Lecture-18
Course: Applied Data Science

Exponential Smoothing

By
Dr. Sibarama Panigrahi
Assistant Professor, Department of Computer Sc. & Engineering
National Institute of Technology, Rourkela, Odisha, 769008, India
Mobile No.: +91-7377302566
Email: panigrahis[at]nitrkl[dot]ac[dot]in
panigrahi[dot]sibarama[at]gmail[dot]com

Exponential Smoothing

- Exponential smoothing was proposed in the late 1950s (Brown, 1959; Holt, 1957; Winters, 1960).

- Forecasts produced using exponential smoothing methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older.

- In other words, the more recent the observation the higher the associated weight.

Brown, R. G. (1959). *Statistical forecasting for inventory control*. McGraw/Hill.
Holt, C. C. (1957). *Forecasting seasonals and trends by exponentially weighted averages* (ONR Memorandum No. 52). Carnegie Institute of Technology, Pittsburgh USA. Reprinted in the *International Journal of Forecasting*, 2004.
Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, *6*(3), 324–342.

Simple Exponential Smoothing

- The simplest of the exponentially smoothing methods is naturally called **simple exponential smoothing** (SES).
- This method is suitable for forecasting data with no clear trend or seasonal pattern.
- The naïve and the average as possible methods for forecasting such data.

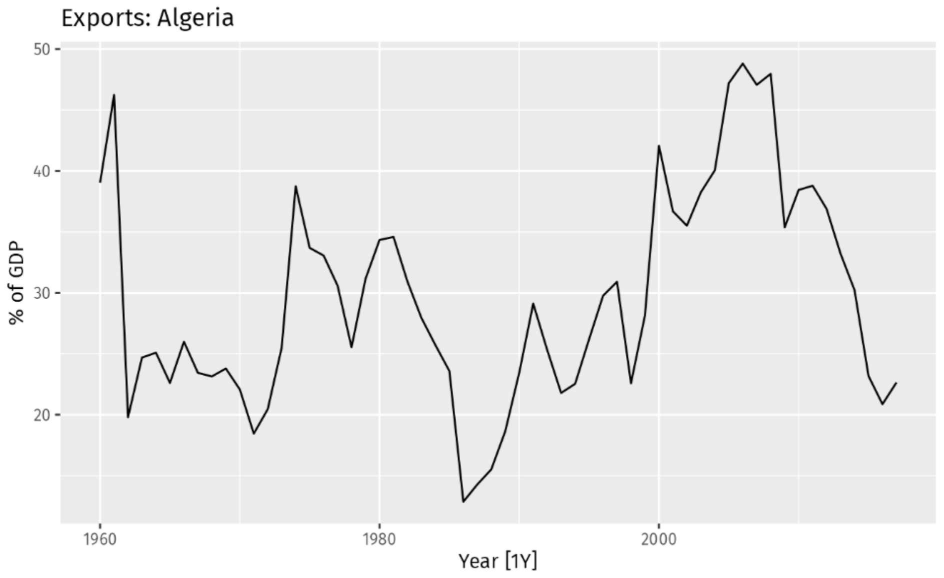


Figure 8.1: Exports of goods and services from Algeria from 1960 to 2017.

Using the naïve method, all forecasts for the future are equal to the last observed value of the series,

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

Using the average method, all future forecasts are equal to a simple average of the observed data,

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

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Simple Exponential Smoothing

- The simplest of the exponentially smoothing methods is naturally called **simple exponential smoothing** (SES).
- This method is suitable for forecasting data with no clear trend or seasonal pattern.
- The naïve and the average as possible methods for forecasting such data.

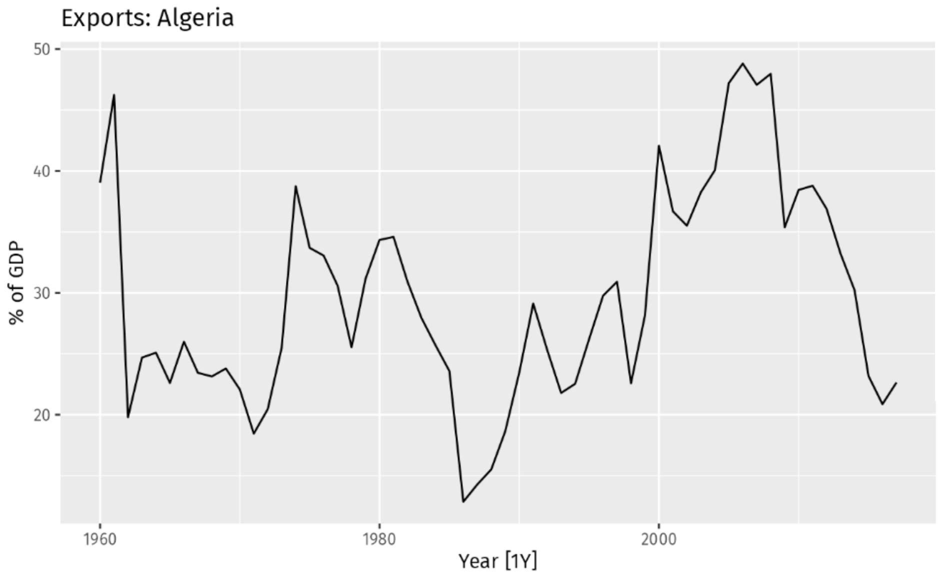


Figure 8.1: Exports of goods and services from Algeria from 1960 to 2017.

We often want something between these two extremes. For example, it may be sensible to attach larger weights to more recent observations than to observations from the distant past. This is exactly the concept behind simple exponential smoothing. Forecasts are calculated using weighted averages, where the weights decrease exponentially as observations come from further in the past — the smallest weights are associated with the oldest observations:

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha) y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots, \tag{8.1}$$

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

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 $0 \le \alpha \le 1$.

Observation	Weights assigned to observations for: $\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$
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}	$(0.2)(0.8)^5$	$(0.4)(0.6)^5$	$(0.6)(0.4)^5$	$(0.8)(0.2)^5$

For any α between 0 and 1, the weights attached to the observations decrease exponentially as we go back in time, hence the name "exponential smoothing". If α is small (i.e., close to 0), more weight is given to observations from the more distant past. If α is large (i.e., close to 1), more weight is given to the more recent observations. For the extreme case where $\alpha = 1$, $\hat{y}_{T+1|T} = y_T$, and the forecasts are equal to the naïve forecasts.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Simple Exponential Smoothing

Component form

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

- ℓ_t is the level (or the smoothed value) of the series at time t.
- $\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$
- $\hat{y}_{T+h|T} = \ell_T$, $h = 2, 3, \ldots$

Iterate to get exponentially weighted moving average form.

Weighted average form

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

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Simple Exponential Smoothing

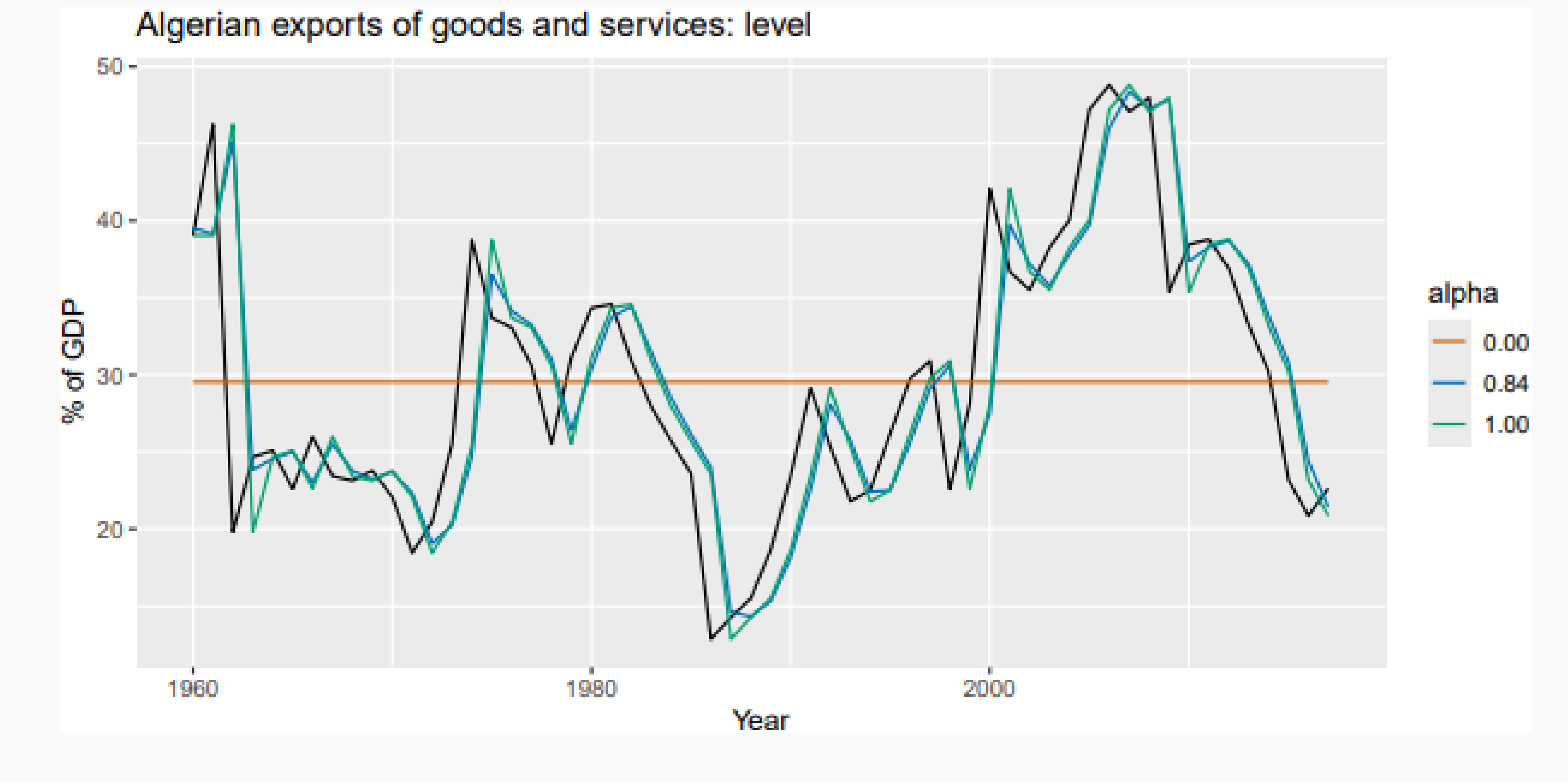
- **Optimising smoothing parameters**
  - Need to choose best values for α and ℓ_0.
  - Similarly to regression, choose optimal parameters by minimising SSE:

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

  - Unlike regression there is no closed form solution — use numerical optimization.
  - For Algerian Exports example:
    - $\hat{\alpha} = 0.8400$
    - $\hat{\ell}_0 = 39.54$

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Simple Exponential Smoothing



Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Holt's linear trend

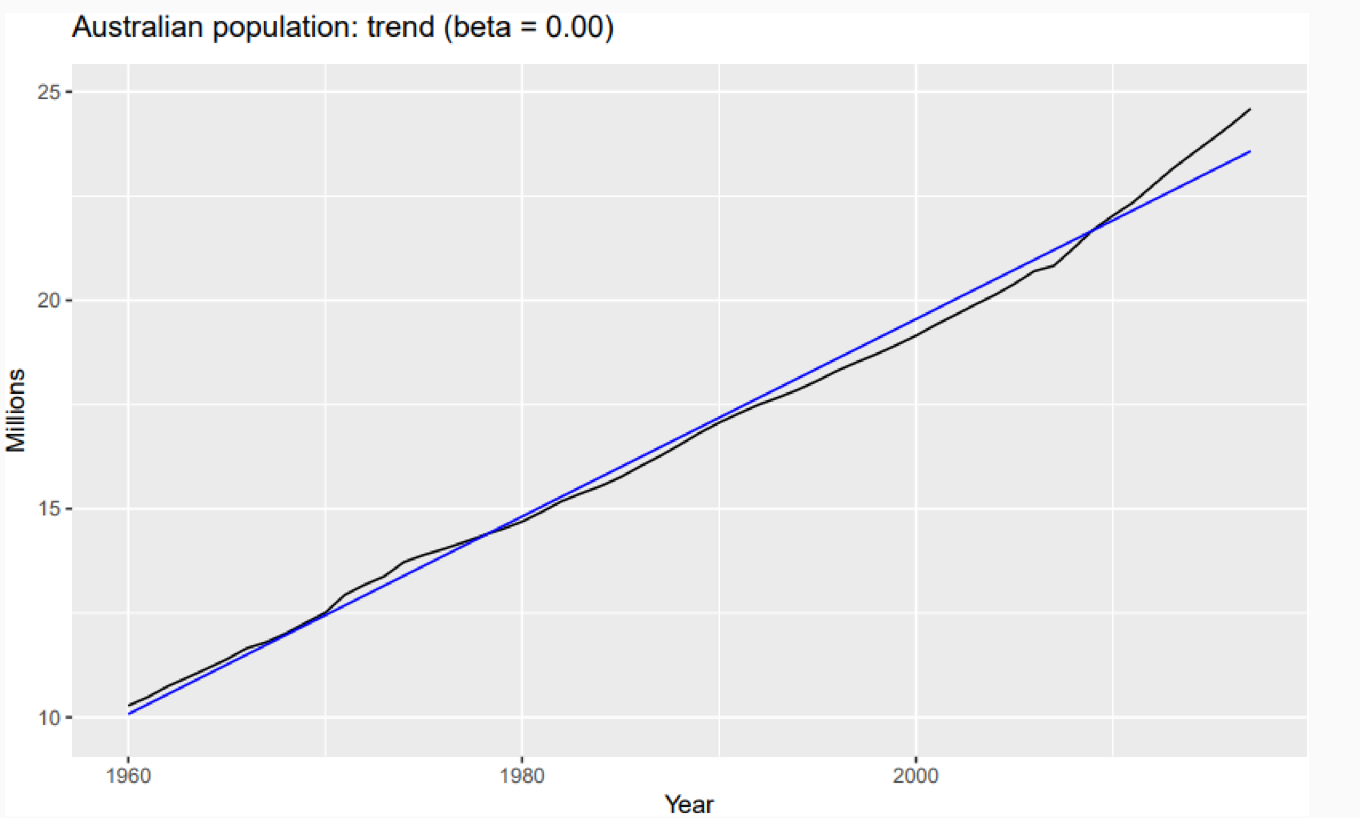
Component form

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

- Two smoothing parameters α and β^* $(0 \leq \alpha, \beta^* \leq 1)$.
- ℓ_t level: weighted average between y_t and one-step ahead forecast for time t, $(\ell_{t-1} + b_{t-1} = \hat{y}_{t|t-1})$
- b_t slope: weighted average of $(\ell_t - \ell_{t-1})$ and b_{t-1}, current and previous estimate of slope.
- Choose $\alpha, \beta^*, \ell_0, b_0$ to minimise SSE.

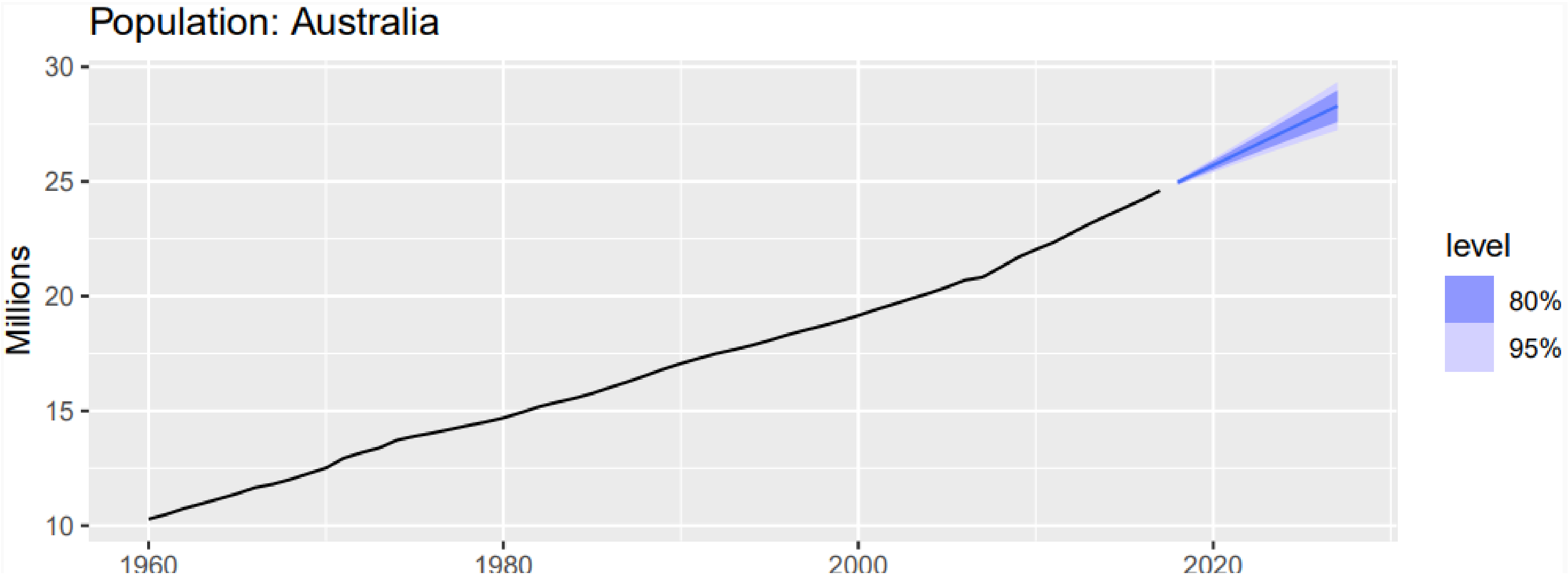
Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Holt’s linear trend



Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Holt’s linear trend



Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Damped trend method

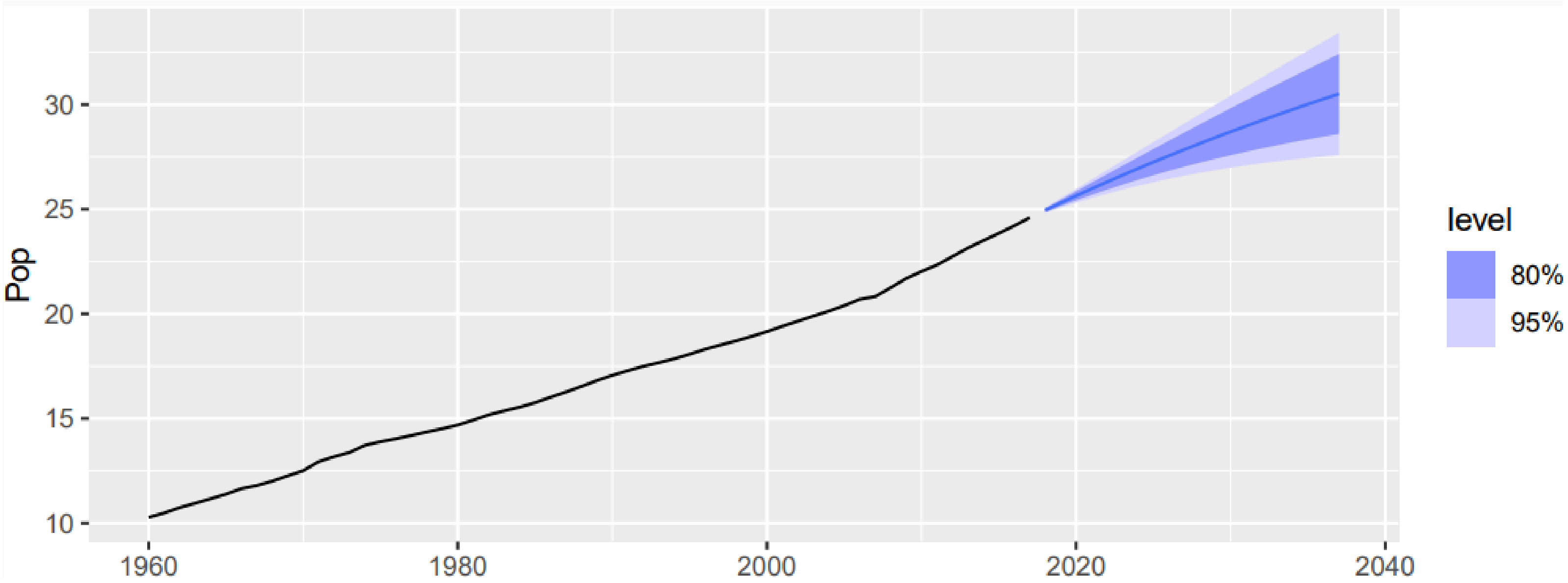
Component form

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

- Damping parameter $0 < \phi < 1$.
- If $\phi = 1$, identical to Holt's linear trend.
- As $h \to \infty$, $\hat{y}_{T+h|T} \to \ell_T + \phi b_T/(1-\phi)$.
- Short-run forecasts trended, long-run forecasts constant.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Damped trend method



Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Comparison: Australian Population

term	SES	Linear trend	Damped trend
α	1.00	1.00	1.00
β^*		0.30	0.40
ϕ			0.98
ℓ_0	10.28	10.05	10.04
b_0		0.22	0.25
Training RMSE	0.24	0.06	0.07
Test RMSE	1.63	0.15	0.21
Test MASE	6.18	0.55	0.75
Test MAPE	6.09	0.55	0.74
Test MAE	1.45	0.13	0.18

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Holt-Winters additive method

Holt and Winters extended Holt's method to capture seasonality.

Component form

$$\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$$

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

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

$$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$$

- k = integer part of $(h-1)/m$. Ensures estimates from the final year are used for forecasting.
- Parameters: $0 \le \alpha \le 1$, $0 \le \beta^* \le 1$, $0 \le \gamma \le 1-\alpha$ and m = period of seasonality (e.g. $m = 4$ for quarterly data).

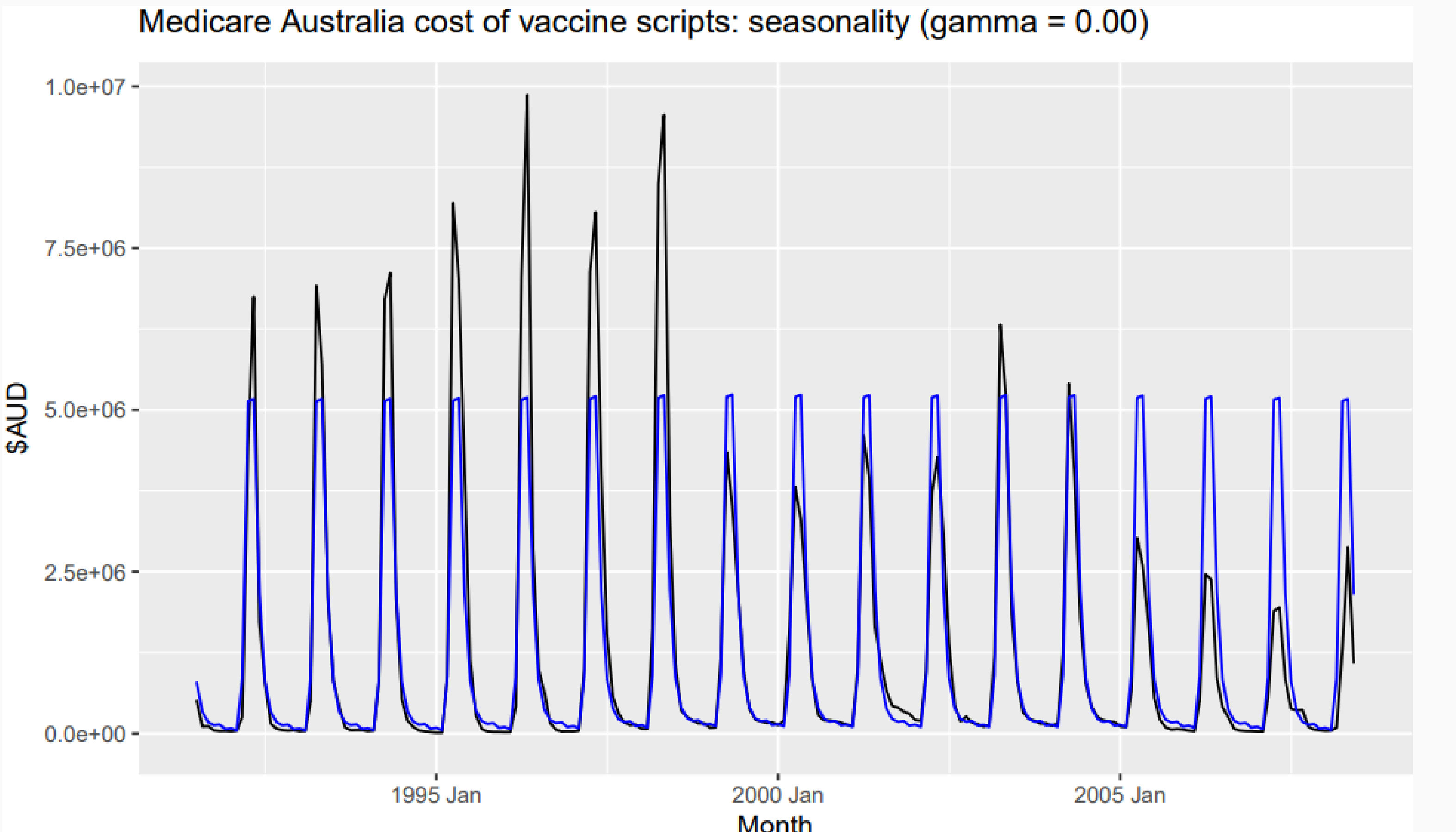
Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Holt-Winters additive method

- Seasonal component is usually expressed as
  $s_t = \gamma^*(y_t - \ell_t) + (1 - \gamma^*)s_{t-m}.$
- Substitute in for ℓ_t:
  $s_t = \gamma^*(1 - \alpha)(y_t - \ell_{t-1} - b_{t-1}) + [1 - \gamma^*(1 - \alpha)]s_{t-m}$
- We set $\gamma = \gamma^*(1 - \alpha)$.
- The usual parameter restriction is $0 \leq \gamma^* \leq 1$, which translates to $0 \leq \gamma \leq (1 - \alpha)$.

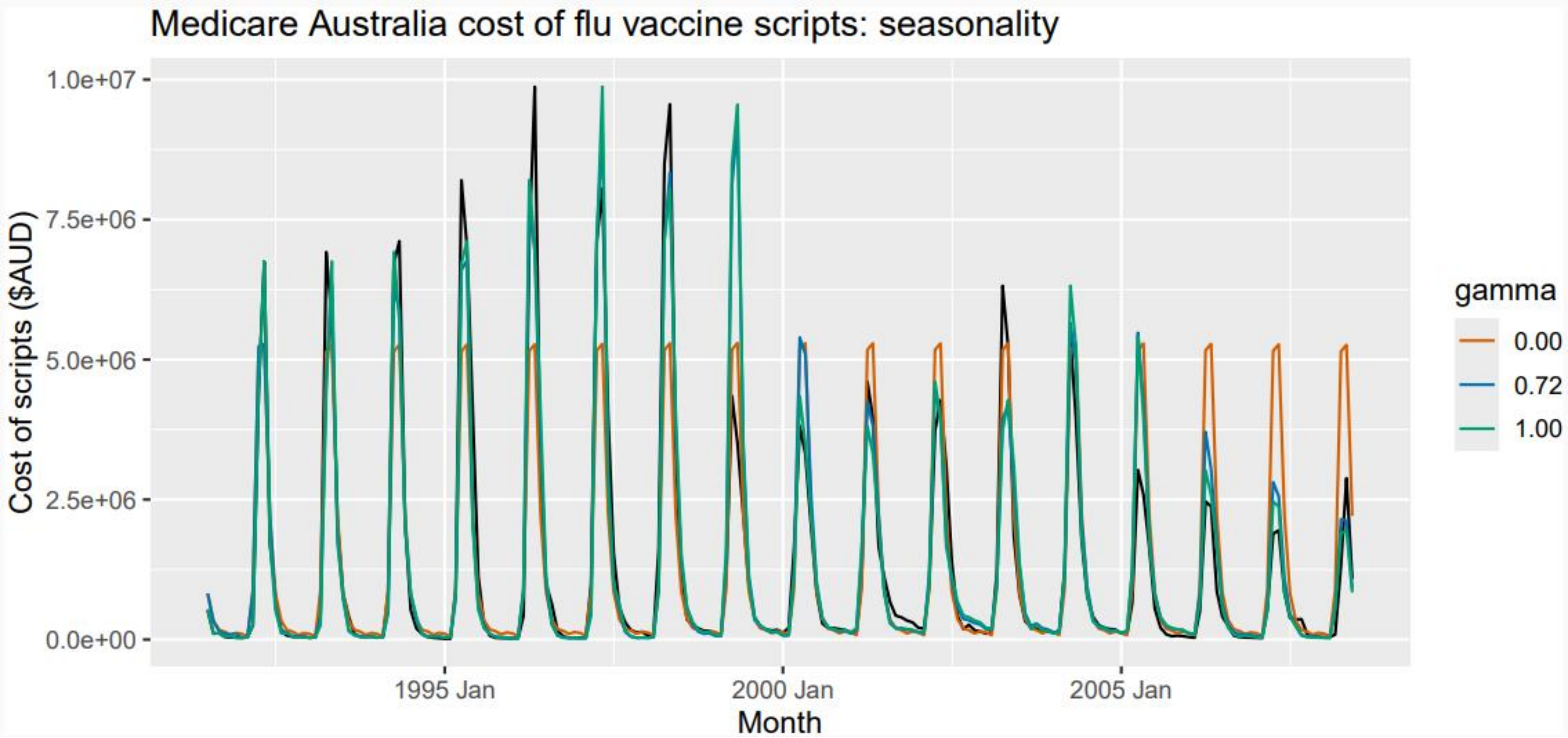
Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Exponential smoothing: seasonality



Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Exponential smoothing: seasonality



Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Holt-Winters multiplicative method

Seasonal variations change in proportion to the level of the series.

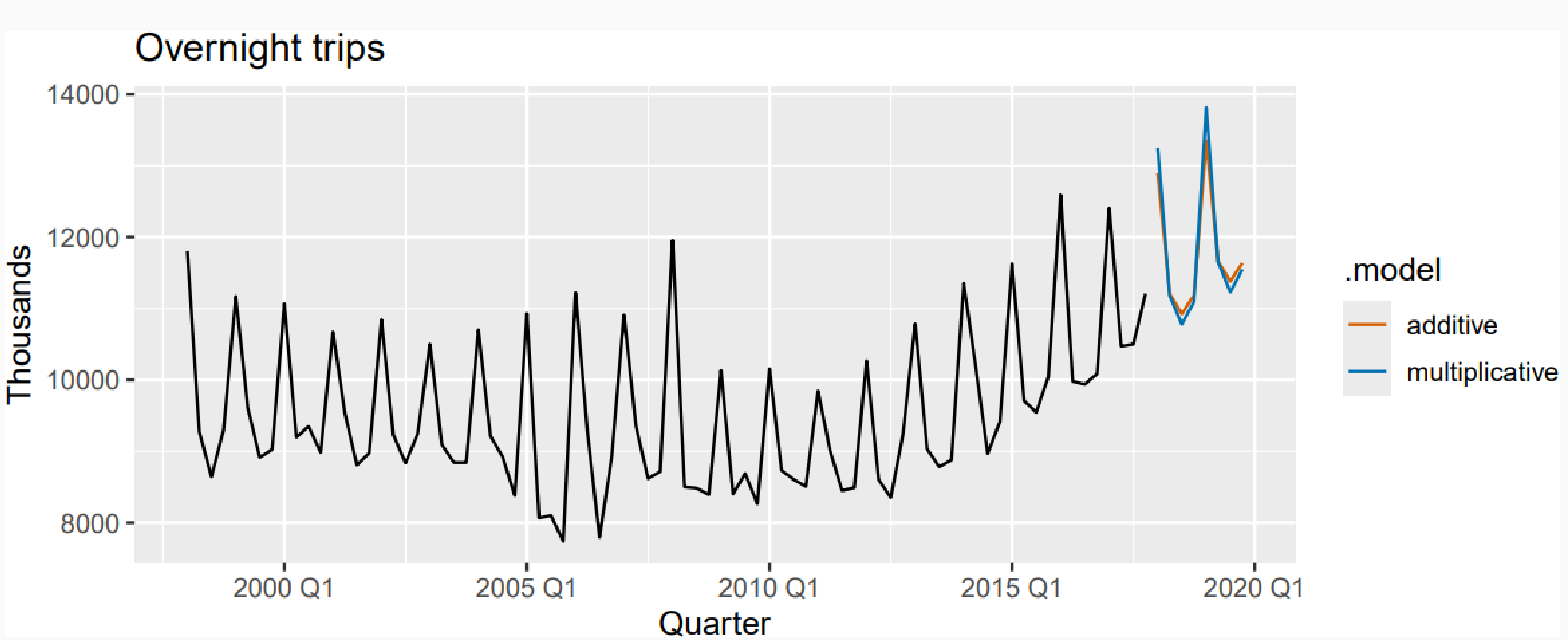
Component form

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

- k is integer part of $(h-1)/m$.
- Additive method: s_t in absolute terms — within each year $\sum_i s_i \approx 0$.
- Multiplicative method: s_t in relative terms — within each year $\sum_i s_i \approx m$.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Example: Australian holiday tourism



Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Holt-Winters damped method

Often the single most accurate forecasting method for seasonal data:

$$\hat{y}_{t+h|t} = [\ell_t + (\phi + \phi^2 + \cdots + \phi^h)b_t]s_{t+h-m(k+1)}$$

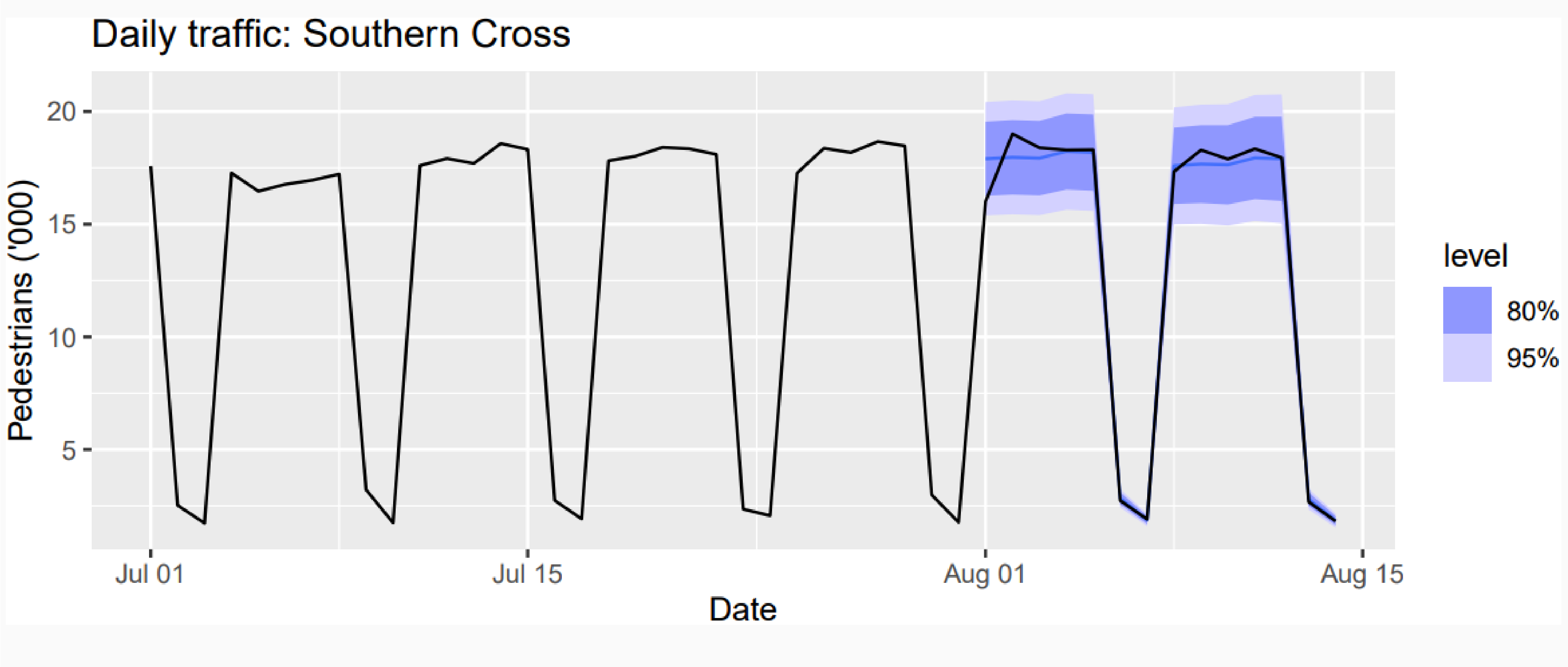
$$\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$$

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

$$s_t = \gamma \frac{y_t}{(\ell_{t-1} + \phi b_{t-1})} + (1-\gamma)s_{t-m}$$

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Holt-Winters with daily data



Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Exponential smoothing methods

Trend Component		Seasonal Component: N (None)	A (Additive)	M (Multiplicative)
N	(None)	(N,N)	(N,A)	(N,M)
A	(Additive)	(A,N)	(A,A)	(A,M)
A_d	(Additive damped)	(A_d,N)	(A_d,A)	(A_d,M)

(N,N): Simple exponential smoothing
(A,N): Holt's linear method
(A_d,N): Additive damped trend method
(A,A): Additive Holt-Winters' method
(A,M): Multiplicative Holt-Winters' method
(A_d,M): Damped multiplicative Holt-Winters' method

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Exponential smoothing methods

Trend	Seasonal N	Seasonal A	Seasonal M
N	$\hat{y}_{t+h\|t} = \ell_t$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t / \ell_{t-1}) + (1-\gamma)s_{t-m}$
A	$\hat{y}_{t+h\|t} = \ell_t + hb_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + hb_t)s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t / (\ell_{t-1} + b_{t-1})) + (1-\gamma)s_{t-m}$
A_d	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + \phi_h b_t)s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t / (\ell_{t-1} + \phi b_{t-1})) + (1-\gamma)s_{t-m}$

$\phi_h = \phi + \phi^2 + \cdots + \phi^h$

k is the integer part of $(h-1)/m$

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Thank You

For Your Valuable Time.