \phi b_T/(1 - \phi)$ as $h \to \infty$.

Multiplicative damped trend method

Taylor (2003) introduced multiplicative damping.

$$\hat{y}_{t+h|t} = \ell_t b_t^{(\phi+\phi^2+\cdots+\phi^h)}$$

$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} b_{t-1}^{\phi})$$

$$b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*) b_{t-1}^{\phi}$$

- $\phi = 1$ gives exponential trend method
- Forecasts converge to $\ell_T + b_T^{\phi/(1-\phi)}$ as $h \to \infty$.

Multiplicative damped trend method

Taylor (2003) introduced multiplicative damping.

$$\hat{y}_{t+h|t} = \ell_t b_t^{(\phi+\phi^2+\dots+\phi^h)}$$
$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} b_{t-1}^{\phi})$$
$$b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*) b_{t-1}^{\phi}$$

- $\phi = 1$ gives exponential trend method
- Forecasts converge to $\ell_T + b_T^{\phi/(1-\phi)}$ as $h \to \infty$.

Multiplicative damped trend method

Taylor (2003) introduced multiplicative damping.

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

- $\phi = 1$ gives exponential trend method
- Forecasts converge to $\ell_T + b_T^{\phi/(1-\phi)}$ as $h \to \infty$.