MONASH University

MONASH BUSINESS SCHOOL

ETF3231/5231
Business forecasting

Week 6: Exponential smoothing

`https://bf.numbat.space/`

AACSB ACCREDITED EQUIS ACCREDITED AMBA ACCREDITED

Outline

Outline

ETS models

General notation E T S : **E**xponen**T**ial **S**moothing

Error **T**rend **S**eason

```
ETS(y ~ error( ) + trend( ) + season( ))
```

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

ETS models

General notation E T S : **E**xponen**T**ial **S**moothing

Error Trend Season

```
ETS(y ~ error( ) + trend( ) + season( ))
```

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

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

ETS models

General notation E T S : **E**xponen**T**ial **S**moothing

Error Trend Season

```
ETS(y ~ 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(A,N,N): SES with additive errors

Observation 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.
- State equation(s): evolution of the state(s) through time.

start from $\hat{y}_{T+1|T} = \alpha y_T + (1-\alpha)\, \hat{y}_{T|T-1}$

$$= \alpha y_T + (1-\alpha)\left[\alpha y_{T-1} + (1-\alpha)\, \hat{y}_{T-1|T-2}\right]$$

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

$$= \alpha y_T + \alpha(1-\alpha)\, y_{T-1} + (1-\alpha)^2\left[\alpha y_{T-2} + (1-\alpha)\, \hat{y}_{T-2|T-3}\right]$$

$$= \alpha y_T + \alpha(1-\alpha)\, y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + (1-\alpha)^3\, \hat{y}_{T-2|T-3}$$

$$\vdots$$

$$= \alpha y_T + \alpha(1-\alpha)\, y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \ldots + (1-\alpha)^T \ell_0$$

we don't have infinite data.

when $\alpha = 1$ $\quad \hat{y}_{T+1|T} = y_T$ $\quad \rightarrow$ only last obs matters

$\alpha = 0$ $\quad \hat{y}_{T+1|T} = \ell_0$ $\quad \rightarrow$ we learn nothing from new info

ETS (A, N, N) setting $\alpha = 1$

$$y_t = l_{t-1} + \varepsilon_t$$

$$l_t = l_{t-1} + \alpha \varepsilon_t$$ let $\alpha = 1$

ETS (A, N, N) setting $\alpha = 1$

$y_t = l_{t-1} + \varepsilon_t$

$l_t = l_{t-1} + \cancel{\alpha}^{1}\, \varepsilon_t$

let $\alpha = 1$

$$\left.\begin{aligned} y_t &= l_{t-1} + \varepsilon_t \\ l_t &= l_{t-1} + \varepsilon_t \end{aligned}\right\| \Rightarrow \quad y_t = l_t \Rightarrow y_{t-1} = l_{t-1}$$

$\begin{pmatrix}\text{Sub in} \\ \text{Obs eqn}\end{pmatrix}$ $y_t = y_{t-1} + \varepsilon_t$ $\varepsilon_t \sim NID(0, \sigma^2)$ Random Walk model

ETS(A,A,N)

Holt's methods method with additive errors.

Forecast equation	$\hat{y}_{t+h\|t} = \ell_t + hb_t$
Observation equation	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$
State equations	$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$
	$b_t = b_{t-1} + \beta\varepsilon_t$

- Forecast errors: $\varepsilon_t = y_t - \hat{y}_{t|t-1}$

$$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$$
$$b_t = b_{t-1} + \beta\varepsilon_t$$

Three interesting cases

(1) $\beta = 0$ $\quad b_t = b_{t-1} = \cdots = b$ slope is constant - all changes through ℓ_t

(2) $\alpha = 0 \Rightarrow \beta = 0 \Rightarrow$ hence $b_t = b_{t-1} = \cdots = b$ slope not changing

also $\ell_t = \ell_{t-1} + b_{t-1}$ level not changing

$\ell_0 + b_0$ becomes important.

Three interesting cases

$$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$$
$$b_t = b_{t-1} + \beta\varepsilon_t$$

(1) $\beta = 0$ $\quad b_t = b_{t-1} = \dots = b$ slope is constant - all changes through ℓ_t

(2) $\alpha = 0 \Rightarrow \beta = 0 \Rightarrow$ hence $b_t = b_{t-1} = \dots = b$ slope not changing

also $\ell_t = \ell_{t-1} + b_{t-1}$ level not changing

$\ell_0 + b_0$ becomes important.

(3) $\beta = 0$ (slope not changing b) and $\alpha = 1$

Three interesting cases

$$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$$
$$b_t = b_{t-1} + \beta\varepsilon_t$$

(1) $\beta = 0$ $\quad b_t = b_{t-1} = \dots = b$ slope is constant - all changes through ℓ_t

(2) $\alpha = 0 \Rightarrow \beta = 0 \Rightarrow$ hence $b_t = b_{t-1} = \dots = b$ slope not changing

also $\ell_t = \ell_{t-1} + b_{t-1}$ level not changing

$\ell_0 + b_0$ becomes important.

(3) $\beta = 0$ (slope not changing b) and $\alpha = 1$

$$\left.\begin{aligned} y_t &= \ell_{t-1} + b_{t-1} + \varepsilon_t \\ \ell_t &= \ell_{t-1} + b_{t-1} + \varepsilon_t \end{aligned}\right\| \Rightarrow y_t = \ell_t \Rightarrow y_{t-1} = \ell_{t-1}$$

Three interesting cases

$$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$$
$$b_t = b_{t-1} + \beta\varepsilon_t$$

(1) $\beta = 0$ $\quad b_t = b_{t-1} = \dots = b$ slope is constant - all changes through ℓ_t

(2) $\alpha = 0 \Rightarrow \beta = 0 \Rightarrow$ hence $b_t = b_{t-1} = \dots = b$ slope not changing

also $\ell_t = \ell_{t-1} + b_{t-1}$ level not changing

$\ell_0 + b_0$ becomes important.

(3) $\beta = 0$ (slope not changing b) ① and $\alpha = 1$

$$\left.\begin{aligned} y_t &= \ell_{t-1} + b_{t-1} + \varepsilon_t \\ \ell_t &= \ell_{t-1} + b_{t-1} + \varepsilon_t \end{aligned}\right\| \Rightarrow y_t = \ell_t \Rightarrow y_{t-1} = \ell_{t-1} \quad ②$$

Rough proof $\quad$ Sub ① & ② in Obs eqn $\quad y_t = b + y_{t-1} + \varepsilon_t$

Three interesting cases

$$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$$
$$b_t = b_{t-1} + \beta\varepsilon_t$$

(1) $\beta = 0$ $\quad b_t = b_{t-1} = \dots = b$ slope is constant - all changes through ℓ_t

(2) $\alpha = 0 \Rightarrow \beta = 0 \Rightarrow$ hence $b_t = b_{t-1} = \dots = b$ slope not changing

also $\ell_t = \ell_{t-1} + b_{t-1}$ level not changing

$\ell_0 + b_0$ becomes important.

(3) $\beta = 0$ (slope not changing b) ① and $\alpha = 1$

$$\left.\begin{aligned} y_t &= \ell_{t-1} + b_{t-1} + \varepsilon_t \\ \ell_t &= \ell_{t-1} + b_{t-1} + \varepsilon_t \end{aligned}\right\| \Rightarrow y_t = \ell_t \Rightarrow y_{t-1} = \ell_{t-1} \text{ — ②}$$

Rough proof $\quad$ Sub ① & ② in Obs eqn $\quad y_t = b + y_{t-1} + \varepsilon_t$

RANDOM WALK WITH DRIFT MODEL

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

To show that $b = \frac{1}{T-1} \sum_{t=2}^{T} (y_t - y_{t-1})^2$

- Consider residuals $\hat{e}_t = y_t - \hat{y}_{t|t-1} = y_t - y_{t-1} - b$ for $t = 2, \dots, T$

- Then min SSE w.r.t to b $\quad \frac{\partial}{\partial b} \sum_{t=2}^{T} (y_t - y_{t-1} - b)^2 = 0$

$$\Rightarrow \quad b = \frac{1}{T-1} \sum_{t=2}^{T} (y_t - y_{t-1})$$

ETS(A,A,A)

Holt-Winters additive method with additive errors.

Forecast equation	$\hat{y}_{t+h\|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$
Observation equation	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$
State equations	$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$
	$b_t = b_{t-1} + \beta\varepsilon_t$
	$s_t = s_{t-m} + \gamma\varepsilon_t$

- Innovation residuals: $\hat{\varepsilon}_t = y_t - \hat{y}_{t|t-1}$
- k is integer part of $(h-1)/m$.

ETS(M,A,M)

Holt-Winters multiplicative method with multiplicative errors.

Forecast equation	$\hat{y}_{t+h\|t} = (\ell_t + hb_t)s_{t+h-m(k+1)}$
	(linear trend) x Seasonal = multiplicative seas.
Observation equation	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$
State equations	$\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$
	$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$
	$s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

- Innovation residuals: $\hat{\varepsilon}_t = (y_t - \hat{y}_{t|t-1})/\hat{y}_{t|t-1}$
- k is integer part of $(h-1)/m$.

ETS(M,A,M)

Holt-Winters multiplicative method with multiplicative errors.

Forecast equation	$\hat{y}_{t+h\|t} = (\ell_t + hb_t)s_{t+h-m(k+1)}$
	(linear trend) x Seasonal = multiplicative seas.
Observation equation	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$
State equations	$\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$
	$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$
	$s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

Recall: we make assumptions about these

- Innovation residuals: $\hat{\varepsilon}_t = (y_t - \hat{y}_{t|t-1})/\hat{y}_{t|t-1}$
- k is integer part of $(h-1)/m$.

$e_t = \hat{\varepsilon}_t \hat{y}_{t|t-1}$ response resids are heteroscedastic

* hence, no need to transform.

ETS model specification

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

initial states

By default, optimal values for α, β, γ, and the states at time 0 are used.

The values for α, β and γ can be specified:

```
trend("A", alpha = 0.5, beta = 0.2)
trend("A", alpha_range = c(0.2, 0.8), beta_range = c(0.1, 0.4))
season("M", gamma = 0.04)
season("M", gamma_range = c(0, 0.3))
```

you probably never do this in practice but you can

Exponential smoothing methods (A taxonomy)

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

* In 1950's, 60's, no optim. so people chose parameters arbitrarily
* Looked at different combinations (James Taylor, Oxford, damped multi. trend, 2003)

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

There are also multiplicative trend methods (not recommended).

ETS models

* monash contribution
Rob, Ralph, co-authors
2000s

Additive Error

Trend Component		Seasonal 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

Trend Component		Seasonal 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

Exponential smoothing models

Additive Error

Trend Component		Seasonal 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~~

* So total 15 models

Multiplicative Error

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

Additive error models

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

Model selection

Akaike's Information Criterion

$$\text{AIC} = \overbrace{-2\log(\text{L})}^{\text{fit of the model}} + \underbrace{2k}_{\to \text{ penalty}}$$

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

Corrected AIC

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

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

Bayesian (Schwatz) Information Criterion

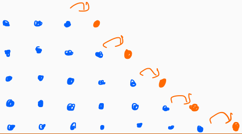
$$\text{BIC} = \text{AIC} + k[\log(T) - 2] = -2\log(\text{L}) + \underbrace{ln(T)}_{\to \text{ greater penalty for } T>8}k$$

AIC and cross-validation

MAGICAL RESULT*

Minimizing the AIC assuming Gaussian residuals is asymptotically equivalent to minimizing one-step time series cross validation MSE.

fitting model to repeated training sets



Magic: you don't need to do this. Just fit the model to the whole data set and compute AIC and you are done.

AIC and cross-validation

Checking for near Normality is good enough.

Minimizing the AIC assuming $\boxed{\text{Gaussian}}$ residuals is $\boxed{\text{asymptotically}}$ equivalent to minimizing one-step time series cross validation MSE.

If you have enough data

** THE ASSUMPTIONS ARE NOT THAT STRONG. **

Automatic forecasting (only because of the models)

From Hyndman et al. (IJF, 2002): ETS( )

- Apply each model that is appropriate to the data. Optimize parameters and initial values using MLE (or some other criterion).
- Select best model using AICc:
- Produce forecasts using best model.
- Obtain forecast intervals using underlying state space model.

Method performed very well in M3 competition.

→ used widely in industry to routinely generate forecasts

Outline

Forecasting with ETS models

Traditional point forecasts: iterate the equations for $t = T + 1, T + 2, \dots, T + h$.

Forecasting with ETS models

Traditional point forecasts: iterate the equations for $t = T+1, T+2, \dots, T+h$.

- Not the same as $\text{E}(y_{t+h}|\boldsymbol{x}_t)$ unless seasonality is additive.
- `fable` uses $\text{E}(y_{t+h}|\boldsymbol{x}_t)$.
- Point forecasts for ETS(A,$*$,$*$) are identical to ETS(M,$*$,$*$) if the parameters are the same. (see example that follows)

Example: ETS(A,A,N)

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

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

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

$$
\begin{aligned}
y_{T+1} &= \ell_T + b_T + \varepsilon_{T+1} \\
\hat{y}_{T+1|T} &= \ell_T + b_T \\
y_{T+2} &= \ell_{T+1} + b_{T+1} + \varepsilon_{T+2} \\
&= (\ell_T + b_T + \alpha\varepsilon_{T+1}) + (b_T + \beta\varepsilon_{T+1}) + \varepsilon_{T+2} \\
\hat{y}_{T+2|T} &= \ell_T + 2b_T
\end{aligned}
$$

etc.

Example: ETS(A,A,N)

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

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

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

$$\begin{aligned}
y_{T+1} &= \ell_T + b_T + \varepsilon_{T+1} \\
\hat{y}_{T+1|T} &= \ell_T + b_T \\
y_{T+2} &= \ell_{T+1} + b_{T+1} + \varepsilon_{T+2} \\
&= (\ell_T + b_T + \alpha\varepsilon_{T+1}) + (b_T + \beta\varepsilon_{T+1}) + \varepsilon_{T+2} \\
\hat{y}_{T+2|T} &= \ell_T + 2b_T
\end{aligned}$$

$\Rightarrow E(y_{T+1|T}) = \ell_T + b_T + \cancel{E(\varepsilon_{T+1|T})}$ 0

etc.

Example: ETS(A,A,N)

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

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

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

$$y_{T+1} = \ell_T + b_T + \varepsilon_{T+1} \quad \Rightarrow \quad E(y_{T+1} \mid T) = \ell_T + b_T + \cancel{E(\varepsilon_{T+1} \mid T)}^{\,0}$$

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

$$y_{T+2} = \ell_{T+1} + b_{T+1} + \varepsilon_{T+2}$$

$$= (\ell_T + b_T + \alpha\varepsilon_{T+1}) + (b_T + \beta\varepsilon_{T+1}) + \varepsilon_{T+2}$$

$$\hat{y}_{T+2|T} = \ell_T + 2b_T$$

(Compare to Holt's linear trend method)

etc.

* You will need to be able to do this for exam
* See past exam papers

Example: ETS(M,A,N)

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

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

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

$$
\begin{aligned}
y_{T+1} &= (\ell_T + b_T)(1 + \varepsilon_{T+1}) \\
\hat{y}_{T+1|T} &= \ell_T + b_T. \\
y_{T+2} &= (\ell_{T+1} + b_{T+1})(1 + \varepsilon_{T+2}) \\
&= \{(\ell_T + b_T)(1 + \alpha\varepsilon_{T+1}) + [b_T + \beta(\ell_T + b_T)\varepsilon_{T+1}]\}\,(1 + \varepsilon_{T+2}) \\
\hat{y}_{T+2|T} &= \ell_T + 2b_T
\end{aligned}
$$

etc.

* Identical point forecasts
* This was known before for ETS models
* The new bit comes next

Forecasting with ETS models

Prediction intervals: can only be generated using the models.

- The prediction intervals will differ between models with additive and multiplicative errors.
- Exact formulae for some models. (see next page)
- More general to simulate future sample paths, conditional on the last estimate of the states, and to obtain prediction intervals from the percentiles of these simulated future paths.

Prediction intervals

No you don't need to know these.

PI for most ETS models: $\hat{y}_{T+h|T} \pm c\sigma_h$, where c depends on coverage probability and σ_h is forecast standard deviation.

(A,N,N) $\quad \sigma_h = \sigma^2\Big[1 + \alpha^2(h-1)\Big]$

(A,A,N) $\quad \sigma_h = \sigma^2\Big[1 + (h-1)\{\alpha^2 + \alpha\beta h + \frac{1}{6}\beta^2 h(2h-1)\}\Big]$

(A,A_d,N) $\quad \sigma_h = \sigma^2\Big[1 + \alpha^2(h-1) + \frac{\beta\phi h}{(1-\phi)^2}\{2\alpha(1-\phi) + \beta\phi\} - \frac{\beta\phi(1-\phi^h)}{(1-\phi)^2(1-\phi^2)}\{2\alpha(1-\phi^2) + \beta\phi(1+2\phi-\phi^h)\}\Big]$

(A,N,A) $\quad \sigma_h = \sigma^2\Big[1 + \alpha^2(h-1) + \gamma k(2\alpha + \gamma)\Big]$

(A,A,A) $\quad \sigma_h = \sigma^2\Big[1 + (h-1)\{\alpha^2 + \alpha\beta h + \frac{1}{6}\beta^2 h(2h-1)\} + \gamma k\{2\alpha + \gamma + \beta m(k+1)\}\Big]$

(A,A_d,A) $\quad \sigma_h = \sigma^2\Big[1 + \alpha^2(h-1) + \frac{\beta\phi h}{(1-\phi)^2}\{2\alpha(1-\phi) + \beta\phi\} - \frac{\beta\phi(1-\phi^h)}{(1-\phi)^2(1-\phi^2)}\{2\alpha(1-\phi^2) + \beta\phi(1+2\phi-\phi^h)\}$
$\quad + \gamma k(2\alpha + \gamma) + \frac{2\beta\gamma\phi}{(1-\phi)(1-\phi^m)}\{k(1-\phi^m) - \phi^m(1-\phi^{mk})\}\Big]$

Eg ETS (A, N, N) (which method underlies this?)

$$y_t = l_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma^2)$$

$$l_t = l_{t-1} + \alpha \varepsilon_t$$

Lets role forward T+h steps

$$y_{T+h} = l_{T+h-1} + \varepsilon_{T+h}$$

iterate backwards

$$= l_{T+h-2} + \alpha \varepsilon_{T+h-1} + \varepsilon_{T+h}$$

$$\vdots$$

$$= l_T + \alpha [\varepsilon_{T+1} + \varepsilon_{T+2} + \dots + \varepsilon_{T+h-1}] + \varepsilon_{T+h}$$

$$var(y_{T+h|T}) = \alpha^2 [\overbrace{\sigma^2 + \sigma^2 + \dots + \sigma^2}^{h-1}] + \sigma^2 = \alpha^2 (h-1) \sigma^2 + \sigma^2$$

$$= \sigma^2 [1 + \alpha^2 (h-1)]$$