3.2 Double Exponential Smoothing

Double exponential smoothing

- Recall simple exponential smoothing, i.e.

$$S_T = \alpha x_T + (1-\alpha) S_{T-1},$$

  is a biased forecast for linear trend process $X_t = b_1 + b_2 t + \epsilon_t$.
- Now we consider *double exponential smoothing* (DES)[3]

$$S_T^{[2]} = \alpha S_T + (1-\alpha) S_{T-1}^{[2]}.$$

- To get an **unbiased** estimate of linear trend model from DES, we estimate b_1 and b_2 separately $\curvearrowright$

[3] SES of SES.

Estimate of b_2.

- **Calculation of $S_T^{[2]}$.** Recall Equation (1):

$$\mathbb{E}[S_T] = \mathbb{E}[X_T] - \frac{1-\alpha}{\alpha} b_2.$$

Intuitively, in (1) replacing X by S and replacing S by $S^{[2]}$, we have

$$\mathbb{E}[S_T^{[2]}] = \mathbb{E}[S_T] - \frac{1-\alpha}{\alpha} b_2. \tag{2}$$

It implies

$$b_2 = \frac{\alpha}{1-\alpha}\left\{\mathbb{E}[S_T] - \mathbb{E}[S_T^{[2]}]\right\}. \tag{3}$$

- **Estimate of b_2.** From (3) it suggests the **unbiased** estimate of b_2 at T is

$$\hat{b}_2(T) = \frac{\alpha}{1-\alpha}(S_T - S_T^{[2]}). \tag{4}$$

Estimate of b_1 and X.

- From (1) we get
$$\mathbb{E}[X_T] = \mathbb{E}[S_T] + \frac{1-\alpha}{\alpha} b_2.$$
- Replacing b_2 by Equation (3) we get
$$\mathbb{E}[X_T] = 2\mathbb{E}[S_T] - \mathbb{E}[S_T^{[2]}]. \tag{5}$$
- **Estimate for X_T.** (5) suggests a reasonable estimate for X at T is
$$\hat{X}_T = 2S_T - S_T^{[2]}. \tag{6}$$
- **Estimate for b_1.** Let $\widehat{b}_1(T)$ be the estimate for b_1 at T, then
$$\widehat{X}_T = \widehat{b}_1(T) + \widehat{b}_2(T) \cdot T,$$
which implies by (6) and (4)
$$\hat{b}_1(T) = 2S_T - S_T^{[2]} - T \cdot \frac{\alpha}{1-\alpha}(S_T - S_T^{[2]}). \tag{7}$$

Determination of S_0 and $S_0^{[2]}$

- To perform the forecast/estimate, we need to know S_0 and $S_0^{[2]}$.
- Letting $T = 0$ in the estimate for b_1 (7) and estimate for b_2 (4), S_0 and $S_0^{[2]}$ can be obtained

$$\begin{cases} S_0 = \hat{b}_1(0) - \left(\dfrac{1-\alpha}{\alpha}\right)\hat{b}_2(0) \\ S_0^{[2]} = \hat{b}_1(0) - 2\left(\dfrac{1-\alpha}{\alpha}\right)\hat{b}_2(0). \end{cases}$$

- The initial estimate $\widehat{b_1}(0)$ and $\widehat{b_2}(0)$ can be usually obtained by fitting he linear trend model to all or part of the available data by least square: $y_t = b_1(0) + b_2(0)t + u_t$, $t = 1, \cdots$
- In exercises, $\widehat{b_1}(0)$ and $\widehat{b_2}(0)$, or S_0 and $S_0^{[2]}$ are usually given.

Forecast of linear trend model by DES

- τ-step (ahead) forecast at time T is given by

$$\begin{aligned}\hat{X}_T(\tau) &= \hat{b}_1(T) + \hat{b}_2(T)(T+\tau) \\ &= 2S_T - S_T^{[2]} - T \cdot \frac{\alpha}{1-\alpha}(S_T - S_T^{[2]}) \\ &\quad + \frac{\alpha}{1-\alpha}(S_T - S_T^{[2]})(T+\tau) \\ &= 2S_T - S_T^{[2]} + \frac{\alpha}{1-\alpha}(S_T - S_T^{[2]})\tau \\ &:= a(T) + b(T) \cdot \tau,\end{aligned}$$

where

$$\begin{cases} a(T) = 2S_T - S_T^{[2]} = \hat{X}_T & \text{by (6)} \\ b(T) = \dfrac{\alpha}{1-\alpha}(S_T - S_T^{[2]}). \end{cases}$$

- The forecast is unbiased: according to the construction of $\widehat{b}_1(T)$ and $\widehat{b}_2(T)$, we have

$$\mathbb{E}[\widehat{X}_T(\tau)] = \mathbb{E}\left[\widehat{b}_1(T) + \widehat{b}_2(T)(T+\tau)\right] = b_1 + b_2(T+\tau).$$

Summary of DES for linear trend model

The forecast is

$$\hat{x}_{T+\tau} = a(T) + b(T) \cdot \tau,$$

where

$$a(T) = 2S_T - S_T^{[2]}, \qquad b(T) = \frac{\alpha}{1-\alpha}(S_T - S_T^{[2]}),$$

$$S_T = \alpha x_T + (1-\alpha) S_{T-1},$$

$$S_T^{[2]} = \alpha S_T + (1-\alpha) S_{T-1}^{[2]}.$$

Higher order ES for general polynomial model (sketch)

- Constant model: SES (first order exponential smoothing),
  Linear trend model: DES (second order exponential smoothing).
- Consider general polynomial model

$$X_t = \beta_0 + \beta_1 t + \frac{\beta_2}{2!}t^2 + \cdots + \frac{\beta_n}{n!}t^n + \epsilon_t.$$

- Model parameters β_0, β_1, $\cdots$, β_n can be estimated by $(n+1)$-order exponential smoothing

$$\begin{aligned} S_T &= \alpha X_T + (1-\alpha) S_{T-1} \\ S_T^{[2]} &= \alpha S_T + (1-\alpha) S_{T-1}^{[2]} \\ &\vdots \\ S_T^{[n+1]} &= \alpha S_T^{[n]} + (1-\alpha) S_{T-1}^{[n+1]} \end{aligned}$$

- Refer to **Fundamental theorem of exponential smoothing** in *Brown, R. G.. Smoothing, Forecasting and Prediction of Discrete Time Series. Prentice-Hall, Englewood Cliffs, NJ.*

Example 3.2 (1/2)

Use DES to do one-step forecast for Jan-June, 1987.

Month of 1987	Period	Actual sales	S_T	$S_T^{[2]}$	$\hat{x}_{T-1}(1)$
Jan	1	538	446.28	352.22	541
Feb	2	570	458.65	362.86	551
March	3	600	472.79	373.85	565
Apr	4	565	482.01	384.67	583
May	5	485	482.31	394.43	590
June	6	604	494.48	404.44	580

Let $S_0 = 436.09$ and $S_0^{[2]} = 341.77$, and let the smoothing constant $\alpha = 0.1$.

Remark: $\hat{x}_{T-1}(1)$ is the one-step ahead forecast for time T made at $T-1$.

Example 3.2 (2/2)

Solution.

- Recall the forecasting equation

$$\widehat{X}_T(\tau) = 2S_T - S_T^{[2]} + \frac{\alpha}{1-\alpha}(S_T - S_T^{[2]})\tau.$$

- **1-step forecast of Jan. 1987.** The 1-step forecast at time 0 (Dec. 1986) is

$$\begin{aligned}\widehat{X}_0(1) &= 2S_0 - S_0^{[2]} + \frac{\alpha}{1-\alpha}(S_0 - S_0^{[2]}) \times 1 \\ &= 2 \times 436.09 - 341.77 + \frac{0.1}{1-0.9} \times (436.09 - 341.77) = 540.89.\end{aligned}$$

- **1-step forecast of Feb. 1987.**

$$\begin{aligned}S_1 &= \alpha X_1 + (1-\alpha)S_0 = 0.1(538) + (1-0.1)(436.09) = 446.28; \\ S_1^{[2]} &= \alpha S_1 + (1-\alpha)S_0^{[2]} = 0.1(446.28) + (1-0.1)(341.77) = 352.22.\end{aligned}$$

Thus,

$$\widehat{X}_1(1) = 2S_1 - S_1^{[2]} + \frac{\alpha}{1-\alpha}(S_1 - S_1^{[2]}) = 550.79.$$

- The other forecasts are obtained in a similar fashion.

3.3 Holt's trend method

- Holt's linear trend method.
- Holt's exponential trend method.

Holt's linear trend method

- Instead of the single parameter double exponential smoothing method, Holt suggested a two-parameter method, which might be more flexible, since we can adjust two smoothing constants rather than one.
- Consider $y_t = c_0 + c_1 t + \epsilon_t$. Let a_T and b_T be the estimate of the level component and the slope component of the linear trend model, at time T.

In Holt's exponential smoothing, a_T is a weighted sum of levels, b_T is a weighted sum of slopes:

1. We obtain an updated estimate a_T of the permanent (level) component by using the equation

$$a_T = \alpha y_T + (1-\alpha)(a_{T-1} + b_{T-1}), \quad \text{where } 0 < \alpha < 1.$$

2. We obtain an updated estimate b_T of the trend (slope) component by using the equation

$$b_T = \beta(a_T - a_{T-1}) + (1-\beta)b_{T-1}, \quad \text{where } 0 < \beta < 1.$$

Holt's linear trend method

3. τ-step (ahead) forecast made at time T is

$$\hat{y}_T(\tau) = a_T + b_T \cdot \tau.$$

Remark: **how to obtain a_0 and b_0?**

- We may fit a regression with linear trend to the first few observations. Then the initial level a_0 is set to the intercept, and the initial slope b_0 is set to the regression slope.
- To make it simple, we may set

$$a_0 = y_1, \quad b_0 = y_2 - y_1.$$

Example 3.3 (1/3)

We demonstrate the application of Holt's linear method to air transportation data for all passengers with an Australian airline. To initialize the method we set $a_0 = y_1$ and $b_0 = y_2 - y_1$.

For demonstration purposes we set $\alpha = 0.8$ and $\beta = 0.2$.

Year	T	y_T	a_T	b_T	$\hat{y}_{T-1}(1)$
	0		17.55	4.31	
1990	1	17.55	18.41	3.62	21.86
1991	2	21.86	21.89	3.59	22.03

Now $a_0 = y_1 = 17.55$ and $b_0 = y_2 - y_1 = 21.86 - 17.55 = 4.31$. Recall the forecast formula:

$$\hat{y}_T(\tau) = a_T + b_T \cdot \tau.$$

So we have

$$\hat{y}_0(1) = a_0 + b_0 = 17.55 + 4.31 = 21.86.$$

Next, $a_1 = \alpha y_1 + (1-\alpha)(a_0 + b_0) = 0.8(17.55) + 0.2(21.86) = 18.41$ and $b_1 = \beta(a_1 - a_0) + (1-\beta)b_0 = 0.2(18.41 - 17.55) + 0.8(4.31) = 3.62$. Then

$$\hat{y}_1(1) = a_1 + b_1 = 18.41 + 3.62 = 22.03.$$

Example 3.3 (2/3)

Year	T	y_T	a_T	b_T	$\hat{y}_{T-1}(1)$
	0		17.55	4.31	
1990	1	17.55	18.41	3.62	21.86
1991	2	21.86	21.89	3.59	22.03
1992	3	23.89	24.21	3.33	25.48
1993	4	26.93	27.05	3.24	27.54
1994	5	26.89	27.57	2.69	30.29
1995	6	28.83	29.12	2.46	30.26
1996	7	30.08	30.38	2.22	31.58
1997	8	30.95	31.28	1.96	32.60
1998	9	30.19	30.80	1.47	33.24
1999	10	31.58	31.72	1.36	32.27
2000	11	32.58	32.68	1.28	33.08
2001	12	33.48	33.57	1.20	33.96
2002	13	39.02	38.17	1.88	34.78
2003	14	41.39	41.12	2.10	40.06
2004	15	41.60	41.92	1.84	43.22

Example 3.3 (3/3)

If we want to forecast the air passengers after 2004 (all following forecasts are made at 2004) then we have

$$\hat{y}_{15}(\tau) = a_{15} + b_{15} \cdot \tau, \quad \tau = 1, \cdots, 5$$

Year	T	y_T	a_T	b_T	$\hat{y}_{T-1}(1)$
2004	15	41.60	41.92	1.84	43.22

Year	τ	y_T	a_T	b_T	$\hat{y}_{15}(\tau)$
2005	1				43.76
2006	2				45.59
2007	3				47.43
2008	4				49.27
2009	5				51.10

Holt's exponential trend method

- Holt's exponential trend method are used to forecast exponential trend model:

$$Y_t = ab^t \epsilon_t.$$

- A variation from Holt's linear trend method is achieved by allowing the level and the slope to be multiplied rather than added:

$$\hat{y}_T(\tau) = a_T \times (b_T)^\tau;$$
$$a_T = \alpha y_T + (1-\alpha)(a_{T-1} \times b_{T-1});$$
$$b_T = \beta\left(\frac{a_T}{a_{T-1}}\right) + (1-\beta)b_{T-1};$$
$$a_0 = y_1, \quad b_0 = y_2/y_1,$$

where b_T now represents an estimated growth rate (in relative terms rather than absolute) which is multiplied rather than added to the estimated level.

Example 3.3, *cont'd*

- We fit an **exponential trend model** to the data in Example 3.3, by using Holt's exponential method.
- The same as Holt's linear method we set $\alpha = 0.8$, $\beta = 0.2$ and $a_0 = y_1$, however, $b_0 = \frac{y_2}{y_1}$.

Year	T	y_T	a_T	b_T	$\hat{y}_{T-1}(1)$
	0		17.55	1.25	
1990	1	17.55	18.41	1.21	21.86
1991	2	21.86	21.93	1.20	22.21

- **Calculate $\widehat{y}_0(1)$.** Since $a_0 = y_1 = 17.55$ and $b_0 = \frac{y_2}{y_1} = \frac{21.86}{17.55} = 1.25$, we have

$$\hat{y}_0(1) = a_0 \times b_0 = 17.55(1.25) = 21.86.$$

- **Calculate $\widehat{y}_1(1)$.** Since
$a_1 = \alpha y_1 + (1-\alpha)(a_0 \times b_0) = 0.8(17.55) + 0.2(21.86) = 18.41$ and
$b_1 = \beta \frac{a_1}{a_0} + (1-\beta) b_0 = 0.2 \frac{18.41}{17.55} + 0.8(1.25) = 1.21$, we have

$$\hat{y}_1(1) = a_1 \times b_1 = 18.41(1.21) = 22.21.$$

Example 3.3, *cont'd*

Year	T	y_T	a_T	b_T	$\hat{y}_{T-1}(1)$
	0		17.55	1.25	
1990	1	17.55	18.41	1.21	21.86
1991	2	21.86	21.93	1.20	22.21
1992	3	23.89	24.39	1.18	26.38
1993	4	26.93	27.32	1.17	28.89
1994	5	26.89	27.91	1.14	32.02
1995	6	28.83	29.44	1.12	31.88
1996	7	30.08	30.68	1.11	33.10
1997	8	30.95	31.56	1.09	33.99
1998	9	30.19	31.04	1.07	34.47
1999	10	31.58	31.91	1.06	33.23
2000	11	32.58	32.84	1.06	33.89
2001	12	33.48	33.71	1.05	34.66
2002	13	39.02	38.29	1.07	35.39
2003	14	41.39	41.28	1.07	40.86
2004	15	41.60	42.10	1.06	44.13

Year	τ	y_T	a_T	b_T	$\hat{y}_{15}(\tau)$
2005	1				44.60
2006	2				47.24
2007	3				50.04
2008	4				53.01
2009	5				56.15

Example 3.3, *cont'd*

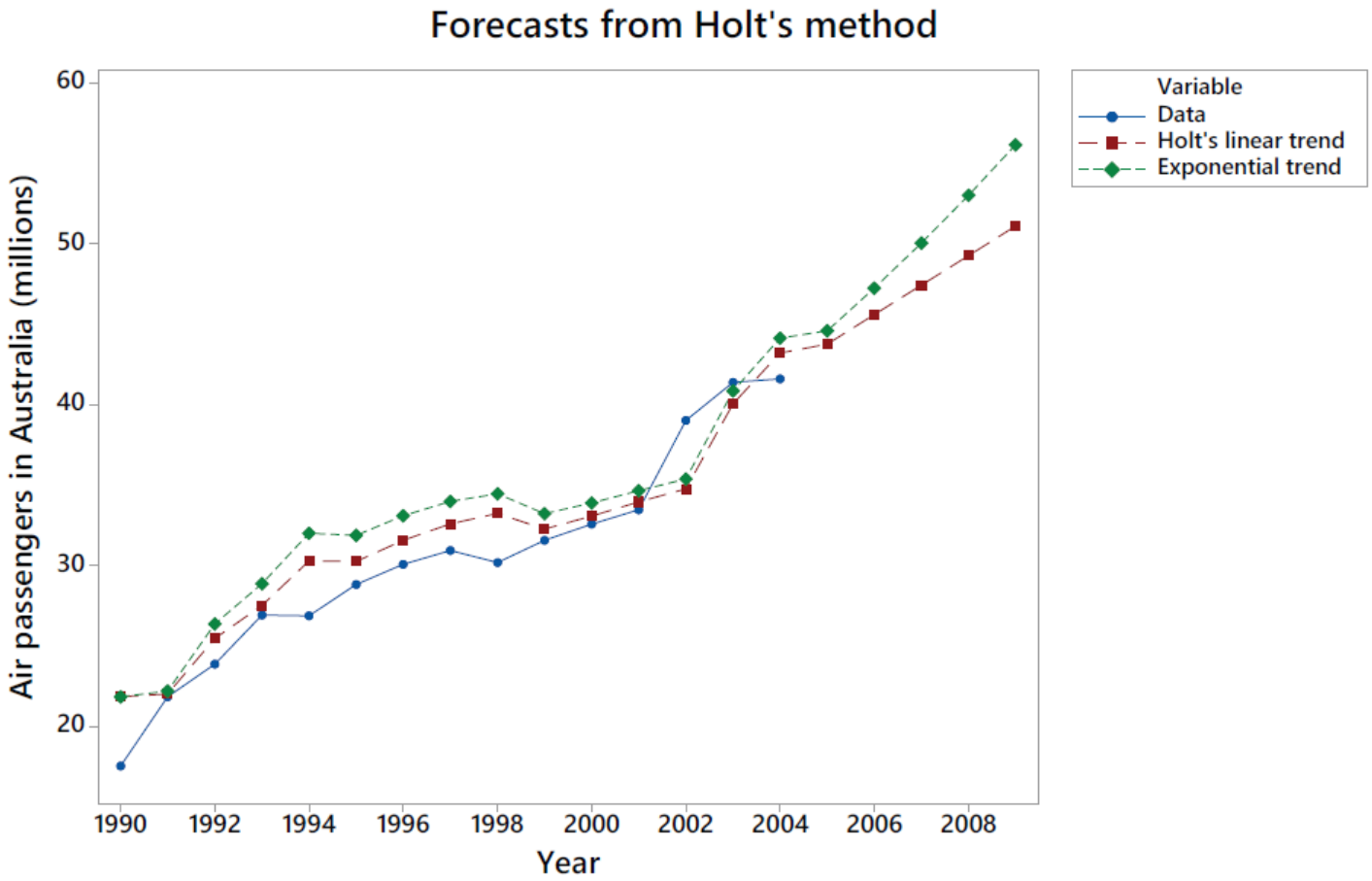


Figure 3.2: Comparison of Holt's linear and exponential trends.

3.4 Damped Trend

Motivation

- Recall the forecast formula of Holt's linear trend:

$$\hat{y}_T(\tau) = a_T + b_T \cdot \tau.$$

- Imagine you are forecasting sales growth for a brand-new futuristic coffee mug. At first, sales increase rapidly because everyone wants one. But eventually:
  - the market gets saturated (not everyone needs 10 coffee mugs),
  - competitors enter,
  - trends slow down (initial hype dies off).
- Holt's linear model overestimates!
- Motivated by this observation, Gardner & McKenzie (1985)[4] introduced one more parameter that "dampens" the trend to a flat line some time in the future.

[4] Gardner, E. S., & McKenzie, E. (1985). Forecasting trends in time series. Management Science, 31(10), 1237-1246.

Additive damped trend

- In addition to the smoothing constants α and β, this method also includes a damping constant $0 < \phi < 1$:

$$\hat{y}_T(\tau) = a_T + (\phi + \phi^2 + \cdots + \phi^\tau) b_T;$$
$$a_T = \alpha y_T + (1-\alpha)(a_{T-1} + \phi b_{T-1});$$
$$b_T = \beta(a_T - a_{T-1}) + (1-\beta)\phi b_{T-1}.$$

- Here ϕ dampens the trend so that it approaches a constant some time in the future.

  In fact the forecasts converge to $a_T + \frac{\phi b_T}{1-\phi}$ as $\tau \to \infty$.

- The effect of this is that short-run forecasts are trended while long-run forecasts are constant.

Remark: If $\phi = 1$ the method is identical to Holt's linear method.

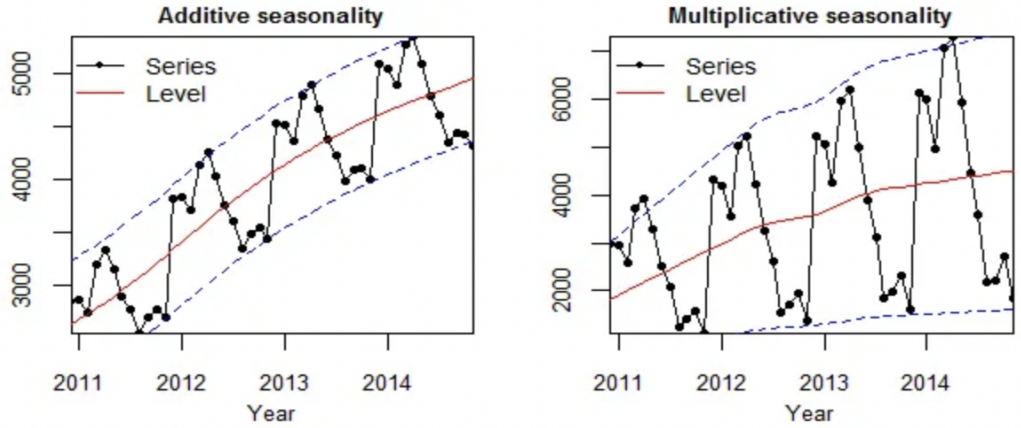
Multiplicative damped trend

- A damping constant can also be included to the exponential trend method resulting to a multiplicative damped trend method:

$$\hat{y}_T(\tau) = a_T \times b_T^{(\phi+\phi^2+\cdots+\phi^\tau)};$$
$$a_T = \alpha y_T + (1-\alpha)(a_{T-1} \times b_{T-1}^{\phi});$$
$$b_T = \beta\left(\frac{a_T}{a_{T-1}}\right) + (1-\beta)b_{T-1}^{\phi}.$$

3.5 Holt-Winters' Seasonal Method

- The Holt-Winters' seasonal method consists of 1 forecast equation and 3 smoothing equations: one for the level ℓ, one for the trend b, and one for the seasonal component s, with corresponding smoothing parameters α, β and γ.
- There are two variations to this method that differ in the nature of the seasonal component. The additive method is preferred when the seasonal variations are roughly constant through the series, while the multiplicative method is preferred when the seasonal variations are changing proportional to the level of the series.



1. Holt-Winters' additive method.

$$\begin{aligned}
\hat{y}_T(\tau) &= \ell_T + \tau \cdot b_T + s_{T+\tau-m(k+1)} \\
\ell_T &= \alpha(y_T - s_{T-m}) + (1-\alpha)(\ell_{T-1} + b_{T-1}) \\
b_T &= \beta(\ell_T - \ell_{T-1}) + (1-\beta)b_{T-1} \\
s_T &= \gamma(y_T - \ell_{T-1} - b_{T-1}) + (1-\gamma)s_{T-m}
\end{aligned}$$

where k is the integer part of $\frac{\tau-1}{m}$.

2. Holt-Winters' multiplicative method.

$$\begin{aligned}
\hat{y}_T(\tau) &= (\ell_T + \tau \cdot b_T)s_{T+\tau-m(k+1)} \\
\ell_T &= \alpha\frac{y_T}{s_{T-m}} + (1-\alpha)(\ell_{T-1} + b_{T-1}) \\
b_T &= \beta(\ell_T - \ell_{T-1}) + (1-\beta)b_{T-1} \\
s_T &= \gamma\frac{y_T}{\ell_{T-1} + b_{T-1}} + (1-\gamma)s_{T-m}
\end{aligned}$$

1. Holt-Winters' damped method: additive damped trend and addtive seasonality

$$\begin{aligned}
\hat{y}_T(\tau) &= \ell_T + (\phi + \phi^2 + \cdots + \phi^\tau) b_T + s_{T+\tau-m(k+1)} \\
\ell_T &= \alpha(y_T - s_{T-m}) + (1-\alpha)(\ell_{T-1} + \phi b_{T-1}) \\
b_T &= \beta(\ell_T - \ell_{T-1}) + (1-\beta)\phi b_{T-1} \\
s_T &= \gamma(y_T - \ell_{T-1} - \phi b_{T-1}) + (1-\gamma) s_{T-m}
\end{aligned}$$

2. Addtive damped trend and multiplicative seasonality...
3. Exponential damped trend and additive seasonality...
4. Exponential damped trend and multiplicative seasonality...