

Time Series Analysis & Forecasting Using R

Exponential smoothing models

Outline

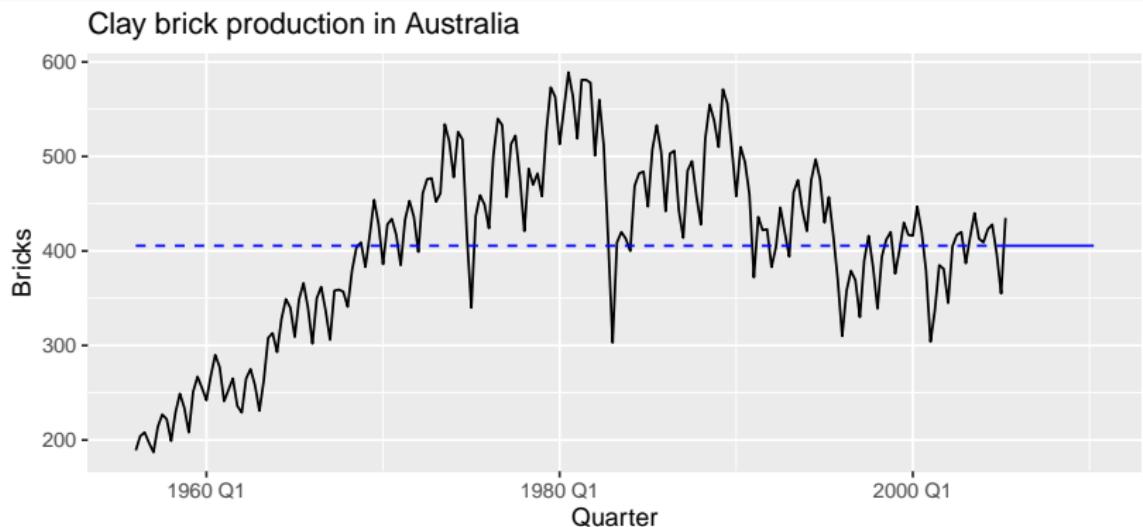
- 1 Exponential Smoothing methods
- 2 Exponential smoothing state space models
- 3 Models with no trend, no seasonality
- 4 Models with trend
- 5 Models with seasonality

Outline

- 1 Exponential Smoothing methods
- 2 Exponential smoothing state space models
- 3 Models with no trend, no seasonality
- 4 Models with trend
- 5 Models with seasonality

Average method

- Forecast of all future values is equal to mean of historical data $\{y_1, \dots, y_T\}$.
 - ▶ Forecasts: $\hat{y}_{T+h|T} = \bar{y} = (y_1 + \dots + y_T)/T$



Average forecasts

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

Average forecasts

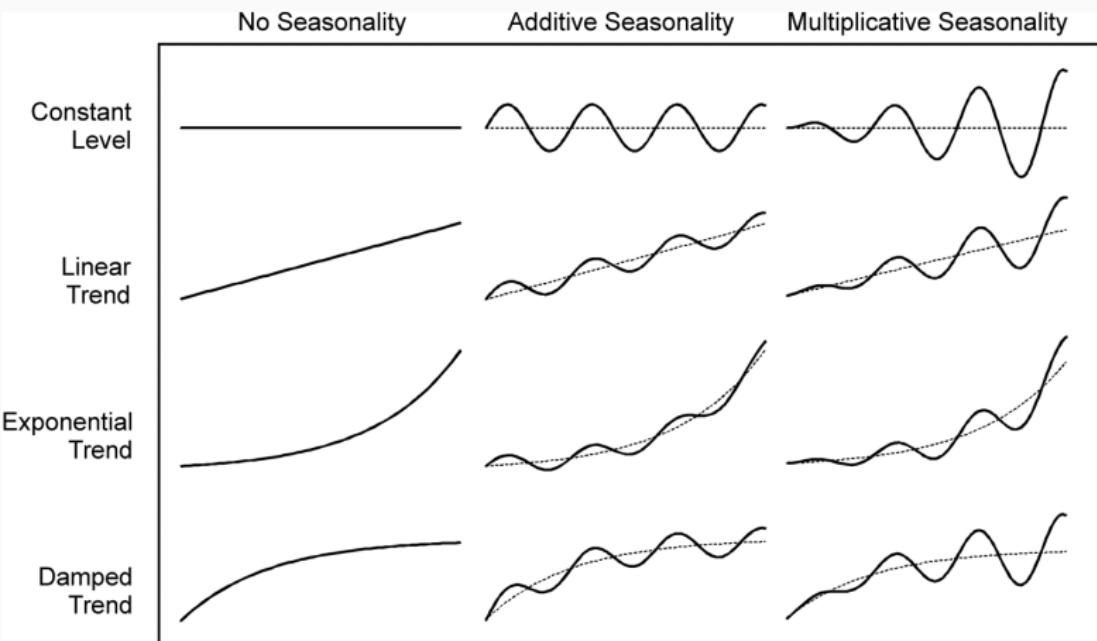
$$\hat{y}_{T+h|T} = \frac{1}{T} \sum_{t=1}^T y_t$$

- Want something in between these methods.
- Most recent data should have more weight.
- Trend and seasonality

Historical perspective

- Developed in the 1950s and 1960s as **methods** (algorithms) to produce point forecasts.
- Combine a “level”, “trend” (slope) and “seasonal” component to describe a time series.
- The rate of change of the components are controlled by “smoothing parameters”: α , β and γ respectively.
- Need to choose best values for the smoothing parameters (and initial states).
- Equivalent ETS state space **models** developed in the 1990s and 2000s.

Pegel's classification



Exponential smoothing methods

		Seasonal Component		
		N	A	M
		(None)	(Additive)	(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

Capturing levels, trends, and seasonalities

We want a method that captures the level (ℓ_t), trend (b_t) and seasonality (s_t).

How do we combine these elements?

Capturing levels, trends, and seasonalities

We want a method that captures the level (ℓ_t), trend (b_t) and seasonality (s_t).

How do we combine these elements?

Additively?

$$y_t = \ell_{t-1} + b_{t-1} + s_{t-m}$$

Capturing levels, trends, and seasonalities

We want a method that captures the level (ℓ_t), trend (b_t) and seasonality (s_t).

How do we combine these elements?

Additively?

$$y_t = \ell_{t-1} + b_{t-1} + s_{t-m}$$

Multiplicatively?

$$y_t = \ell_{t-1} b_{t-1} s_{t-m}$$

Capturing levels, trends, and seasonalities

We want a method that captures the level (ℓ_t), trend (b_t) and seasonality (s_t).

How do we combine these elements?

Additively?

$$y_t = \ell_{t-1} + b_{t-1} + s_{t-m}$$

Multiplicatively?

$$y_t = \ell_{t-1} b_{t-1} s_{t-m}$$

Perhaps a mix of both?

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

Exponential smoothing methods

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

Exponential smoothing methods

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

How do the level, trend and seasonal components evolve over time?

Big idea: control the rate of change (smoothing)

α controls the flexibility of the **level**

- If $\alpha = 0$, the level never updates (mean)
- If $\alpha = 1$, the level updates completely (naive)

β controls the flexibility of the **trend**

- If $\beta = 0$, the trend is linear (regression trend)
- If $\beta = 1$, the trend updates every observation

γ controls the flexibility of the **seasonality**

- If $\gamma = 0$, the seasonality is fixed (seasonal means)
- If $\gamma = 1$, the seasonality updates completely (seasonal naive)

Parameter estimation

- Need to choose value for smoothing and initial values.
- A robust and objective way to obtain values for the unknown parameters included in any exponential smoothing method is to estimate them from the observed data.
- Initial and smoothing values are chosen by minimising sum of squared errors (SSE):

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

Outline

- 1 Exponential Smoothing methods
- 2 Exponential smoothing state space models
- 3 Models with no trend, no seasonality
- 4 Models with trend
- 5 Models with seasonality

Exponential smoothing state space models

methods

- ES methods presented so far are algorithms that generate **point forecast**.
- Each exponential smoothing method can be written as an **Innovation state space model**

models

- Generate same point forecasts but can also generate forecast intervals.
- A stochastic (or random) data generating process that can generate an entire forecast distribution.
- Allow for “proper” model selection. Parameters are estimated using maximizing likelihood, the

Exponential smoothing methods

		Seasonal Component		
		N	A	M
		(None)	(Additive)	(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

ETS models

- Two models for each method: one with additive and one with multiplicative errors (multiplicative error means the noise increases with level of series)
- The possibilities for each component are:
 - ▶ Error ={A,M}
 - ▶ Trend ={N,A, A_d ,M, M_d }
 - ▶ Seasonal ={N,A,M} where N =none, A = additive, M = multiplicative, and d =damped.
- Each state space model can be labeled as ETS (Error, Trend, Seasonal).

ETS models

- Each model consists of an *observation* equation that describes the observed data and *transition*, one for each state (level, trend, seasonal), i.e., state space models, that describe how the unobserved components or states change over time. Hence these are referred to as *state space models*.

ETS taxonomy

General notation

ETS : ExponenTial Smoothing

The diagram shows the acronym 'ETS' followed by a colon and the words 'ExponenTial Smoothing'. Above the colon, there are three arrows pointing downwards from the letters 'E', 'T', and 'S' respectively, towards the words 'Error', 'Trend', and 'Season' which are positioned below the acronym.

Error Trend Season

Error: Additive ("A") or multiplicative ("M")

ETS taxonomy

General notation

ETS : ExponenTial Smoothing
Error Trend Season

Error: Additive ("A") or multiplicative ("M")

Trend: None ("N"), additive ("A"), multiplicative ("M"), or damped ("Ad" or "Md").

ETS taxonomy

General notation

ETS : ExponenTial Smoothing
Error Trend Season

Error: Additive ("A") or multiplicative ("M")

Trend: None ("N"), additive ("A"), multiplicative ("M"), or damped ("Ad" or "Md").

Seasonality: None ("N"), additive ("A") or multiplicative ("M")

ETS taxonomy

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

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

Additive error models

Trend	Seasonal		
	N	A	M
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = \ell_{t-1} s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / \ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + b_{t-1})$
A _d	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = \phi b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + \phi b_{t-1})$

Multiplicative error models

Trend	Seasonal		
	N	A	M
N	$y_t = \ell_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A _d	$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

Innovations state space models

Let $\mathbf{x}_t = (\ell_t, b_t, s_t, s_{t-1}, \dots, s_{t-m+1})$ and $\varepsilon_t \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$.

$$y_t = \underbrace{h(\mathbf{x}_{t-1})}_{\mu_t} + \underbrace{k(\mathbf{x}_{t-1})\varepsilon_t}_{e_t}$$

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}) + g(\mathbf{x}_{t-1})\varepsilon_t$$

Additive errors $K(x) = 1$. $y_t = \mu_t + \varepsilon_t$.

Multiplicative errors $K(\mathbf{x}_{t-1}) = \mu_t$. $y_t = \mu_t(1 + \varepsilon_t)$.

$\varepsilon_t = (y_t - \mu_t)/\mu_t$ is relative error.

Innovations state space models

Estimation

$$\begin{aligned} L^*(\theta, \mathbf{x}_0) &= n \log \left(\sum_{t=1}^n \varepsilon_t^2 / k^2(\mathbf{x}_{t-1}) \right) + 2 \sum_{t=1}^n \log |k(\mathbf{x}_{t-1})| \\ &= -2 \log(\text{Likelihood}) + \text{constant} \end{aligned}$$

Estimating ETS models

- Smoothing parameters α, β, γ and ϕ , and the initial states $\ell_0, b_0, s_0, s_{-1}, \dots, s_{-m+1}$ are estimated by maximising the “likelihood” = the probability of the data arising from the specified model.
- For models with additive errors equivalent to minimising SSE.
- For models with multiplicative errors, **not** equivalent to minimising SSE.

Model selection

Akaike's Information Criterion

$$AIC = -2 \log(L) + 2k$$

where L is the likelihood and k is the number of parameters initial states estimated in the model.

Model selection

Akaike's Information Criterion

$$AIC = -2 \log(L) + 2k$$

where L is the likelihood and k is the number of parameters initial states estimated in the model.

Corrected AIC

$$AIC_c = AIC + \frac{2(k+1)(k+2)}{T-k}$$

which is the AIC corrected (for small sample bias).

Model selection

Akaike's Information Criterion

$$AIC = -2 \log(L) + 2k$$

where L is the likelihood and k is the number of parameters initial states estimated in the model.

Corrected AIC

$$AIC_c = AIC + \frac{2(k+1)(k+2)}{T-k}$$

which is the AIC corrected (for small sample bias).

Bayesian Information Criterion

$$BIC = AIC + k(\log(T) - 2).$$

ETS(A,N,N): Specifying the model

```
ETS(y ~ error("A") + trend("N") + season("N"))
```

Outline

- 1 Exponential Smoothing methods
- 2 Exponential smoothing state space models
- 3 Models with no trend, no seasonality
- 4 Models with trend
- 5 Models with seasonality

ETS(A,N,N): SES with additive errors

Forecast equation $\hat{y}_{T+h|T} = \ell_T$

Measurement equation $y_t = \ell_{t-1} + \varepsilon_t$

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

where $\varepsilon_t \sim \text{NID}(0, \sigma^2)$.

ETS(A,N,N): SES with additive errors

Forecast equation $\hat{y}_{T+h|T} = \ell_T$

Measurement equation $y_t = \ell_{t-1} + \varepsilon_t$

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

where $\varepsilon_t \sim \text{NID}(0, \sigma^2)$.

- “innovations” or “single source of error” because equations have the same error process, ε_t .
- Measurement equation: relationship between observations and states.
- Transition/state equation(s): evolution of the state(s) through time.

ETS(M,N,N): SES with multiplicative errors

Forecast equation $\hat{y}_{T+h|T} = \ell_T$

Measurement equation $y_t = \ell_{t-1}(1 + \varepsilon_t)$

State equation $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$

where $\varepsilon_t \sim \text{NID}(0, \sigma^2)$.

ETS(M,N,N): SES with multiplicative errors

Forecast equation $\hat{y}_{T+h|T} = \ell_T$

Measurement equation $y_t = \ell_{t-1}(1 + \varepsilon_t)$

State equation $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$

where $\varepsilon_t \sim \text{NID}(0, \sigma^2)$.

- Models with additive and multiplicative errors with the same parameters generate the same point forecasts but different prediction intervals.

Outline

- 1 Exponential Smoothing methods**
- 2 Exponential smoothing state space models**
- 3 Models with no trend, no seasonality**
- 4 Models with trend**
- 5 Models with seasonality**

Holt's linear trend

Additive errors: ETS(A,A,N)

Forecast equation $\hat{y}_{T+h|T} = \ell_T + h b_T$

Measurement equation $y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$

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

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

Holt's linear trend

Additive errors: ETS(A,A,N)

Forecast equation $\hat{y}_{T+h|T} = \ell_T + h b_T$

Measurement equation $y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$

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

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

Multiplicative errors: ETS(M,A,N)

Forecast equation $\hat{y}_{T+h|T} = \ell_T + h b_T$

Measurement equation $y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$

State equations $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha \varepsilon_t)$

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

Damped trend additive

- Holt's linear model displays a constant trend (increasing or decreasing) indefinitely into the future.
- The exponential trend model also includes exponential growth or decline.
- Empirical evidence indicates that these methods tend to over-forecast, especially for longer forecast horizons.
- Motivated by this, a *dampening* parameter is introduced so that the trend approaches a flat line some time in the future.

Damped trend additive

Component form

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

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

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

Damped trend additive

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 $h \rightarrow \infty$, $\hat{y}_{T+h|T} \rightarrow \ell_T + \phi b_T / (1 - \phi)$.
- Short-run forecasts trended, long-run forecasts constant.

ETS(A,Ad,N): Damped trend multiplicative

Additive errors

Forecast equation $\hat{y}_{T+h|T} = \ell_T + b_T^l \phi + \dots + \phi^{h-1})$

Measurement equation $y_t = (\ell_{t-1} + \phi b_{t-1}) + (1 + \varepsilon_t)$

State equations $\ell_t = (\ell_{t-1} + \phi b_{t-1}) + (1 + \alpha \varepsilon_t)$

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

ETS(A,Ad,N): Damped trend multiplicative

Additive errors

Forecast equation $\hat{y}_{T+h|T} = \ell_T + b_T^l \phi + \dots + \phi^{h-1})$

Measurement equation $y_t = (\ell_{t-1} + \phi b_{t-1}) + (1 + \varepsilon_t)$

State equations $\ell_t = (\ell_{t-1} + \phi b_{t-1}) + (1 + \alpha \varepsilon_t)$

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

Outline

- 1 Exponential Smoothing methods**
- 2 Exponential smoothing state space models**
- 3 Models with no trend, no seasonality**
- 4 Models with trend**
- 5 Models with seasonality**

Holt-Winters method

- Holt and Winters extended Holt's method to capture seasonality.
- There are two variations to this method that differ in the nature of the seasonal component: -
Additive : The additive method is preferred when the seasonal variations are roughly constant through the series, - **Multiplicative** : the multiplicative method is preferred when the seasonal variations are changing proportional to the level of the series.

ETS(A,A,A): Holt-Winters additive

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

Observation equation $y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$

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

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

$s_t = s_{t-m} + \gamma \varepsilon_t$

- $k = \text{integer part of } (h - 1)/m.$
- $\sum_i s_i \approx 0.$
- Parameters: $0 \leq \alpha \leq 1$, $0 \leq \beta^* \leq 1$, $0 \leq \gamma \leq 1 - \alpha$ and $m = \text{period of seasonality}$ (e.g. $m = 4$ for quarterly data).

ETS(M,A,M): Holt-Winters multiplicative

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

Observation equation $y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$

State equations $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$

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

$s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

- k is integer part of $(h - 1)/m$.
- $\sum_i s_i \approx m$.
- Parameters: $0 \leq \alpha \leq 1$, $0 \leq \beta^* \leq 1$, $0 \leq \gamma \leq 1 - \alpha$ and m = period of seasonality (e.g. $m = 4$ for quarterly data).

Holt-Winters damped method

Often the single most accurate forecasting method for seasonal data:

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

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

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

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