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

- Proposed in the late 1950s (Brown 1959, Holt 1957 and Winters 1960 are key pioneering works) as methods (algorithms) to produce point forecasts.
- Forecasts are weighted averages of past observations, with the weights decaying 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.

# Combine components

- Combine components: level $\ell_t$ , 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": $\alpha$, $\beta$ and $\gamma$ respectively.
- Need to choose best values for the smoothing parameters (and initial states).
- Add error $\varepsilon_t$ to get equivalent ETS state space models developed in the 1990s and 2000s.

# Big idea: control the rate of change (smoothing)

$\alpha$ controls the flexibility of the level $\ell_t$

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

$\beta$ controls the flexibility of the trend $b_t$

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

$\gamma$ 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 ($\ell_t$), 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 ($\ell_t$), 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?**

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

# A model for levels, trends, and seasonalities

We want a model that captures the level ($\ell_t$), 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?

# ETS models

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

$\nearrow \uparrow \nwarrow$

**E**rror **T**rend **S**eason

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

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

# ETS models

**General notation** $\underset{\text{Error}}{E}\ \underset{\text{Trend}}{T}\ \underset{\text{Season}}{S}$ : **E**xponen**T**ial **S**moothing

```
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"`)

# Models and methods

## Methods

- Algorithms that return point forecasts.

# Models and methods

## Methods

- Algorithms that return point forecasts.

## 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.
- Allow for “proper” model selection.

# Outline

1. Exponential smoothing
2. Simple exponential smoothing
3. Models with trend

# 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

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

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

# 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

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

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

# 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

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

# 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

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

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

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

| | |
|---|---|
| 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, $\varepsilon_t$.
- Measurement equation: relationship between observations and states.
- State equation(s): evolution of the state(s) through time.

# 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:
  - $y_t = \ell_{t-1} + \ell_{t-1}\varepsilon_t$
  - $e_t = y_t - \hat{y}_{t|t-1} = \ell_{t-1}\varepsilon_t$

| | |
|---|---|
| Measurement 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:
  - $y_t = \ell_{t-1} + \ell_{t-1}\varepsilon_t$
  - $e_t = y_t - \hat{y}_{t|t-1} = \ell_{t-1}\varepsilon_t$

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

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

# Residuals

## Residuals (response)

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

# Residuals

## Residuals (response)

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

## Innovation residuals

Additive error model:

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

Multiplicative error model:

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

# Outline

# Holt's linear trend method

## Component form

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

# 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 $\alpha$ and $\beta^*$ ($0 \le \alpha, \beta^* \le 1$).
- $\ell_t$ 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.

# 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

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

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

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

# ETS(M,A,N)

Holt's linear method with multiplicative errors.

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

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

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

# Damped trend method

## Component form

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

# Damped trend method

## Component form

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

- 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)