# STAT481/581: Introduction to Time Series Analysis

Exponential smoothing

OTexts.org/fpp3/

# Outline

# Outline

# 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": $\alpha$, $\beta$ and $\gamma$ respectively.
- Need to choose best values for the smoothing parameters (and initial states).
- 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**

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

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

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

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

$\gamma$ controls the flexibility of the **seasonality**

- 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?

**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?**

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

# ETS models

General notation E T S : **E**xponen**T**ial **S**moothing
↗ ↑ ↖
**E**rror **T**rend **S**eason

**E**rror: Additive (`"A"`) or multiplicative (`"M"`)

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

**S**easonality: None (`"N"`), additive (`"A"`) or multiplicative (`"M"`)

# Outline

# Simple methods

Time series $y_1, y_2, \dots, y_T$.

## Random walk forecasts

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

## 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.

# Simple Exponential Smoothing

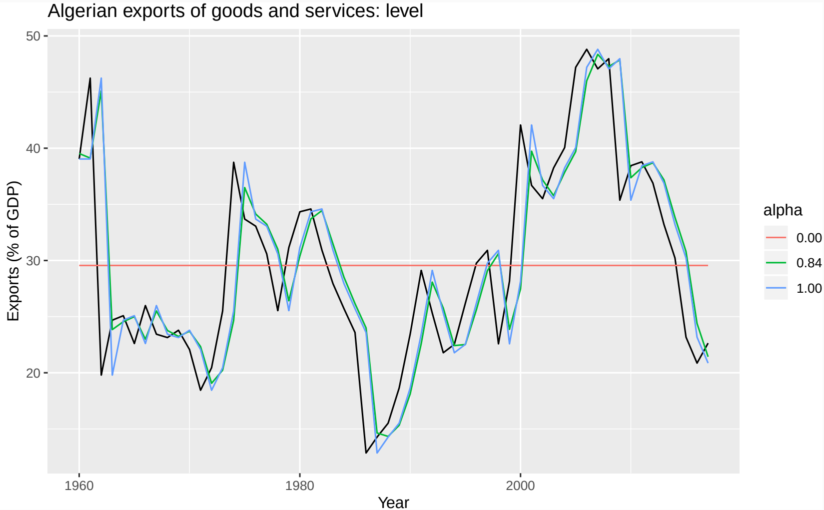
## Forecast equation

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

where $0 \leq \alpha \leq 1$.

| | Weights assigned to observations for: | | | |
|---|---|---|---|---|
| Observation | $\alpha = 0.2$ | $\alpha = 0.4$ | $\alpha = 0.6$ | $\alpha = 0.8$ |
| $y_T$ | 0.2 | 0.4 | 0.6 | 0.8 |
| $y_{T-1}$ | 0.16 | 0.24 | 0.24 | 0.16 |
| $y_{T-2}$ | 0.128 | 0.144 | 0.096 | 0.032 |
| $y_{T-3}$ | 0.1024 | 0.0864 | 0.0384 | 0.0064 |
| $y_{T-4}$ | $(0.2)(0.8)^4$ | $(0.4)(0.6)^4$ | $(0.6)(0.4)^4$ | $(0.8)(0.2)^4$ |
| $y_{T-5}$ | $(0.2)(0.8)^5$ | $(0.4)(0.6)^5$ | $(0.6)(0.4)^5$ | $(0.8)(0.2)^5$ |

# Simple Exponential Smoothing



# Simple Exponential Smoothing

## 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.

# ETS(A,N,N)

State space form:

| | |
|---|---|
| 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.
- Transition/state equation(s): evolution of the state(s) through time.

# ETS(A,N,N)

# ETS(A,N,N)

## Component form

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

- $\ell_t$ is the level (or the smoothed value) of the series at time t.
- $\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$
  Iterate to get exponentially weighted moving average form.

## Weighted average form

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

# Optimising smoothing parameters

- Need to choose best values for $\alpha$ and $\ell_0$.
- Similarly to regression, choose optimal parameters by minimising SSE:

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

- Unlike regression there is no closed form solution — use numerical optimization.

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

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

By default, an optimal value for $\alpha$ and $\ell_0$ is used.

$\alpha$ can be chosen manually in `trend()`.

```
trend("N", alpha = 0.5)
trend("N", alpha_range = c(0.2, 0.8))
```

# Algeria economy data

```
algeria_economy <- tsibbledata::global_economy %>%
  filter(Country == "Algeria")
fit <- algeria_economy %>%
  model(ETS(Exports ~ error("A") + trend("N")
              +season("N"),opt_crit = "mse"))
tidy(fit)$term
```

```
## [1] "alpha" "l"
```

```
tidy(fit)$estimate
```

```
## [1]  0.8398 39.5401
```

# Algeria economy data

- For Algerian Exports example:
  - $\hat{\alpha} = 0.8398$
  - $\hat{\ell}_0 = 39.54$
- Here $l_0$

# Example: Algerian Exports

```
report(fit)
```

```
## Series: Exports
## Model: ETS(A,N,N)
##   Smoothing parameters:
##     alpha = 0.8398
##
##   Initial states:
##      l
##  39.54
##
##   sigma^2:  35.63
##
##   AIC  AICc   BIC
## 446.7 447.2 452.9
```

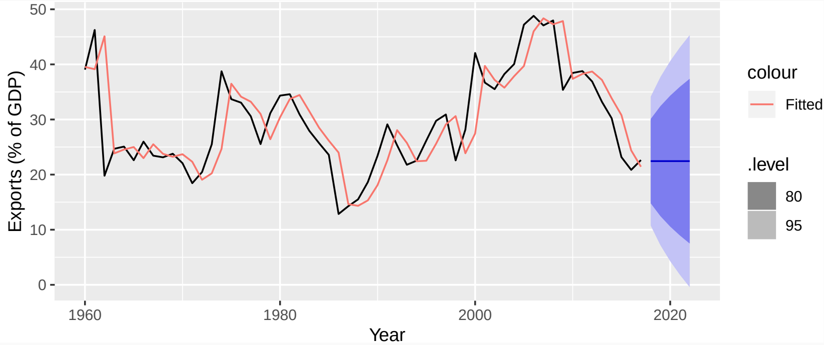
# Example: Algerian Exports

```
components(fit) %>%
  left_join(fitted(fit), by = c("Country", ".model", "Year"))
```

```
## # A tsibble: 59 x 7 [1Y]
## # Key:       Country, .model [1]
##    Country .model  Year Exports level remainder .fitted
##    <fct>   <chr>  <dbl>   <dbl> <dbl>     <dbl>   <dbl>
##  1 Algeria ANN     1959    NA    39.5    NA        NA
##  2 Algeria ANN     1960    39.0  39.1    -0.496    39.5
##  3 Algeria ANN     1961    46.2  45.1     7.12     39.1
##  4 Algeria ANN     1962    19.8  23.8   -25.3      45.1
##  5 Algeria ANN     1963    24.7  24.6     0.841    23.8
##  6 Algeria ANN     1964    25.1  25.0     0.534    24.6
##  7 Algeria ANN     1965    22.6  23.0    -2.39     25.0
##  8 Algeria ANN     1966    26.0  25.5     3.00     23.0
##  9 Algeria ANN     1967    23.4  23.8    -2.07     25.5
## 10 Algeria ANN     1968    23.1  23.2    -0.630    23.8
## # ... with 49 more rows
```

# Simple Exponential Smoothing is the initial state.

```
fc <- fit %>% forecast(h = 5)
fc %>%autoplot(algeria_economy) + geom_line(aes(y = .fitted,
colour = "Fitted"), data = augment(fit)) +
  ylab("Exports (% of GDP)") + xlab("Year")
```



# ETS(M,N,N)

SES with multiplicative errors.

| | |
|---|---|
| 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.

# ETS(M,N,N)

# Outline

# ETS(A,A,N)

Holt's linear method with additive errors.

- $b_t$: slope
- State space form:

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

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

# Holt's linear trend

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

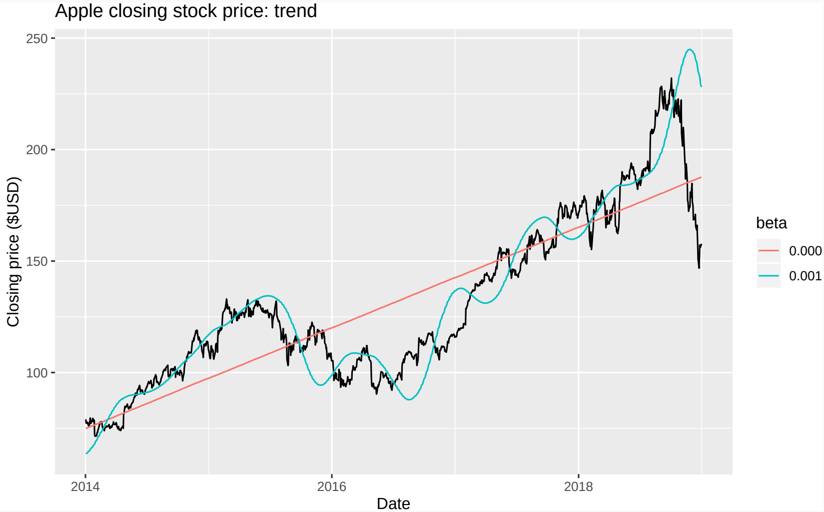
- Two smoothing parameters $\alpha$ and $\beta^*$ $(0 \le \alpha, \beta^* \le 1)$.

# Holt's linear trend

- $\ell_t$ 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.

# Holt's linear trend

# Exponential smoothing: trend/slope



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

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

By default, optimal values for $\beta$ and $b_0$ are used.

$\beta$ can be chosen manually in `trend()`.

```
trend("A", beta = 0.004)
trend("A", beta_range = c(0, 0.1))
```

# AUS pupulation data

```
aus_economy <- global_economy %>%filter(Code == "AUS") %>%
  mutate(Pop = Population / 1e6)
fit <- aus_economy %>%
  model(AAN = ETS(Pop ~ error("A") + trend("A") +
                          season("N")))
fc <- fit %>% forecast(h = 10)
```

# AUS pupulation data

```r
report(fit)
```

```
## Series: Pop
## Model: ETS(A,A,N)
##   Smoothing parameters:
##     alpha = 0.9999
##     beta  = 0.3266
##
##   Initial states:
##        l      b
##   10.05 0.2225
##
##   sigma^2:  0.0041
##
##     AIC    AICc     BIC
## -76.99 -75.83 -66.68
```

# AUS pupulation data

```r
components(fit) %>%
  left_join(fitted(fit), by = c("Country", ".model", "Year"))
```

```
## # A tsibble: 59 x 8 [1Y]
## # Key:       Country, .model [1]
##    Country .model  Year   Pop level slope remainder
##    <fct>   <chr>  <dbl> <dbl> <dbl> <dbl>     <dbl>
##  1 Austra~ AAN     1959  NA    10.1 0.222 NA
##  2 Austra~ AAN     1960  10.3  10.3 0.222 -0.000145
##  3 Austra~ AAN     1961  10.5  10.5 0.217 -0.0159
##  4 Austra~ AAN     1962  10.7  10.7 0.231  0.0418
##  5 Austra~ AAN     1963  11.0  11.0 0.223 -0.0229
##  6 Austra~ AAN     1964  11.2  11.2 0.221 -0.00641
##  7 Austra~ AAN     1965  11.4  11.4 0.221 -0.000314
##  8 Austra~ AAN     1966  11.7  11.7 0.235  0.0418
##  9 Austra~ AAN     1967  11.8  11.8 0.206 -0.0869
## 10 Austra~ AAN     1968  12.0  12.0 0.208  0.00350
## # ... with 49 more rows, and 1 more variable:
```

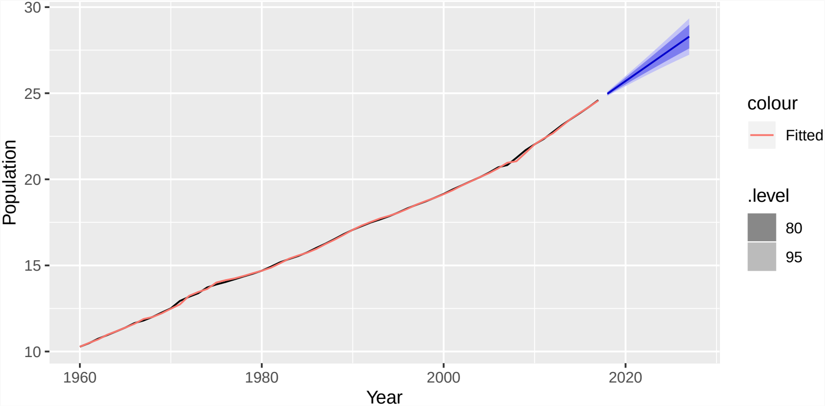
# AUS pupulation data

```
fc
```

```
## # A fable: 10 x 5 [1Y]
## # Key:     Country, .model [1]
##    Country   .model  Year   Pop .distribution
##    <fct>     <chr>  <dbl> <dbl> <dist>
##  1 Australia AAN     2018  25.0 N(25, 0.0041)
##  2 Australia AAN     2019  25.3 N(25, 0.0114)
##  3 Australia AAN     2020  25.7 N(26, 0.0227)
##  4 Australia AAN     2021  26.1 N(26, 0.0389)
##  5 Australia AAN     2022  26.4 N(26, 0.0609)
##  6 Australia AAN     2023  26.8 N(27, 0.0895)
##  7 Australia AAN     2024  27.2 N(27, 0.1257)
##  8 Australia AAN     2025  27.6 N(28, 0.1704)
##  9 Australia AAN     2026  27.9 N(28, 0.2243)
## 10 Australia AAN     2027  28.3 N(28, 0.2885)
```

# AUS pupulation data

```
fit %>%forecast(h = 10) %>%autoplot(aus_economy) +
  geom_line(aes(y=.fitted,colour="Fitted"),data =
augment(fit))+ylab("Population") + xlab("Year")
```



# ETS(M,A,N)

Holt's linear method with multiplicative errors.

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

# ETS(M,A,N) component form

# Damped trend method

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

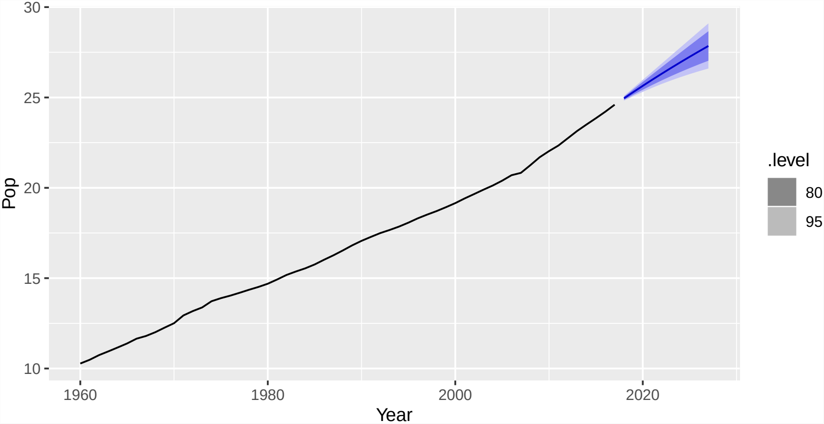
## Your turn

- Write down the state space model for ETS(A,Ad,N)
- Derive the forecast equation and the smoothing equations.

# Your turn

# Example: Australian population

```
aus_economy %>%
  model(holt = ETS(Pop ~ error("A") + trend("Ad") + season("N")))
  forecast(h = 10) %>%
  autoplot(aus_economy)
```



# Example: Australian population

```r
fit <- aus_economy %>%
  filter(Year <= 2010) %>%
  model(
    ses = ETS(Pop ~ error("A") + trend("N") + season("N")),
    holt = ETS(Pop ~ error("A") + trend("A") + season("N")),
    damped = ETS(Pop ~ error("A") + trend("Ad") + season("N"))
  )
```

```r
tidy(fit)
accuracy(fit)
```

# Example: Australian population

| term | SES | Linear trend | Damped trend |
|---|---|---|---|
| $\alpha$ | 1.00 | 1.00 | 1.00 |
| $\beta^*$ | | 0.30 | 0.40 |
| $\phi$ | | | 0.98 |
| $\ell_0$ | 10.28 | 10.05 | 10.04 |
| $b_0$ | | 0.22 | 0.25 |
| Training RMSE | 0.24 | 0.06 | 0.07 |
| Test RMSE | 1.63 | 0.15 | 0.21 |
| Test MASE | 6.18 | 0.55 | 0.75 |
| Test MAPE | 6.09 | 0.55 | 0.74 |
| Test MAE | 1.45 | 0.13 | 0.18 |

# Your turn

`fma::eggs` contains the price of a dozen eggs in the United States from 1900–1993

1. Use simple exponential smoothing (SES) and Holt's method (with and without damping) to forecast "future" data.
   [Hint: use h=100 so you can clearly see the differences between the options when plotting the forecasts.]
2. Which method gives the best training RMSE?
3. Are these RMSE values comparable?
4. Do the residuals from the best fitting method

# Outline

# ETS(A,A,A)

- Holt and Winters extended Holt's method to capture seasonality.
- Holt-Winters additive method in state space form:

$$\begin{aligned}
\text{Observation equation} \quad & y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t \\
\text{State equations} \quad & \ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t \\
& b_t = b_{t-1} + \beta\varepsilon_t \\
& s_t = s_{t-m} + \gamma\varepsilon_t
\end{aligned}$$

- $k$ is integer part of $(h-1)/m$.
- $s_t$: seasonal state.

# Holt-Winters additive method component form

## Component form

$$\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + hb_t + s_{t+h-m(k+1)} \\
\ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}
\end{aligned}$$

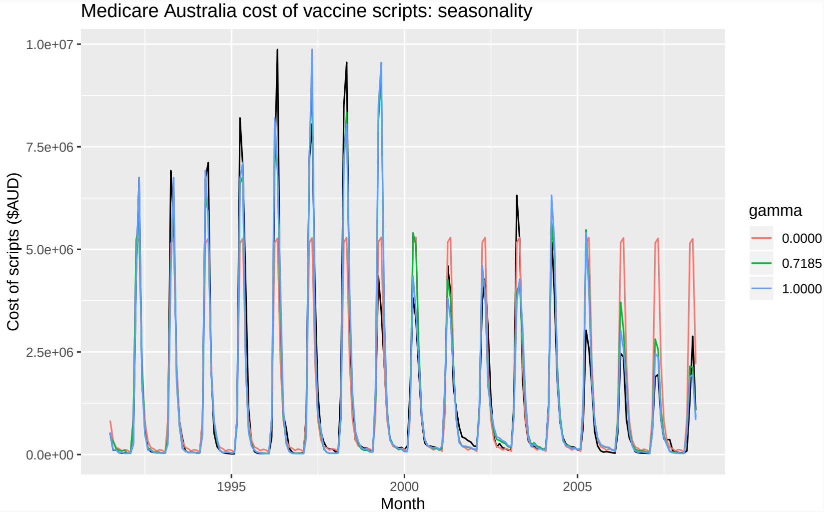
- $k =$ integer part of $(h-1)/m$. Ensures estimates from the final year are used for forecasting.
- Parameters: $0 \le \alpha \le 1$, $0 \le \beta^* \le 1$, $0 \le \gamma \le 1-\alpha$ and $m =$ period of seasonality (e.g. $m = 4$ for quarterly data).

# Holt-Winters additive method component form

# Holt-Winters additive method

- Seasonal component is usually expressed as $s_t = \gamma^*(y_t - \ell_t) + (1 - \gamma^*)s_{t-m}$.
- Substitute in for $\ell_t$: $s_t = \gamma^*(1-\alpha)(y_t - \ell_{t-1} - b_{t-1}) + [1 - \gamma^*(1-\alpha)]s_{t-m}$
- We set $\gamma = \gamma^*(1 - \alpha)$.
- The usual parameter restriction is $0 \le \gamma^* \le 1$, which translates to $0 \le \gamma \le (1 - \alpha)$.

# Exponential smoothing: seasonality



# Your turn

- Write down the state space model and component form for ETS(A,N,A).

# Your turn

# ETS(M,A,M)

- Holt-Winters multiplicative method with multiplicative errors for when seasonal variations are changing proportional to the level of the series.

$$\begin{aligned}
\text{Observation equation} \quad y_t &= (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t) \\
\text{State equations} \quad \ell_t &= (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t) \\
b_t &= b_{t-1}(1 + \beta\varepsilon_t) \\
s_t &= s_{t-m}(1 + \gamma\varepsilon_t)
\end{aligned}$$

- $k$ is integer part of $(h-1)/m$.

# Holt-Winters multiplicative method

## Component form

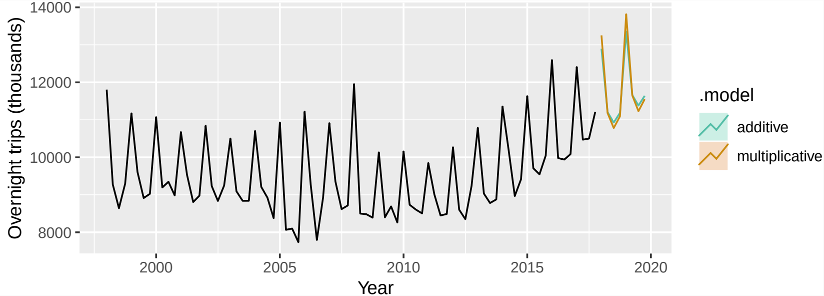
$$\begin{aligned}
\hat{y}_{t+h|t} &= (\ell_t + hb_t)s_{t+h-m(k+1)} \\
\ell_t &= \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
s_t &= \gamma \frac{y_t}{(\ell_{t-1} + b_{t-1})} + (1-\gamma)s_{t-m}
\end{aligned}$$

- $k$ is integer part of $(h-1)/m$.
- With additive method $s_t$ is in absolute terms: within each year $\sum_i s_i \approx 0$.
- With multiplicative method $s_t$ is in relative terms: within each year $\sum_i s_i \approx m$.

# Holt-Winters multiplicative method component form

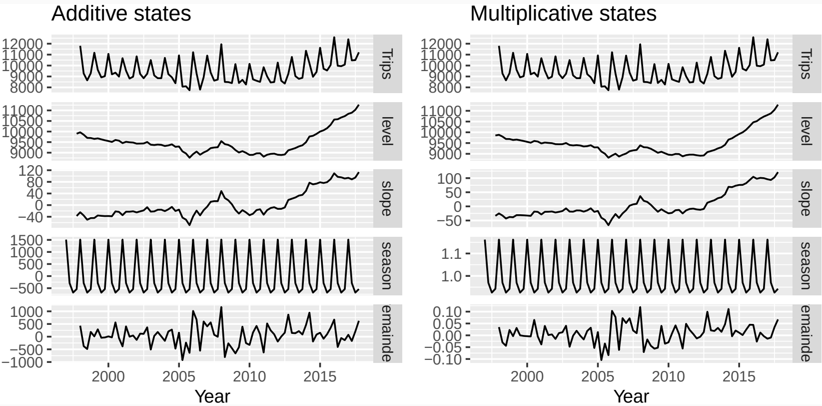
# Example: Australian holiday tourism

```
aus_holidays <- tourism %>% filter(Purpose == "Holiday") %>%
  summarise(Trips = sum(Trips))
fit <- aus_holidays %>%
  model(
    additive = ETS(Trips ~ error("A") + trend("A") + season("A")),
    multiplicative = ETS(Trips ~ error("M") + trend("A") + season("M"))
  )
fc <- fit %>% forecast()
```



# Estimated components

```
components(fit)
```



# Holt-Winters damped method

Often the single most accurate forecasting method for seasonal data:

$$
\begin{aligned}
\hat{y}_{t+h|t} &= [\ell_t + (\phi + \phi^2 + \cdots + \phi^h)b_t]s_{t+h-m(k+1)} \\
\ell_t &= \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1} \\
s_t &= \gamma \frac{y_t}{(\ell_{t-1} + \phi b_{t-1})} + (1-\gamma)s_{t-m}
\end{aligned}
$$

# Your turn

Apply Holt-Winters' multiplicative method to the Gas data from `aus_production`.

1. Why is multiplicative seasonality necessary here?
2. Experiment with making the trend damped.
3. Check that the residuals from the best method look like white noise.

# Outline

# 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

# ETS models

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

# Multiplicative error models

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

# Estimating ETS models

- Smoothing parameters $\alpha$, $\beta$, $\gamma$ and $\phi$, 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.

# Innovations state space models

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

$$\begin{aligned} y_t &= \underbrace{h(\boldsymbol{x}_{t-1})}_{\mu_t} + \underbrace{k(\boldsymbol{x}_{t-1})\varepsilon_t}_{e_t} \\ \boldsymbol{x}_t &= f(\boldsymbol{x}_{t-1}) + g(\boldsymbol{x}_{t-1})\varepsilon_t \end{aligned}$$

## Additive errors

$$k(x) = 1. \qquad y_t = \mu_t + \varepsilon_t.$$

## Multiplicative errors

$$k(\boldsymbol{x}_{t-1}) = \mu_t. \qquad y_t = \mu_t(1 + \varepsilon_t).$$

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

# Innovations state space models

## Estimation

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

- Estimate parameters $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \phi)$ and initial states $\boldsymbol{x}_0 = (\ell_0, b_0, s_0, s_{-1}, \dots, s_{-m+1})$ by minimizing $L^*$.

# Parameter restrictions

## *Usual* region

- Traditional restrictions in the methods
  $0 < \alpha, \beta^*, \gamma^*, \phi < 1$
  (equations interpreted as weighted averages).
- In models we set $\beta = \alpha\beta^*$ and $\gamma = (1 - \alpha)\gamma^*$.
- Therefore $0 < \alpha < 1, \quad 0 < \beta < \alpha \quad$ and $0 < \gamma < 1 - \alpha$.
- $0.8 < \phi < 0.98$ — to prevent numerical difficulties.

# Parameter restrictions

## *Admissible* region

- To prevent observations in the distant past having a continuing effect on current forecasts.
- Usually (but not always) less restrictive than the *traditional* region.
- For example for ETS(A,N,N):
  *traditional* $0 < \alpha < 1$ — *admissible* is $0 < \alpha < 2$.

# Model selection

## Akaike's Information Criterion

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

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

## Corrected AIC

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

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

## Bayesian Information Criterion

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

# AIC and cross-validation

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

# Automatic forecasting

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

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

Method performed very well in M3 competition.

# Some unstable models

- Some of the combinations of (Error, Trend, Seasonal) can lead to numerical difficulties; see equations with division by a state.
- These are: ETS(A,N,M), ETS(A,A,M), ETS(A,$A_d$,M).
- Models with multiplicative errors are useful for strictly positive data, but are not numerically stable with data containing zeros or negative values. In that case only the six fully additive models will be applied.

# Exponential smoothing models

| **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 |

# Example: Australian holiday tourism

```
fit <- aus_holidays %>% model(ETS(Trips))
report(fit)
```

```
## Series: Trips
## Model: ETS(M,N,M)
##   Smoothing parameters:
##     alpha = 0.3578
##     gamma = 0.0009686
##
##   Initial states:
##     l     s1     s2     s3    s4
##  9667 0.943 0.9268 0.9684 1.162
##
##   sigma^2:  0.0022
##
##  AIC AICc  BIC
## 1331 1333 1348
```

# Example: Australian holiday tourism

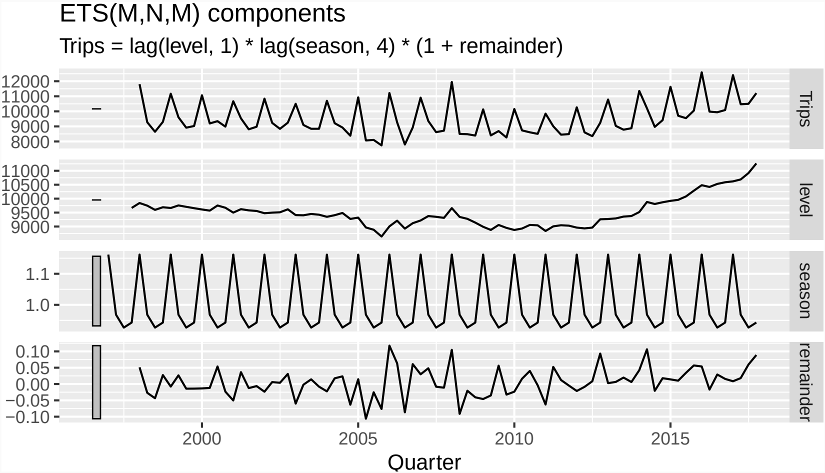
Model selected: ETS(M,N,M)

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

$\hat{\alpha} = 0.3578$, and $\hat{\gamma} = 0.000969$.

# Example: Australian holiday tourism

```
components(fit)
```



# Residuals

## Response residuals

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

## Innovation residuals

Additive error model:

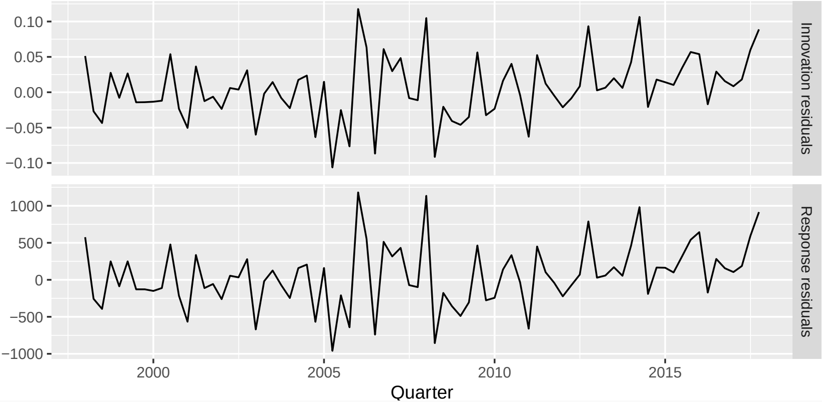
$$\hat{\varepsilon}_t = y_t - \hat{y}_{t|t-1}$$

Multiplicative error model:

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

# Example: Australian holiday tourism

```r
residuals(fit)
residuals(fit, type = "response")
```



# Outline

# Forecasting with ETS models

Point forecasts: iterate the equations for $t = T + 1, T + 2, \ldots, T + h$ and set all $\varepsilon_t = 0$ for $t > T$.

- Not the same as $\mathrm{E}(y_{t+h}|\boldsymbol{x}_t)$ unless trend and seasonality are both additive.
- Point forecasts for ETS(A,*,*) are identical to ETS(M,*,*) if the parameters are the same.

# Example: ETS(A,A,N)

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

etc.

# Example: ETS(M,A,N)

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

etc.

# Forecasting with ETS models

Prediction intervals: can only generated using the models.

- The prediction intervals will differ between models with additive and multiplicative errors.
- Exact formulae for some models.
- 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.

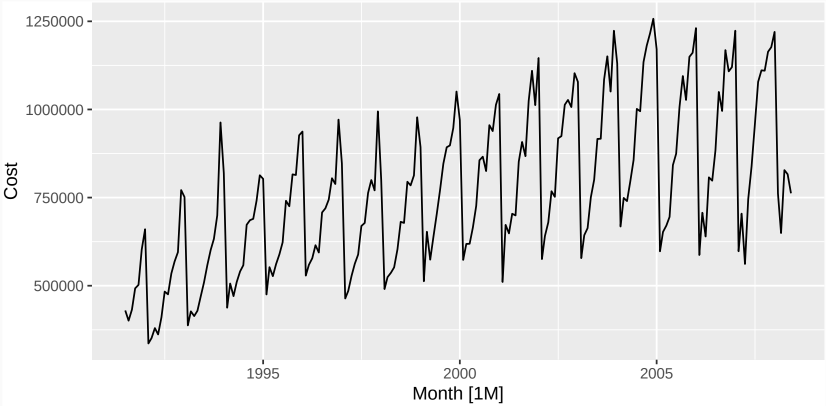
# Prediction intervals

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

| | |
|---|---|
| (A,N,N) | $\sigma_h = \sigma^2\big[1 + \alpha^2(h-1)\big]$ |
| (A,A,N) | $\sigma_h = \sigma^2\Big[1 + (h-1)\big\{\alpha^2 + \alpha\beta h + \frac{1}{6}\beta^2 h(2h-1)\big\}\Big]$ |
| (A,A$_d$,N) | $\sigma_h = \sigma^2\Big[1 + \alpha^2(h-1) + \frac{\beta\phi h}{(1-\phi)^2}\left\{2\alpha(1-\phi) + \beta\phi\right\} - \frac{\beta\phi(1-\phi^h)}{(1-\phi)^2(1-\phi^2)}\left\{2\alpha(1-\phi^2) + \beta\phi(1+2\phi-\phi^h)\right\}\Big]$ |
| (A,N,A) | $\sigma_h = \sigma^2\Big[1 + \alpha^2(h-1) + \gamma k(2\alpha+\gamma)\Big]$ |
| (A,A,A) | $\sigma_h = \sigma^2\Big[1 + (h-1)\big\{\alpha^2 + \alpha\beta h + \frac{1}{6}\beta^2 h(2h-1)\big\} + \gamma k\big\{2\alpha + \gamma + \beta m(k$ |
| (A,A$_d$,A) | $\sigma_h = \sigma^2\Big[1 + \alpha^2(h-1) + \frac{\beta\phi h}{(1-\phi)^2}\left\{2\alpha(1-\phi) + \beta\phi\right\} - \frac{\beta\phi(1-\phi^h)}{(1-\phi)^2(1-\phi^2)}\left\{2\alpha(1-\phi^2) + \beta\phi(1+2\phi-\phi^h)\right\} + \gamma k(2\alpha+\gamma) + \frac{2\beta\gamma\phi}{(1-\phi)(1-\phi^m)}\left\{k(1-\phi^m) - \phi^m(1-\phi^{mk})\right\}\Big]$ |

# Example: Corticosteroid drug sales

```
h02 <- tsibbledata::PBS %>%
  filter(ATC2 == "H02") %>%
  summarise(Cost = sum(Cost))
h02 %>%
  autoplot(Cost)
```



# Example: Corticosteroid drug sales

```r
h02 %>% model(ETS(Cost)) %>% report
```

```
## Series: Cost
## Model: ETS(M,Ad,M)
##   Smoothing parameters:
##     alpha = 0.3071
##     beta  = 0.0001007
##     gamma = 0.0001007
##     phi   = 0.9775
##
##   Initial states:
##        l    b     s1    s2     s3     s4     s5    s6
##   417269 8206 0.8717 0.826 0.7563 0.7733 0.6872 1.284
##     s7   s8    s9   s10   s11    s12
##   1.325 1.18 1.164 1.105 1.048 0.9806
##
##   sigma^2:  0.0046
##
##   AIC AICc  BIC
## 5515 5519 5575
```

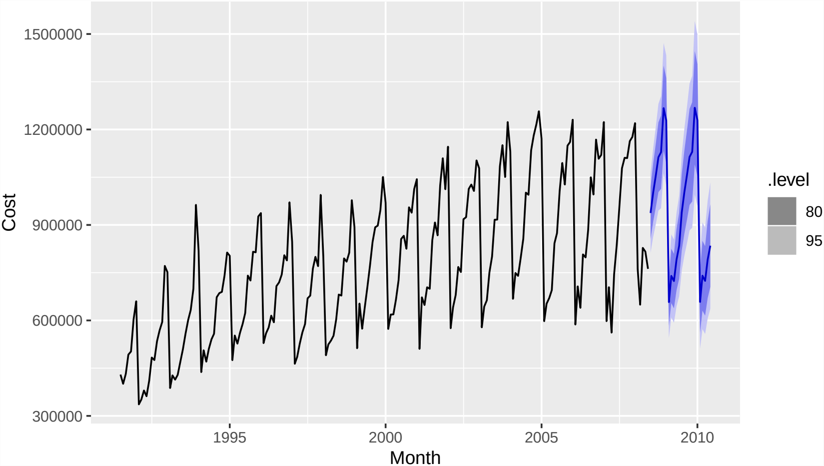
# Example: Corticosteroid drug sales

```
h02 %>% model(ETS(Cost ~ error("A") + trend("A") + season("A"))) %>% report
```

```
## Series: Cost
## Model: ETS(A,A,A)
##   Smoothing parameters:
##     alpha = 0.1702
##     beta  = 0.006311
##     gamma = 0.4546
##
##   Initial states:
##       l    b     s1      s2      s3      s4      s5
##  409706 9097 -99075 -136602 -191496 -174531 -241437
##      s6     s7     s8     s9    s10    s11    s12
##  210644 244644 145368 130570 84458 39132 -11674
##
##   sigma^2:  3.499e+09
##
##  AIC AICc  BIC
## 5585 5589 5642
```

# Example: Corticosteroid drug sales

```
h02 %>% model(ETS(Cost)) %>% forecast() %>% autoplot(h02)
```



# Example: Corticosteroid drug sales

```
h02 %>%
  model(
    auto = ETS(Cost),
    AAA = ETS(Cost ~ error("A") + trend("A") + season("A"))
  ) %>%
  accuracy()
```

| Model | ME | MAE | RMSE | MAPE | MASE |
|---|---|---|---|---|---|
| auto | 2461 | 38649 | 51102 | 4.989 | 0.6376 |
| AAA | -5780 | 43378 | 56784 | 6.048 | 0.7156 |

# Your turn

- Use `ETS()` on some of these series:

  *`tourism`*, *`gafa_stock`*, *`pelt`*

- Does it always give good forecasts?
- Find an example where it does not work well. Can you figure out why?