MONASH University

MONASH BUSINESS SCHOOL

ETF3231/5231
Business forecasting

Week 5: Exponential smoothing

https://bf.numbat.space/

AACSB ACCREDITED EQUIS ACCREDITED ASSOCIATION OF MBAS AMBA ACCREDITED

Outline

Outline

Historical perspective

Developed in the US navy for forecasting spare parts

- Proposed in the late 1950s (Brown 1959, Holt 1957 and Winters 1960 are key pioneering works) as methods (algorithms) to produce point forecasts.

  for this reason not popular with statisticians.

KEY IDEA

- Forecasts are weighted averages of past observations, with the weights decaying smoothly exponentially as the observations get older.
- Framework generates reliable forecasts quickly and for a wide spectrum of time series. A great advantage and of major importance to applications in industry.

  - now used everywhere in business
  - strong benchmarks

Combine components

- Combine components: level ℓ_t , trend (slope) b_t and seasonal s_t to describe a time series

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

- 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).
- Add error ε_t to get equivalent ETS state space models developed in the 1990s and 2000s.

Combine components

- Combine components: level ℓ_t , trend (slope) b_t and seasonal s_t to describe a time series

$$y_t = f(\ell_{t-1}, b_{t-1}, s_{t-m}) \rightarrow \hat{y}_{T+h|T} = f(\ell_T, b_T, s_{T-m+1})$$

- The rate of change of the components are controlled by "smoothing parameters": α, β and γ respectively. → next slide
- Need to choose best values for the smoothing parameters (and initial states).
- Add error ε_t to get equivalent ETS state space models developed in the 1990s and 2000s.

  * Monash very famous about these
  * Pioneer Ralph Snyder (textbook with Rob Hyndman, Anne Koehler & Keith Ord, 2008).

Big idea: control the rate of change (smoothing)

α controls the flexibility of the level ℓ_t (height, overall position of the series)

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

β controls the flexibility of the trend b_t (slope)

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

usually

$0 \leq \alpha, \beta, \gamma \leq 1$

(more to follow)

γ controls the flexibility of the seasonality s_t

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

A model for levels, trends, and seasonalities

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

How do we combine these elements?

A model for levels, trends, and seasonalities

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

How do we combine these elements?

Additively?

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

$\varepsilon_t \sim \text{iid } N(0, \sigma^2)$

Multiplicatively?

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

Perhaps a mix of both?

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

A model for levels, trends, and seasonalities

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

How do we combine these elements?

Additively?

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

Multiplicatively?

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

Perhaps a mix of both?

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

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

× so more than one equation

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)

* Hence many combinations of these
* There are two more we no longer consider Trend: M, Md
* In theory 30 models (2 x 5 x 3) → In practice we consider half of these

Models and methods

Methods

- Algorithms that return point forecasts.

Models and methods

Methods

- Algorithms that return point forecasts.

* You should now be familiar with the difference between methods & models

Models

- Generate same point forecasts but can also generate forecast distributions.
- A stochastic (or random) data generating process that can generate an entire forecast distribution.

  → can generate something that looks like data.
- Allow for "proper" model selection.

  - Ord, Koehler, Snyder (1997, JASA)

* machine learning methods (NNs) are in their algorithmic phase.

WA 1

Outline

Simple Exponential Smoothing - SES

Iterative form

$$\hat{y}_{t+1|t} = \alpha y_t + (1 - \alpha)\hat{y}_{t|t-1}$$

Simple Exponential Smoothing - SES

Iterative form

$$\hat{y}_{t+1|t} = \alpha y_t + (1 - \alpha)\hat{y}_{t|t-1}$$

$$\hat{y}_{2|1} = \alpha y_1 + (1-\alpha)\hat{y}_{1|0}$$

Simple Exponential Smoothing - SES

Iterative form

$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$$

$$\hat{y}_{2|1} = \alpha y_1 + (1-\alpha)\cancel{\hat{y}_{1|0}}\ \ell_0$$

Simple Exponential Smoothing - SES

Iterative form

$$\hat{y}_{t+1|t} = \alpha y_t + (1 - \alpha)\hat{y}_{t|t-1}$$

$$\hat{y}_{2|1} = \alpha y_1 + (1-\alpha)\cancel{\hat{y}_{1|0}}\ \ell_0$$

$$\hat{y}_{3|2} = \alpha y_2 + (1-\alpha)\hat{y}_{2|1}$$

Simple Exponential Smoothing - SES

Iterative form

$$\hat{y}_{t+1|t} = \alpha y_t + (1 - \alpha)\hat{y}_{t|t-1}$$

$$\hat{y}_{2|1} = \alpha y_1 + (1-\alpha)\cancel{\hat{y}_{1|0}}\ \ell_0$$

$$\hat{y}_{3|2} = \alpha y_2 + (1-\alpha)\hat{y}_{2|1}$$

$$\hat{y}_{4|3} = \alpha y_3 + (1-\alpha)\hat{y}_{3|2}$$

Simple Exponential Smoothing - SES

Iterative form

$$\hat{y}_{t+1|t} = \alpha y_t + (1 - \alpha)\hat{y}_{t|t-1}$$

$$\hat{y}_{2|1} = \alpha y_1 + (1-\alpha)\cancel{\hat{y}_{1|0}} \; \ell_0$$

$$\hat{y}_{3|2} = \alpha y_2 + (1-\alpha)\hat{y}_{2|1}$$

$$\hat{y}_{4|3} = \alpha y_3 + (1-\alpha)\hat{y}_{3|2}$$

$$\vdots$$

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

Simple Exponential Smoothing - SES

Iterative form

$$\hat{y}_{t+1|t} = \alpha y_t + (1 - \alpha)\hat{y}_{t|t-1}$$

Weighted average form

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

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

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

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

start from

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

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

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

Start from

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

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

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

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

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

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

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

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

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

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

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

we don't have infinite data.

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

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

Simple Exponential Smoothing - SES

Iterative form

$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$$

Weighted average form

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

Simple Exponential Smoothing - SES

Iterative form

$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$$

Weighted average form

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

Component form

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

Simple Exponential Smoothing - SES

Iterative form

$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$$

Let $\hat{y}_{t+1|t} = \ell_t \Rightarrow \hat{y}_{t|t-1} = \ell_{t-1}$

Weighted average form

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

Component form

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

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

Component form

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

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

Component form

$$\begin{aligned} \text{Forecast equation} \quad & \hat{y}_{t+1|t} = \ell_t \\ \text{Smoothing equation} \quad & \ell_t = \alpha y_t + (1-\alpha)\ell_{t-1} \end{aligned}$$

for any model

Residual: $\boxed{e_t = y_t - \hat{y}_{t|t-1}} = y_t - \ell_{t-1}$.

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

Component form

$$\begin{aligned} \text{Forecast equation} \qquad \hat{y}_{t+1|t} &= \ell_t \\ \text{Smoothing equation} \qquad \ell_t &= \alpha y_t + (1-\alpha)\ell_{t-1} \end{aligned}$$

Residual: $e_t = y_t - \hat{y}_{t|t-1} = y_t - \ell_{t-1}$. → re-arrange

Error correction form

$$\begin{aligned} y_t &= \ell_{t-1} + e_t \\ \ell_t &= \ell_{t-1} + \alpha(y_t - \ell_{t-1}) \\ &= \ell_{t-1} + \alpha e_t \end{aligned}$$

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

Component form

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

$\Rightarrow \ell_t = \alpha y_t + \ell_{t-1} + \alpha \ell_{t-1}$

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

Residual: $e_t = y_t - \hat{y}_{t|t-1} = y_t - \ell_{t-1}$.

Error correction form

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

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

$$= \ell_{t-1} + \alpha e_t$$

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

Component form

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

Residual: $e_t = y_t - \hat{y}_{t|t-1} = y_t - \ell_{t-1}$.

Error correction form

$$\begin{aligned} y_t &= \ell_{t-1} + e_t \\ \ell_t &= \ell_{t-1} + \alpha(y_t - \ell_{t-1}) \\ &= \ell_{t-1} + \alpha e_t \end{aligned}$$

* KEY RESULT *

Specify probability distribution for e_t, we assume $e_t = \varepsilon_t \sim \mathrm{NID}(0, \sigma^2)$.

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

an ETS model

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

- innovations or single source of error because equations have the same error process, ε_t.
- Observation equation: relationship between observations and states. or measurement equation
- State equation(s): evolution of the state(s) through time.

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

an ETS model

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

- innovations or single source of error because equations have the same error process, ε_t.
- Observation equation (or measurement equation): relationship between observations and states.
- State equation(s): evolution of the state(s) through time.

QUESTION / HOMEWORK: what happens when $\alpha = 1$?

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

- Specify relative errors $\varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}} \sim \text{NID}(0, \sigma^2)$
- Substituting $\hat{y}_{t|t-1} = \ell_{t-1}$ gives:
  - $y_t = \ell_{t-1} + \ell_{t-1}\varepsilon_t$
  - $e_t = y_t - \hat{y}_{t|t-1} = \ell_{t-1}\varepsilon_t$

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

- Specify relative errors $\varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}} \sim \text{NID}(0, \sigma^2)$
- Substituting $\hat{y}_{t|t-1} = \ell_{t-1}$ gives:
  - $y_t = \ell_{t-1} + \ell_{t-1}\varepsilon_t = \ell_{t-1}(1+\varepsilon_t)$
  - $e_t = y_t - \hat{y}_{t|t-1} = \ell_{t-1}\varepsilon_t$

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

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

- Specify relative errors $\varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}} \sim \text{NID}(0, \sigma^2)$
- Substituting $\hat{y}_{t|t-1} = \ell_{t-1}$ gives:
  - $y_t = \ell_{t-1} + \ell_{t-1}\varepsilon_t$ $= \ell_{t-1}(1+\varepsilon_t)$
  - $e_t = y_t - \hat{y}_{t|t-1} = \ell_{t-1}\varepsilon_t$ Recall $\ell_t = \ell_{t-1} + \alpha e_t \Rightarrow \ell_t = \ell_{t-1} + \alpha \ell_{t-1} + \varepsilon_t$
    $= \ell_{t-1}(1+\alpha\varepsilon_t)$

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

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

- Specify relative errors $\varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}} \sim \text{NID}(0, \sigma^2)$
- Substituting $\hat{y}_{t|t-1} = \ell_{t-1}$ gives:
  - $y_t = \ell_{t-1} + \ell_{t-1}\varepsilon_t$
  - $e_t = y_t - \hat{y}_{t|t-1} = \ell_{t-1}\varepsilon_t$

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

they are different models

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

(ℓ_0, α)

Residuals

Residuals (response)

$$e_t = y_t - \hat{y}_{t|t-1}$$

* for <u>all</u> methods & models

* .resid = e_t

Residuals

Residuals (response)

$$e_t = y_t - \hat{y}_{t|t-1}$$

* for <u>all</u> methods & models

* .resid = e_t

Innovation residuals

Additive error model:

$$\hat{\varepsilon}_t = y_t - \hat{y}_{t|t-1} = e_t$$

* these are attached to the model & the fit

* we make assumptions about these

Multiplicative error model:

$$\hat{\varepsilon}_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}} \neq e_t$$

* .innov = $\hat{\varepsilon}_t$

WA 2

Outline

SES Component form

Forecast equation : $\hat{y}_{t+h|t} = \ell_t$

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

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

Holt's trend method

~~SES~~ Component form

Forecast equation : $\hat{y}_{t+h|t} = \ell_t + \underset{\text{h-step}}{h} \; \underset{\text{slope}}{b_t}$ || linear / straight

Smoothing equation

level : $\ell_t = \alpha y_t + (1-\alpha)\,\hat{y}_{t|t-1}$

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

trend : $b_t = \beta^* \underbrace{(\ell_t - \ell_{t-1})}_{\text{current slope}} + (1-\beta^*) \underbrace{b_{t-1}}_{\text{last estimated slope}}$

(change in estimated level)

Holt's trend method

~~SES~~ Component form

Forecast equation : $\hat{y}_{t+h|t} = \ell_t + \underset{\text{h-step}}{h}\, \underset{\text{slope}}{b_t}$ || linear / straight

Smoothing equation level : $\ell_t = \alpha y_t + (1-\alpha)\, \hat{y}_{t|t-1}$

$= \alpha y_t + (1-\alpha)\, \ell_{t-1}$

Holt's linear trend method

Component form

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

Holt's linear trend method

Component form

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

- 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.

* we now have 2 parameters and 2 initial states.

Error Correction form

$$e_t = y_t - \hat{y}_{t|t-1} \implies y_t = \hat{y}_{t|t-1} + e_t$$

sub in $\hat{y}_{t|t-1} = l_{t-1} + b_{t-1}$

$$\implies y_t = l_{t-1} + b_{t-1} + \cancel{e_t}$$

$\varepsilon_t \sim NID(0, \sigma^2)$

(Obs eqn)

Error Correction form

$$e_t = y_t - \hat{y}_{t|t-1} \Rightarrow y_t = \hat{y}_{t|t-1} + e_t$$

sub in $\hat{y}_{t|t-1} = l_{t-1} + b_{t-1}$

$$\Rightarrow y_t = l_{t-1} + b_{t-1} + \cancel{e_t} \rightarrow \varepsilon_t \sim NID(0, \sigma^2)$$

(Obs eqn)

$$l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) = \alpha y_t + l_{t-1} + b_{t-1} - \alpha l_{t-1} - \alpha b_{t-1}$$

$$= l_{t-1} + b_{t-1} + \alpha(y_t - (l_{t-1} + b_{t-1}))$$

$$\Rightarrow l_t = l_{t-1} + b_{t-1} + \alpha \cancel{e_t} \rightarrow \varepsilon_t \sim NID(0, \sigma^2)$$

(Level eqn)

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

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

sub in ℓ_t

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

$$= \cancel{\beta^* \ell_{t-1}} + \cancel{\beta^* b_{t-1}} + \alpha \beta^* e_t - \cancel{\beta^* \ell_{t-1}} + b_{t-1} - \cancel{\beta^* b_{t-1}}$$

$$\Rightarrow \quad b_t = b_{t-1} + \alpha \beta^* \cancel{e_t}$$

$\varepsilon_t \sim NID(0, \sigma^2)$ (Trend eqn)

ETS(A,A,N)

Holt's linear method with additive errors.

- Assume $\varepsilon_t = y_t - \ell_{t-1} - b_{t-1} \sim \text{NID}(0, \sigma^2)$.
- Substituting into the error correction equations for Holt's linear method

$$
\begin{aligned}
y_t &= \ell_{t-1} + b_{t-1} + \varepsilon_t \\
\ell_t &= \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t \\
b_t &= b_{t-1} + \alpha\beta^*\varepsilon_t
\end{aligned}
$$

* Developed in early 2000 at Monash

- For simplicity, set $\beta = \alpha\beta^*$.

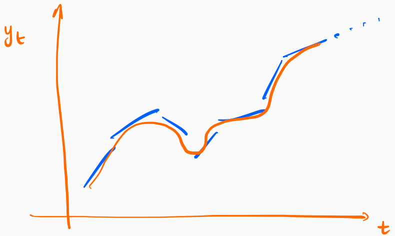
$0 < \beta^* < 1 \Rightarrow 0 < \beta < \alpha$

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



ETS(M,A,N)

Holt's linear method with multiplicative errors.

- Assume $\varepsilon_t = \frac{y_t-(\ell_{t-1}+b_{t-1})}{(\ell_{t-1}+b_{t-1})}$ $\Rightarrow \hat{\varepsilon}_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}} \neq e_t$
- Following a similar approach as above, the innovations state space model underlying Holt's linear method with multiplicative errors is specified as

$$
\begin{aligned}
y_t &= (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t) \\
\ell_t &= (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t) \\
b_t &= b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t
\end{aligned}
$$

* So now we multiply by the error component

where again $\beta = \alpha\beta^*$ and $\varepsilon_t \sim \text{NID}(0, \sigma^2)$.

Damped trend method

(Eve Gardner & Ed Mckenzie, 1985)

Component form

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

→ to produce more conservative forecasts

Assume $\phi = 0.9$

$h=1$ $\quad \phi b_T \quad 0.9\, b_T$

$h=2$ $\quad (\phi+\phi^2) b_T \quad (\text{''} + 0.81)\, b_T$

$h=3$ $\quad (\phi+\phi^2+\phi^3) b_T \quad (\text{''} + \text{''} + 0.729)\, b_T$

as $h \to \infty$ $\quad \phi^\infty \to 0 \quad \phi+\phi^2+\phi^3+\ldots = \frac{\phi}{1-\phi} \longrightarrow \ell_T + \frac{\phi}{1-\phi} b_T$ (flat/Constant)

Damped trend method

Component form

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

- 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.

Over to you

- Write down the model for ETS(A,Ad,N)

Recall you need error correction form. Start with

$$e_t = \hat{y}_t - \hat{y}_{t|t-1} = y_t - (\ell_{t-1} + \phi b_{t-1})$$

Re-arrange $\quad y_t = \ell_{t-1} + \phi b_{t-1} + e_t \quad \varepsilon_t \sim NID(0, \sigma^2)$

WA 3