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

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

Error: Additive (A), Multiplicative (M)

ETS models

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

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

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

Error: Additive (A), Multiplicative (M)

Trend: None (N), Additive (A), Additive damped (Ad).

ETS models

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

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

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

Error: Additive (A), Multiplicative (M)

Trend: None (N), Additive (A), Additive damped (Ad).

Seasonality: None (N), Additive (A), Multiplicative (M)

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

Forecast equation	$\hat{y}_{t+h\|t} = \ell_t$
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.
- Obervation equation: relationship between observations and states.
- State equation(s): evolution of the state(s) through time.

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

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

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

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

$$= \alpha y_T + \alpha(1-\alpha) y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + (1-\alpha)^3 \hat{y}_{T-2|T-3} + \dots + (1-\alpha)^T \hat{y}_{1|0}$$

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

we don't have infinite data.

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

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

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

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

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

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

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

$$l_t = l_{t-1} + \overset{1}{\cancel{\alpha}}\, \varepsilon_t \qquad \text{let } \alpha = 1$$

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

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

ETS(A,A,N)

Holt's methods method with additive errors.

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

- Innovation residuals: $\hat{\varepsilon}_t = y_t - \hat{y}_{t|t-1}$

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

Three interesting cases

(1) $\beta = 0$ $\quad b_t = b_{t-1} = \dots = b$ 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.

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

(2) $\alpha = 0 \Rightarrow \beta = 0 \Rightarrow$ hence $b_t = b_{t-1} = \ldots = b$ 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 $\alpha = 1$

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

Three interesting cases

(1) $\beta = 0$ $\quad b_t = b_{t-1} = \dots = b$ 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 $\alpha = 1$

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

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 $\alpha = 1$

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

RANDOM WALK WITH DRIFT MODEL

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

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

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

ETS(M,A,M)

Holt-Winters multiplicative method with multiplicative errors.

linear trend × seasonal = multi seas.

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	$\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$
	$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$
	$s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

- 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.

linear trend × seasonal = multi seas.

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	$\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$
	$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$
	$s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

recall we make assumptions about these

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

ETS(M,A,M)

Holt-Winters multiplicative method with multiplicative errors.

linear trend × seasonal = multi seas.

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	$\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$
	$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$
	$s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

recall we make assumptions about these

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

* Hence, no need to transform

ETS model specification

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

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 if you want to.

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

(Eve Gardner once told me this is the most useful method in his consulting practice)

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

* Introduced in 1950-60
* No optim - people used parameters arbitrarily
* Looked at different combinations (James Taylor, Oxford, damped multi trend in 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).

very unstable

ETS 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

* Monash Contribution
  Rob, Ralph & co-authors in 2000s.
* Several PhD students worked on various aspects of these.

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

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

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	$y_t = (\ell_{t-1} + b_{t-1})(1+\varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1+\alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1+\varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1+\varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1+\alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1+\gamma\varepsilon_t)$
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

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

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

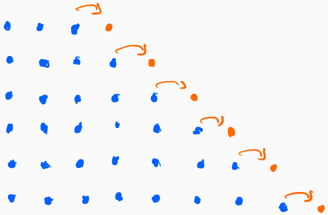
$$\text{BIC} = \text{AIC} + k[\log(T) - 2] = -2\log(\text{L}) + \underbrace{\ln(T)k}_{\text{greater penalty for } T>8}$$

AIC and cross-validation

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

* MAGIC RESULT *

fitting model to repeated training sets



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

AIC and cross-validation

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

checking for near normality is good enough

** THE ASSUMPTIONS ~~ARE~~ ARE NOT THAT STRONG **

Automatic forecasting

From Hyndman et al. (IJF, 2002):

ETS( ): only possible because of the models.

- 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 $t = T+1, T+2, \ldots, T+h$.

Forecasting with ETS models

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

- Not the same as $\mathrm{E}(y_{t+h}|\boldsymbol{x}_t)$ unless seasonality is additive. (recall we do not consider multi trend)
- `fable` uses $\mathrm{E}(y_{t+h}|\boldsymbol{x}_t)$.
- Point forecasts for ETS(A,*,*) are identical to ETS(M,*,*) if the parameters are the same. (see the 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)

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

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

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

$$y_{T+1} = \ell_T + b_T + \varepsilon_{T+1}$$

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

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

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

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

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

etc.

Example: ETS(A,A,N)

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

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

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

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

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

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

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

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

etc.

Example: ETS(A,A,N)

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

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

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

$$y_{T+1} = \ell_T + b_T + \varepsilon_{T+1} \Rightarrow E(y_{T+1} \mid T) = \ell_T + b_T + \underbrace{E(\varepsilon_{T+1} \mid T)}_{0}$$

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

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

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

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

(Compare to Holt's linear trend method)

etc.

Example: ETS(A,A,N)

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

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

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

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

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

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

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

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

(Compare to Holt's linear trend method)

etc.

* You will need to be able to do this for the exam
* Check past exams

Example: ETS(M,A,N)

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

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

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

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

etc.

Example: ETS(M,A,N)

$E(y_{T+1|T}) = (\ell_T + b_T)[1 + E(\varepsilon_{T+1|T})]$ $\xrightarrow{} 0$

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

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

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

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

etc.

* Identical point forecasts with ETS(A,A,N)
* This was known before ETS models
* The 'innovation' 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

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

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

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

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

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

Prediction intervals

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.

* NO you don't need to know these

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

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

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

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

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

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

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

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

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

Let's role forward T+h steps

$$y_{T+h} = \underbrace{l_{T+h-1}}_{\text{iterate backwards}} + \varepsilon_{T+h}$$

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

$$\vdots$$

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

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

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