



ETC3550

Applied forecasting for business and economics

Ch7. Exponential smoothing

OTexts.org/fpp3/

Outline

- 1 Exponential smoothing
- 2 The level
- 3 The trend
- 4 Seasonal methods
- 5 Innovations state space models
- 6 Forecasting with ETS models

Outline

- 1 Exponential smoothing
- 2 The level
- 3 The trend
- 4 Seasonal methods
- 5 Innovations state space models
- 6 Forecasting with ETS models

Simple methods

Time series y_1, y_2, \dots, y_T .

Random walk forecasts

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

Simple methods

Time series y_1, y_2, \dots, y_T .

Random walk forecasts

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

Average forecasts

$$\hat{y}_{T+h|T} = \frac{1}{T} \sum_{t=1}^T y_t$$

Simple methods

Time series y_1, y_2, \dots, y_T .

Random walk forecasts

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

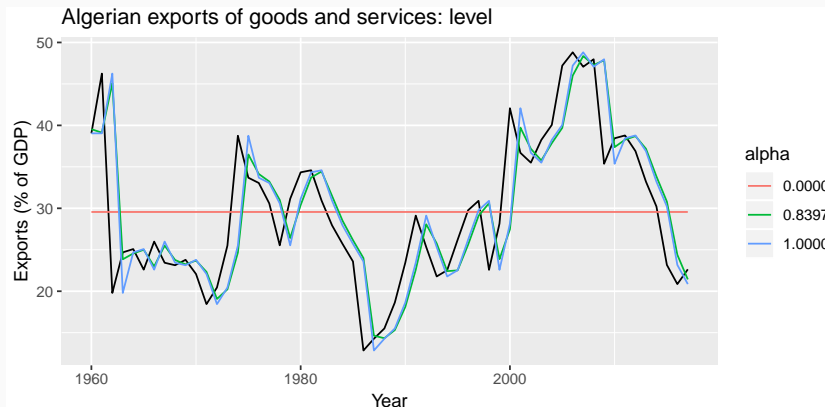
Average forecasts

$$\hat{y}_{T+h|T} = \frac{1}{T} \sum_{t=1}^T y_t$$

- Want something in between these methods.
- Most recent data should have more weight.

Exponential smoothing: level/intercept

Exponential smoothing: level/intercept



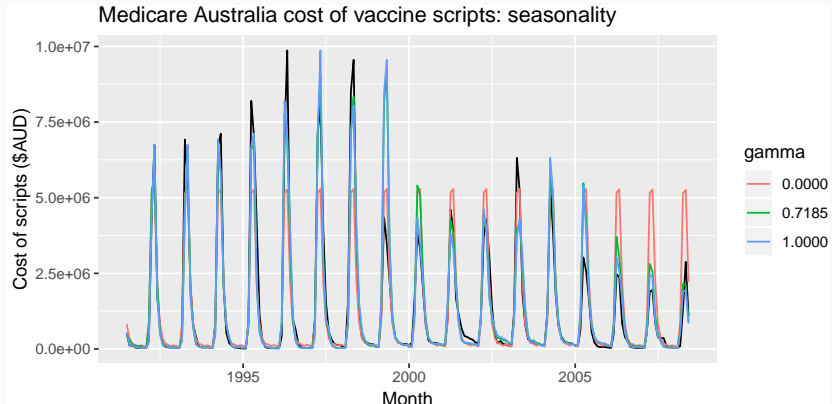
Exponential smoothing: trend/slope

Exponential smoothing: trend/slope



Exponential smoothing: seasonality

Exponential smoothing: seasonality



Big idea: control the rate of change (smoothing)

α controls the flexibility of the **level**

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

β controls the flexibility of the **trend**

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

γ controls the flexibility of the **seasonality**

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

Optimising smoothing parameters

- Need to choose best values for the smoothing parameters (and initial states).
- 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.

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?

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

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?

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

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?

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

Exponential smoothing

General notation

E T S : ExponenTial Smoothing
 ↗ ↑ ↖
Error Trend Season

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

Exponential smoothing

General notation

E T S : ExponenTial Smoothing
 ↑ ↑ ↑
Error Trend Season

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

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

Exponential smoothing

General notation

ETS : ExponenTial Smoothing
 ↑ ↑ ↑
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")

Outline

- 1 Exponential smoothing
- 2 The level
- 3 The trend
- 4 Seasonal methods
- 5 Innovations state space models
- 6 Forecasting with ETS models

Starting simple

Suppose our model contains no trend or seasonality.
This is compactly represented by ETS(A,N,N).

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

We design ℓ_t to weight recent observations more.

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

Need to choose α and ℓ_0 .

Simple Exponential Smoothing

Component form

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

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

- ℓ_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

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

Simple Exponential Smoothing

Forecast equation

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

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

Simple Exponential Smoothing

Forecast equation

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

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

| Observation | Weights assigned to observations for: | | | |
|-------------|---------------------------------------|----------------|----------------|----------------|
| | $\alpha = 0.2$ | $\alpha = 0.4$ | $\alpha = 0.6$ | $\alpha = 0.8$ |
| 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} | $(0.2)(0.8)^5$ | $(0.4)(0.6)^5$ | $(0.6)(0.4)^5$ | $(0.8)(0.2)^5$ |

Methods v Models

Methods

- Algorithms that return point forecasts.

Methods v Models

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): A model for SES

Component form

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

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

ETS(A,N,N): A model for SES

Component form

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

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

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

ETS(A,N,N): A model for SES

Component form

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

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

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

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): A model for SES

Component form

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

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

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

Error correction form

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

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

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

Specify probability distribution for e_t , we assume

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

ETS(A,N,N)

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, ε_t .
- Measurement equation: relationship between observations and states.
- Transition/state equation(s): evolution of the state(s) through time.

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 $\hat{y}_{t|t-1} = l_{t-1}$ gives:
 - ▶ $y_t = l_{t-1} + l_{t-1}\varepsilon_t$
 - ▶ $e_t = y_t - \hat{y}_{t|t-1} = l_{t-1}\varepsilon_t$

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 $\hat{y}_{t|t-1} = l_{t-1}$ gives:
 - ▶ $y_t = l_{t-1} + l_{t-1}\varepsilon_t$
 - ▶ $e_t = y_t - \hat{y}_{t|t-1} = l_{t-1}\varepsilon_t$

Measurement equation

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

State equation

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

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 $\hat{y}_{t|t-1} = l_{t-1}$ gives:
 - ▶ $y_t = l_{t-1} + l_{t-1}\varepsilon_t$
 - ▶ $e_t = y_t - \hat{y}_{t|t-1} = l_{t-1}\varepsilon_t$

| | |
|----------------------|------------------------------------|
| Measurement equation | $y_t = l_{t-1}(1 + \varepsilon_t)$ |
|----------------------|------------------------------------|

| | |
|----------------|--|
| State equation | $l_t = l_{t-1}(1 + \alpha\varepsilon_t)$ |
|----------------|--|

- Models with additive and multiplicative errors with the same parameters generate the same point forecasts but different prediction intervals. 22

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

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

By default, an optimal value for α and ℓ_0 is used.

α can be chosen manually in `trend()`.

```
trend("N", alpha = 0.5)
```

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

Example: Algerian Exports

```
algeria_economy <- tsibbledata::global_economy %>%  
  filter(Country == "Algeria")  
fit <- algeria_economy %>%  
  model(ANN = ETS(Exports ~ error("A") + trend("N") + season("N")))  
report(fit)
```

```
## Series: Exports  
## Model: ETS(A,N,N)  
## Smoothing parameters:  
##   alpha = 0.84  
##  
## Initial states:  
##   l  
## 39.54  
##  
##  
## sigma: 5.969  
##  
## AIC AICc BIC
```

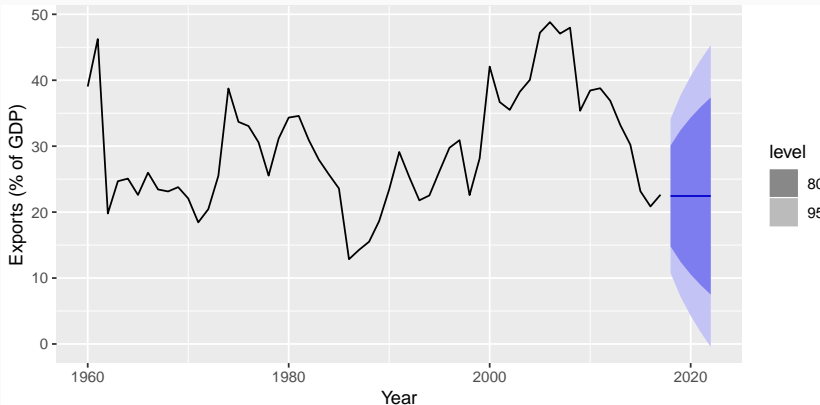
Example: Algerian Exports

```
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
```

Example: Algerian Exports

```
fit %>%  
  forecast(h = 5) %>%  
  autoplot(algeria_economy) +  
  ylab("Exports (% of GDP)") + xlab("Year")
```



Outline

- 1 Exponential smoothing
- 2 The level
- 3 The trend
- 4 Seasonal methods
- 5 Innovations state space models
- 6 Forecasting with ETS models

Adding a trend

What if our data is trended? Add b_t to the model.

This is compactly represented by ETS(A,A,N).

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

We design b_t to weight recent slopes more.

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

Need to choose α , β^* , ℓ_0 and b_0 .

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

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 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.

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

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

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

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

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

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

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

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

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

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

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

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

By default, an optimal value for β and b_0 is used.

β can be chosen manually in `trend()`.

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

Example: Australian population

```
aus_economy <- global_economy %>% filter(Code == "AUS")
fit <- aus_economy %>%
  model(AAN = ETS(Population ~
    error("A") + trend("A") + season("N")))
report(fit)
```

```
## Series: Population
## Model: ETS(A,A,N)
## Smoothing parameters:
##   alpha = 0.9999
##   beta  = 0.3257
##
## Initial states:
##      l      b
## 10067191 228013
##
##
## sigma: 64240
```

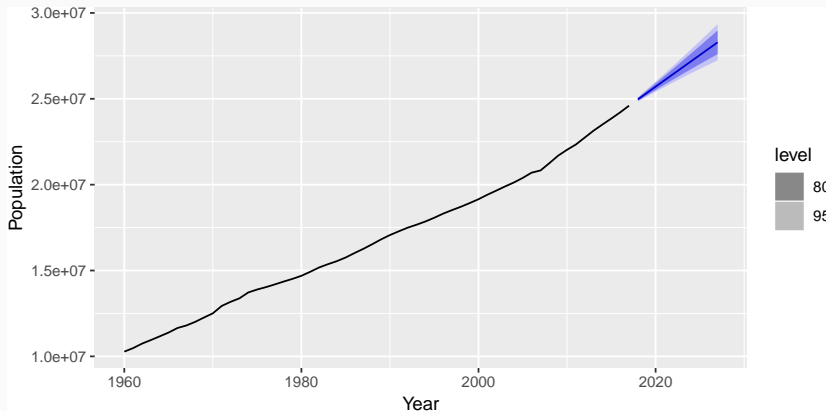

Example: Australian population

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

```
## # A tsibble: 59 x 8 [1Y]  
## # Key:      Country, .model [1]  
##   Country .model Year Population level slope  
##   <fct>   <chr>  <dbl>      <dbl>  <dbl>  <dbl>  
## 1 Austr~ AAN    1959         NA  1.01e7 2.28e5  
## 2 Austr~ AAN    1960    10276477  1.03e7 2.22e5  
## 3 Austr~ AAN    1961    10483000  1.05e7 2.17e5  
## 4 Austr~ AAN    1962    10742000  1.07e7 2.31e5  
## 5 Austr~ AAN    1963    10950000  1.10e7 2.23e5  
## 6 Austr~ AAN    1964    11167000  1.12e7 2.21e5  
## 7 Austr~ AAN    1965    11388000  1.14e7 2.21e5  
## 8 Austr~ AAN    1966    11651000  1.17e7 2.35e5  
## 9 Austr~ AAN    1967    11799000  1.18e7 2.07e5  
## 10 Austr~ AAN    1968    12009000  1.20e7 2.08e5  
## # with 49 more rows and 2 more variables:
```

Example: Australian population

```
fit %>%  
  forecast(h = 10) %>%  
  autoplot(aus_economy) +  
  ylab("Population") + xlab("Year")
```



Damped trend method

Component form

$$\hat{y}_{t+h|t} = \ell_t + (\phi + \phi^2 + \dots + \phi^h)b_t$$

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

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

Damped trend method

Component form

$$\hat{y}_{t+h|t} = \ell_t + (\phi + \phi^2 + \dots + \phi^h)b_t$$

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

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

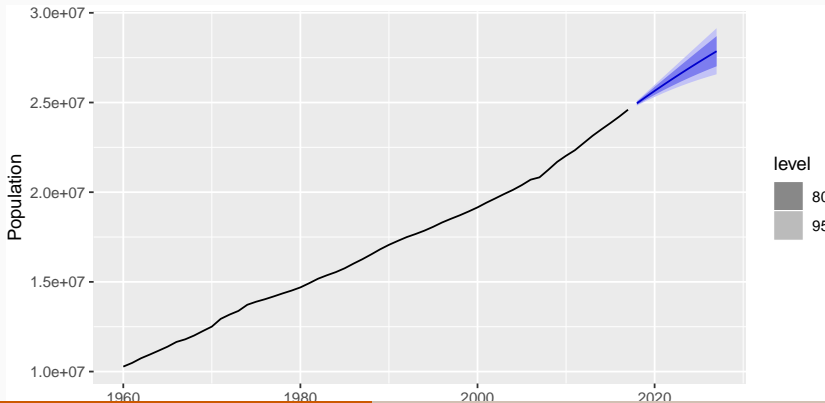
- Damping parameter $0 < \phi < 1$.
- If $\phi = 1$, identical to Holt's linear trend.
- As $h \rightarrow \infty$, $\hat{y}_{T+h|T} \rightarrow \ell_T + \phi b_T / (1 - \phi)$.
- Short-run forecasts trended, long-run forecasts constant.

Your turn

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

Example: Australian Population

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



Example: Australian Population

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

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

Example: Sheep in Asia

| term | | SES | Linear trend | Damped trend |
|---------------|-----------|-------------|--------------|--------------|
| | α | 1.00 | 1.00 | 1.00 |
| | β^* | | 0.30 | 0.42 |
| | ϕ | | | 0.98 |
| | ℓ_0 | 11272847.32 | 10067190.61 | 10067190.04 |
| | b_0 | | 228458.45 | 277893.91 |
| Training RMSE | | 279368.61 | 64642.01 | 67111.29 |
| Test RMSE | | 1632602.00 | 147969.40 | 196836.92 |
| Test MASE | | 6.18 | 0.55 | 0.71 |
| Test MAPE | | 6.09 | 0.55 | 0.70 |
| Test MAE | | 1453007.44 | 130252.30 | 166362.49 |

Your turn

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

- 1 Use 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 look like white noise?

Outline

- 1 Exponential smoothing
- 2 The level
- 3 The trend
- 4 Seasonal methods
- 5 Innovations state space models
- 6 Forecasting with ETS models

Adding seasonality

What if our data is seasonal? Add s_t to the model.

This is compactly represented by ETS(A,A,A).

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

We design s_t to weight slopes more.

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

Need to choose $\alpha, \beta^*, \gamma, \ell_0, b_0, s_0, s_{-1}, \dots, s_{-m+1}$.

Holt-Winters additive method

Holt and Winters extended Holt's method to capture seasonality.

Component form

$$\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$$

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

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

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

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

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 ℓ_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 $0 \leq \gamma^* \leq 1$,
which translates to $0 \leq \gamma \leq (1 - \alpha)$.

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

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

$$s_t = s_{t-m} + \gamma\varepsilon_t$$

- Forecast errors: $\varepsilon_t = y_t - \hat{y}_{t|t-1}$
- k is integer part of $(h - 1)/m$.

Your turn

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

Holt-Winters multiplicative method

For when seasonal variations are changing proportional to the level of the series.

Component form

$$\hat{y}_{t+h|t} = (\ell_t + hb_t)s_{t+h-m(k+1)}$$

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

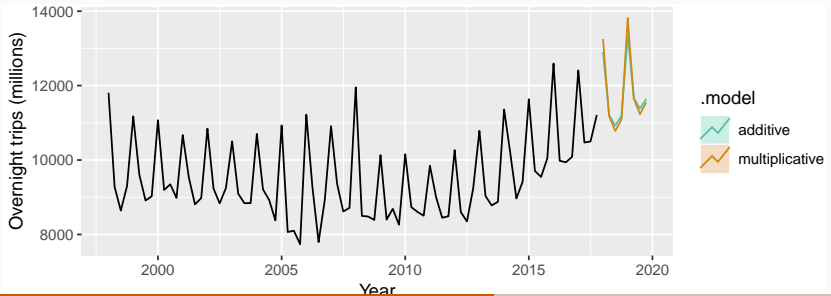
$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$$

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

- 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 $\sum_i s_i \approx m$.

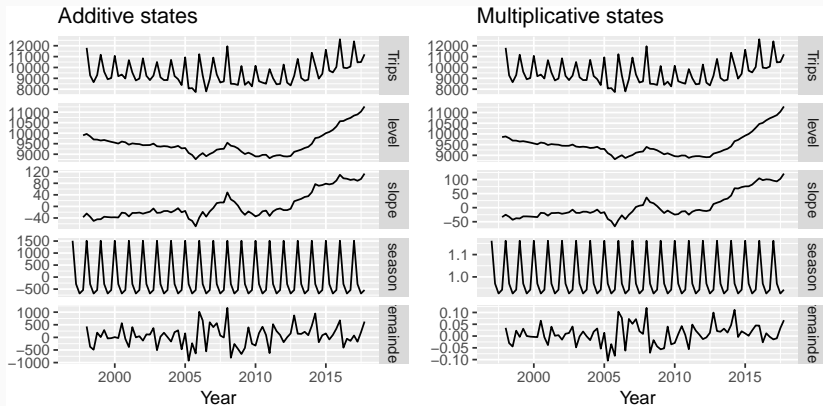
Example: Australian holiday tourism

```
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:

$$\hat{y}_{t+h|t} = [\ell_t + (\phi + \phi^2 + \dots + \phi^h)b_t]s_{t+h-m(k+1)}$$

$$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$$

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

$$s_t = \gamma \frac{y_t}{(\ell_{t-1} + \phi b_{t-1})} + (1 - \gamma)s_{t-m}$$

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

- 1 Exponential smoothing
- 2 The level
- 3 The trend
- 4 Seasonal methods
- 5 Innovations state space models
- 6 Forecasting with ETS models

Exponential smoothing methods

| | | Seasonal Component | | |
|-----------------|-------------------|---------------------|---------------------|---------------------|
| | | N | A | M |
| Trend Component | | (None) | (Additive) | (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

- Each model has an *observation* equation and *transition* equations, one for each state (level, trend, seasonal), i.e., state space models.
- Two models for each method: one with additive and one with multiplicative errors, i.e., in total 18 models.
- ETS(Error,Trend,Seasonal):
 - ▶ Error = $\{A, M\}$
 - ▶ Trend = $\{N, A, A_d\}$
 - ▶ Seasonal = $\{N, A, M\}$.

Additive error models

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

Multiplicative error models

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

Estimating ETS models

- Smoothing parameters α , β , γ and ϕ , and the initial states ℓ_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 $\mathbf{x}_t = (\ell_t, b_t, s_t, s_{t-1}, \dots, s_{t-m+1})$ and $\varepsilon_t \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$.

$$y_t = \underbrace{h(\mathbf{x}_{t-1})}_{\mu_t} + \underbrace{k(\mathbf{x}_{t-1})\varepsilon_t}_{e_t}$$

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}) + g(\mathbf{x}_{t-1})\varepsilon_t$$

Additive errors

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

Multiplicative errors

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

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

Innovations state space models

Estimation

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

- Estimate parameters $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \phi)$ and initial states $\mathbf{x}_0 = (\ell_0, b_0, s_0, s_{-1}, \dots, s_{-m+1})$ 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 $0 < \alpha < 1$, $0 < \beta < \alpha$ and $0 < \gamma < 1 - \alpha$.
- $0.8 < \phi < 0.98$ — to prevent numerical difficulties.

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 $0 < \alpha < 1$, $0 < \beta < \alpha$ and $0 < \gamma < 1 - \alpha$.
- $0.8 < \phi < 0.98$ — to prevent numerical difficulties.

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

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

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

Model selection

Akaike's Information Criterion

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

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

Corrected AIC

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

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

Model selection

Akaike's Information Criterion

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

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

Corrected AIC

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

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 | | |
|-----------------|----------------------------------|---------------------|---------------------|------------------------------|
| | | N (None) | A (Additive) | M (Multiplicative) |
| Trend Component | 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 | | |
|-----------------|----------------------------------|---------------------|---------------------|-----------------------|
| | | N (None) | A (Additive) | M (Multiplicative) |
| Trend Component | 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 |

The ETS() function

- Automatically chooses a model by default using the AIC, AICc or BIC.
- Can handle any combination of trend, seasonality and damping
- Ensures the parameters are admissible (equivalent to invertible)

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: 0.0464  
##  
## AIC AICc BIC  
## 1331 1333 1348
```

Example: International tourists

Model selected: ETS(M,N,M)

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

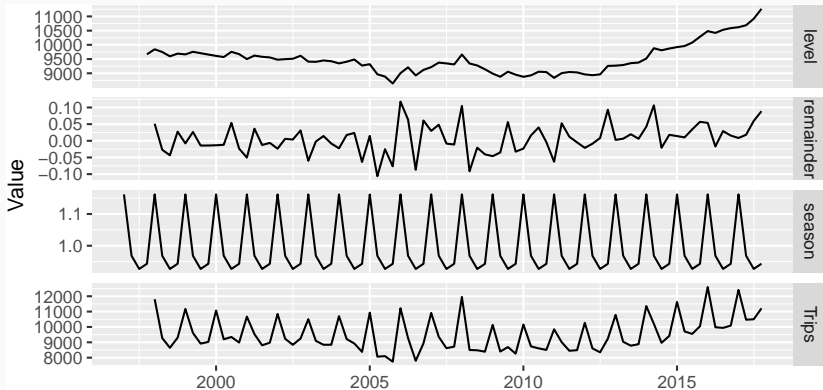
$$l_t = l_{t-1}(1 + \alpha\varepsilon_t)$$

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

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

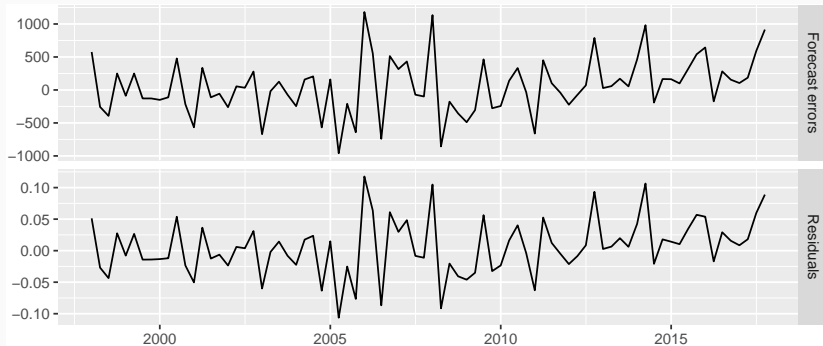
Example: International tourists

```
components(fit) %>%  
  gather("Term", "Value", -.model, -Quarter) %>%  
  ggplot(aes(x = Quarter, y = Value)) +  
  geom_line() +  
  facet_grid(Term ~ ., scales = "free_y")
```



Example: International tourists

```
residuals(fit) %>%  
  mutate(Type = "Residuals") %>%  
  bind_rows(residuals(fit, type = "response") %>%  
    mutate(Type = "Forecast errors")) %>%  
  ggplot(aes(x = Quarter, y = .resid)) +  
    geom_line() + ylab(NULL) +  
    facet_grid(Type ~ ., scales = "free_y")
```



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: drug sales

```
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: 0.0678
##
```

Example: 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: 59151  
##
```

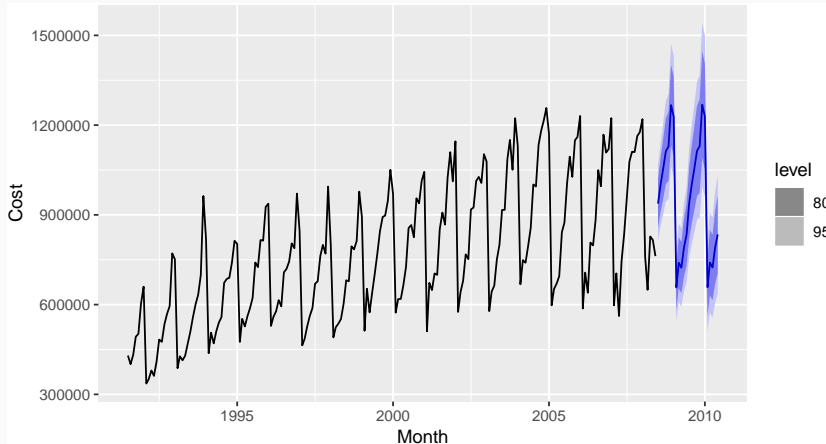
Example: drug sales

```
h02 %>% model(ETS(Cost)) %>% components()
```

```
## # A dable:                216 x 7 [1M]
## # Key:                    .model [1]
## # ETS(M,Ad,M) Decomposition: Cost = (lag(level, 1) +
## #   0.977528003332348 * lag(slope, 1)) * lag(season,
## #   12) * (1 + remainder)
##   .model      Month  Cost level slope season remainder
##   <chr>        <mth> <dbl> <dbl> <dbl>  <dbl>      <dbl>
## 1 ETS(C~      1990 Jul   NA    NA    NA    0.981      NA
## 2 ETS(C~      1990 Aug   NA    NA    NA    1.05       NA
## 3 ETS(C~      1990 Sep   NA    NA    NA    1.10       NA
## 4 ETS(C~      1990 Oct   NA    NA    NA    1.16       NA
## 5 ETS(C~      1990 Nov   NA    NA    NA    1.18       NA
## 6 ETS(C~      1990 Dec   NA    NA    NA    1.32       NA
## 7 ETS(C~      1991 Jan   NA    NA    NA    1.28       NA
## 8 ETS(C~      1991 Feb   NA    NA    NA    0.687      NA
## 9 ETS(C~      1991 Mar   NA    NA    NA    0.773      NA
## 10 ETS(C~     1991 Apr   NA    NA    NA    0.756      NA
```

Example: drug sales

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



Example: drug sales

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

```
## # A tibble: 2 x 9  
##   .model .type      ME    RMSE    MAE      MPE    MAPE    MASE  
##   <chr>  <chr>    <dbl>  <dbl>   <dbl>   <dbl>  <dbl>  <dbl>  
## 1 auto   Trai~  2461. 51102. 38649. -0.0127  4.99  0.638  
## 2 AAA    Trai~ -5780. 56784. 43378. -1.30    6.05  0.716  
## # ... with 1 more variable: ACF1 <dbl>
```


Outline

- 1 Exponential smoothing
- 2 The level
- 3 The trend
- 4 Seasonal methods
- 5 Innovations state space models
- 6 Forecasting with ETS models

Forecasting with ETS models

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

Forecasting with ETS models

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

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

Example: ETS(A,A,N)

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

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

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

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

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

etc.

Example: ETS(M,A,N)

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

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

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

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

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

etc.

Forecasting with ETS models

Prediction intervals: cannot be generated using the methods, only 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 σ_h is forecast standard deviation.

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

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

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

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

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

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

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?