Functional time series
with applications in demography

2. Automatic time series forecasting
1. Functional time series

2. Exponential smoothing

3. ARIMA modelling

4. Forecasting functional time series

5. References
y_t(x_i) = s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i},

s_t(x) = \mu(x) + \sum_{k=1}^{T-1} \beta_{t,k} \phi_k(x)

1. Estimate smooth functions \( s_t(x) \) using weighted penalized regression splines.
2. Compute \( \mu(x) \) as \( \bar{s}(x) \) across years.
3. Compute \( \beta_{t,k} \) and \( \phi_k(x) \) using functional principal components.
4. To forecast \( y_t(x_i) \), we need forecasts of \( \{\beta_{t,k}\} \).
Functional principal components

Main effects

Interaction

French male mortality

Functional time series with applications in demography

2. Automatic time series forecasting
Functional principal components

Main effects

Interaction

Australian fertility

Functional time series with applications in demography

2. Automatic time series forecasting
1. Functional time series

2. Exponential smoothing

3. ARIMA modelling

4. Forecasting functional time series

5. References
## Exponential smoothing methods

### Trend Component

<table>
<thead>
<tr>
<th>Trend Component</th>
<th>Seasonal Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (None)</td>
<td>N,N, N,A, N,M</td>
</tr>
<tr>
<td>A (Additive)</td>
<td>A,N, A,A, A,M</td>
</tr>
<tr>
<td>Ad (Additive damped)</td>
<td>Ad,N, Ad,A, Ad,M</td>
</tr>
<tr>
<td>M (Multiplicative)</td>
<td>M,N, M,A, M,M</td>
</tr>
<tr>
<td>Md (Multiplicative damped)</td>
<td>Md,N, Md,A, Md,M</td>
</tr>
</tbody>
</table>

### Functional time series with applications in demography

2. Automatic time series forecasting
## Exponential smoothing methods

<table>
<thead>
<tr>
<th>Trend Component</th>
<th>Seasonal Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (None)</td>
<td>N,N</td>
</tr>
<tr>
<td>A (Additive)</td>
<td>A,N</td>
</tr>
<tr>
<td>Ad (Additive damped)</td>
<td>Ad,N</td>
</tr>
<tr>
<td>M (Multiplicative)</td>
<td>M,N</td>
</tr>
<tr>
<td>Md (Multiplicative damped)</td>
<td>Md,N</td>
</tr>
<tr>
<td></td>
<td>N,A</td>
</tr>
<tr>
<td></td>
<td>A,A</td>
</tr>
<tr>
<td></td>
<td>Ad,A</td>
</tr>
<tr>
<td></td>
<td>M,A</td>
</tr>
<tr>
<td></td>
<td>M,A</td>
</tr>
<tr>
<td></td>
<td>Md,A</td>
</tr>
<tr>
<td></td>
<td>N,M</td>
</tr>
<tr>
<td></td>
<td>A,M</td>
</tr>
<tr>
<td></td>
<td>Ad,M</td>
</tr>
<tr>
<td></td>
<td>M,M</td>
</tr>
<tr>
<td></td>
<td>Md,M</td>
</tr>
</tbody>
</table>

### Trend Component
- **N**: Simple exponential smoothing
- **A**: (Additive)
- **Ad**: (Additive damped)
- **M**: (Multiplicative)
- **Md**: (Multiplicative damped)

### Seasonal Component
- **None**: N
- **Additive**: A
- **Multiplicative**: M
## Exponential smoothing methods

<table>
<thead>
<tr>
<th>Trend Component</th>
<th>Seasonal Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (None)</td>
<td>N,N (None)</td>
</tr>
<tr>
<td>A (Additive)</td>
<td>A,N (Additive)</td>
</tr>
<tr>
<td>Ad (Additive damped)</td>
<td>Ad,N (Additive)</td>
</tr>
<tr>
<td>M (Multiplicative)</td>
<td>M,N (Multiplicative)</td>
</tr>
<tr>
<td>Md (Multiplicative damped)</td>
<td>Md,N (Multiplicative)</td>
</tr>
</tbody>
</table>

- **N,N:** Simple exponential smoothing
- **A,N:** Holt’s linear method
## Exponential smoothing methods

<table>
<thead>
<tr>
<th>Trend Component</th>
<th>Seasonal Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (None)</td>
<td>N,N</td>
</tr>
<tr>
<td>A (Additive)</td>
<td>N,A</td>
</tr>
<tr>
<td>Ad (Additive damped)</td>
<td>A_{d},A</td>
</tr>
<tr>
<td>M (Multiplicative)</td>
<td>M,N</td>
</tr>
<tr>
<td>Md (Multiplicative damped)</td>
<td>M_{d},N</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>N (None)</th>
<th>A (Additive)</th>
<th>M (Multiplicative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N,N</td>
<td>N,N</td>
<td>N,A</td>
<td>N,M</td>
</tr>
<tr>
<td>A,N</td>
<td>A,N</td>
<td>A,A</td>
<td>A,M</td>
</tr>
<tr>
<td>A_{d},N</td>
<td>A_{d},A</td>
<td>A_{d},A</td>
<td>A_{d},M</td>
</tr>
<tr>
<td>M,N</td>
<td>M,N</td>
<td>M,A</td>
<td>M,M</td>
</tr>
<tr>
<td>M_{d},N</td>
<td>M_{d},N</td>
<td>M_{d},A</td>
<td>M_{d},M</td>
</tr>
</tbody>
</table>

- **N,N:** Simple exponential smoothing
- **A,N:** Holt’s linear method
- **A_{d},N:** Additive damped trend method
### Exponential smoothing methods

<table>
<thead>
<tr>
<th>Trend Component</th>
<th>Seasonal Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (None)</td>
<td>N,N</td>
</tr>
<tr>
<td>A (Additive)</td>
<td>A,N</td>
</tr>
<tr>
<td>A_d (Additive damped)</td>
<td>A_d,N</td>
</tr>
<tr>
<td>M (Multiplicative)</td>
<td>M,N</td>
</tr>
<tr>
<td>M_d (Multiplicative damped)</td>
<td>M_d,N</td>
</tr>
</tbody>
</table>

- **N,N**: Simple exponential smoothing
- **A,N**: Holt’s linear method
- **A_d,N**: Additive damped trend method
- **M,N**: Exponential trend method

**Functional time series with applications in demography**

2. Automatic time series forecasting
# Exponential smoothing methods

<table>
<thead>
<tr>
<th>Trend Component</th>
<th>Seasonal Component</th>
<th>( N ) (None)</th>
<th>( A ) (Additive)</th>
<th>( M ) (Multiplicative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N ) (None)</td>
<td>( N,N )</td>
<td>( N,A )</td>
<td>( N,M )</td>
<td></td>
</tr>
<tr>
<td>( A ) (Additive)</td>
<td>( A,N )</td>
<td>( A,A )</td>
<td>( A,M )</td>
<td></td>
</tr>
<tr>
<td>( A_d ) (Additive damped)</td>
<td>( A_d,N )</td>
<td>( A_d,A )</td>
<td>( A_d,M )</td>
<td></td>
</tr>
<tr>
<td>( M ) (Multiplicative)</td>
<td>( M,N )</td>
<td>( M,A )</td>
<td>( M,M )</td>
<td></td>
</tr>
<tr>
<td>( M_d ) (Multiplicative damped)</td>
<td>( M_d,N )</td>
<td>( M_d,A )</td>
<td>( M_d,M )</td>
<td></td>
</tr>
</tbody>
</table>

- \( N,N \): Simple exponential smoothing
- \( A,N \): Holt’s linear method
- \( A_d,N \): Additive damped trend method
- \( M,N \): Exponential trend method
- \( M_d,N \): Multiplicative damped trend method
### Trend Component

<table>
<thead>
<tr>
<th>Trend Component</th>
<th>Seasonal Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (None)</td>
<td>N,N</td>
</tr>
<tr>
<td>A (Additive)</td>
<td>A,N</td>
</tr>
<tr>
<td>Ad (Additive damped)</td>
<td>Ad,N</td>
</tr>
<tr>
<td>M (Multiplicative)</td>
<td>M,N</td>
</tr>
<tr>
<td>Md (Multiplicative damped)</td>
<td>Md,N</td>
</tr>
</tbody>
</table>

### Seasonal Component

- **N (None)**
- **A (Additive)**
- **M (Multiplicative)**
- **Ad (Additive damped)**
- **Md (Multiplicative damped)**

### Explanation

- **N,N**: Simple exponential smoothing
- **A,N**: Holt’s linear method
- **Ad,N**: Additive damped trend method
- **M,N**: Exponential trend method
- **Md,N**: Multiplicative damped trend method
- **A,A**: Additive Holt-Winters’ method

| N,N: Simple exponential smoothing |
| A,N: Holt’s linear method |
| Ad,N: Additive damped trend method |
| M,N: Exponential trend method |
| Md,N: Multiplicative damped trend method |
| A,A: Additive Holt-Winters’ method |
### Exponential smoothing methods

<table>
<thead>
<tr>
<th>Trend Component</th>
<th>Seasonal Component</th>
<th>N (None)</th>
<th>A (Additive)</th>
<th>M (Multiplicative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (None)</td>
<td>N,N</td>
<td>N,A</td>
<td>N,M</td>
<td></td>
</tr>
<tr>
<td>A (Additive)</td>
<td>A,N</td>
<td>A,A</td>
<td>A,M</td>
<td></td>
</tr>
<tr>
<td>Ad (Additive damped)</td>
<td>Ad,N</td>
<td>Ad,A</td>
<td>Ad,M</td>
<td></td>
</tr>
<tr>
<td>M (Multiplicative)</td>
<td>M,N</td>
<td>M,A</td>
<td>M,M</td>
<td></td>
</tr>
<tr>
<td>Md (Multiplicative damped)</td>
<td>Md,N</td>
<td>Md,A</td>
<td>Md,M</td>
<td></td>
</tr>
</tbody>
</table>

- **N,N**: Simple exponential smoothing
- **A,N**: Holt’s linear method
- **Ad,N**: Additive damped trend method
- **M,N**: Exponential trend method
- **Md,N**: Multiplicative damped trend method
- **A,A**: Additive Holt-Winters’ method
- **A,M**: Multiplicative Holt-Winters’ method
### Exponential smoothing methods

<table>
<thead>
<tr>
<th>Trend Component</th>
<th>Seasonal Component</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N (None)</td>
</tr>
<tr>
<td>N (None)</td>
<td>N,N</td>
</tr>
<tr>
<td>A (Additive)</td>
<td>A,N</td>
</tr>
<tr>
<td>A (Additive damped)</td>
<td>A_d,N</td>
</tr>
<tr>
<td>M (Multiplicative)</td>
<td>M,N</td>
</tr>
<tr>
<td>M (Multiplicative damped)</td>
<td>M_d,N</td>
</tr>
<tr>
<td></td>
<td>A (Additive)</td>
</tr>
<tr>
<td>N (None)</td>
<td>N,A</td>
</tr>
<tr>
<td>A (Additive)</td>
<td>A,A</td>
</tr>
<tr>
<td>A (Additive damped)</td>
<td>A_d,A</td>
</tr>
<tr>
<td>M (Multiplicative)</td>
<td>M,A</td>
</tr>
<tr>
<td>M (Multiplicative damped)</td>
<td>M_d,A</td>
</tr>
<tr>
<td></td>
<td>M (Multiplicative)</td>
</tr>
<tr>
<td>N (None)</td>
<td>N,M</td>
</tr>
<tr>
<td>A (Additive)</td>
<td>A,M</td>
</tr>
<tr>
<td>A (Additive damped)</td>
<td>A_d,M</td>
</tr>
<tr>
<td>M (Multiplicative)</td>
<td>M,M</td>
</tr>
<tr>
<td>M (Multiplicative damped)</td>
<td>M_d,M</td>
</tr>
</tbody>
</table>

- There are 15 separate exponential smoothing methods.
## Exponential smoothing methods

There are 15 separate exponential smoothing methods.

Each can have an additive or multiplicative error, giving 30 separate models.

<table>
<thead>
<tr>
<th>Trend Component</th>
<th>Seasonal Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (None)</td>
<td>N,N</td>
</tr>
<tr>
<td>A (Additive)</td>
<td>A,N</td>
</tr>
<tr>
<td>A_d (Additive damped)</td>
<td>A_d,N</td>
</tr>
<tr>
<td>M (Multiplicative)</td>
<td>M,N</td>
</tr>
<tr>
<td>M_d (Multiplicative damped)</td>
<td>M_d,N</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Seasonal Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (None)</td>
</tr>
<tr>
<td>N,N</td>
</tr>
<tr>
<td>A,N</td>
</tr>
<tr>
<td>A_d,N</td>
</tr>
<tr>
<td>M,N</td>
</tr>
<tr>
<td>M_d,N</td>
</tr>
</tbody>
</table>
# Exponential smoothing methods

<table>
<thead>
<tr>
<th>Trend Component</th>
<th>Seasonal Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (None)</td>
<td>N,N, N,A, N,M</td>
</tr>
<tr>
<td>A (Additive)</td>
<td>A,N, A,A, A,M</td>
</tr>
<tr>
<td>A&lt;sub&gt;d&lt;/sub&gt; (Additive damped)</td>
<td>A&lt;sub&gt;d&lt;/sub&gt;,N, A&lt;sub&gt;d&lt;/sub&gt;,A, A&lt;sub&gt;d&lt;/sub&gt;,M</td>
</tr>
<tr>
<td>M (Multiplicative)</td>
<td>M,N, M,A, M,M</td>
</tr>
<tr>
<td>M&lt;sub&gt;d&lt;/sub&gt; (Multiplicative damped)</td>
<td>M&lt;sub&gt;d&lt;/sub&gt;,N, M&lt;sub&gt;d&lt;/sub&gt;,A, M&lt;sub&gt;d&lt;/sub&gt;,M</td>
</tr>
</tbody>
</table>

## General notation

ETS: Exponential Smoothing

- A,N,N: Simple exponential smoothing with additive errors
- A,A,N: Holt's linear method with additive errors
- M,A,M: Multiplicative Holt-Winters' method with multiplicative errors

## Functional time series with applications in demography

## 2. Automatic time series forecasting
### Exponential smoothing methods

<table>
<thead>
<tr>
<th>Trend Component</th>
<th>Seasonal Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (None)</td>
<td>N,N</td>
</tr>
<tr>
<td>A (Additive)</td>
<td>A,N</td>
</tr>
<tr>
<td>A&lt;sub&gt;d&lt;/sub&gt; (Additive damped)</td>
<td>A&lt;sub&gt;d&lt;/sub&gt;,N</td>
</tr>
<tr>
<td>M (Multiplicative)</td>
<td>M,N</td>
</tr>
<tr>
<td>M&lt;sub&gt;d&lt;/sub&gt; (Multiplicative damped)</td>
<td>M&lt;sub&gt;d&lt;/sub&gt;,N</td>
</tr>
<tr>
<td></td>
<td>N,A</td>
</tr>
<tr>
<td></td>
<td>A,A</td>
</tr>
<tr>
<td></td>
<td>A&lt;sub&gt;d&lt;/sub&gt;,A</td>
</tr>
<tr>
<td></td>
<td>M,A</td>
</tr>
<tr>
<td></td>
<td>M&lt;sub&gt;d&lt;/sub&gt;,A</td>
</tr>
<tr>
<td></td>
<td>N,M</td>
</tr>
<tr>
<td></td>
<td>A,M</td>
</tr>
<tr>
<td></td>
<td>A&lt;sub&gt;d&lt;/sub&gt;,M</td>
</tr>
<tr>
<td></td>
<td>M,M</td>
</tr>
<tr>
<td></td>
<td>M&lt;sub&gt;d&lt;/sub&gt;,M</td>
</tr>
</tbody>
</table>

**General notation**

ETS : Exponential Smoothing

**Examples:**
- A,N,N: Simple exponential smoothing with additive errors
- A,A,N: Holt’s linear method with additive errors
- M,A,N: Multiplicative Holt-Winters’ method with multiplicative errors
### Exponential smoothing methods

<table>
<thead>
<tr>
<th>Trend Component</th>
<th>Seasonal Component</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
</tr>
<tr>
<td>N (None)</td>
<td>N,N</td>
</tr>
<tr>
<td>A (Additive)</td>
<td>A,N</td>
</tr>
<tr>
<td>Ad (Additive damped)</td>
<td>Ad,N</td>
</tr>
<tr>
<td>M (Multiplicative)</td>
<td>M,N</td>
</tr>
<tr>
<td>Md (Multiplicative damped)</td>
<td>Md,N</td>
</tr>
</tbody>
</table>

**General notation**

ETS : Exponential Smoothing

↑ Trend

**Examples:**

- A,N,N: Simple exponential smoothing with additive errors
- A,A,N: Holt’s linear method with additive errors
- M,A,M: Multiplicative Holt-Winters’ method with multiplicative errors
### Exponential smoothing methods

<table>
<thead>
<tr>
<th>Trend Component</th>
<th>Seasonal Component</th>
<th>N (None)</th>
<th>A (Additive)</th>
<th>M (Multiplicative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (None)</td>
<td></td>
<td>N,N</td>
<td>N,A</td>
<td>N,M</td>
</tr>
<tr>
<td>A (Additive)</td>
<td></td>
<td>A,N</td>
<td>A,A</td>
<td>A,M</td>
</tr>
<tr>
<td>A_d (Additive damped)</td>
<td></td>
<td>A_d,N</td>
<td>A_d,A</td>
<td>A_d,M</td>
</tr>
<tr>
<td>M (Multiplicative)</td>
<td></td>
<td>M,N</td>
<td>M,A</td>
<td>M,M</td>
</tr>
<tr>
<td>M_d (Multiplicative damped)</td>
<td></td>
<td>M_d,N</td>
<td>M_d,A</td>
<td>M_d,M</td>
</tr>
</tbody>
</table>

**General notation**

ETS: Exponential Smoothing

↑ ↖

Trend Seasonal

**Examples:**

A,N,N: Simple exponential smoothing with additive errors
A,A,N: Holt’s linear method with additive errors
M,A,M: Multiplicative Holt-Winters’ method with multiplicative errors
# Exponential smoothing methods

<table>
<thead>
<tr>
<th>Trend Component</th>
<th>Seasonal Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (None)</td>
<td>N,N</td>
</tr>
<tr>
<td>A (Additive)</td>
<td>A,N</td>
</tr>
<tr>
<td>A_d (Additive damped)</td>
<td>A_d,N</td>
</tr>
<tr>
<td>M (Multiplicative)</td>
<td>M,N</td>
</tr>
<tr>
<td>M_d (Multiplicative damped)</td>
<td>M_d,N</td>
</tr>
</tbody>
</table>

**General notation**  
ETS : Exponential Smoothing

**Examples:**
- A,N,N: Simple exponential smoothing with additive errors
- A,A,N: Holt’s linear method with additive errors
- M,A,M: Multiplicative Holt-Winters’ method with multiplicative errors
Exponential smoothing methods

<table>
<thead>
<tr>
<th>Trend Component</th>
<th>Seasonal Component</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
</tr>
<tr>
<td>N (None)</td>
<td>N,N</td>
</tr>
<tr>
<td>A (Additive)</td>
<td>A,N</td>
</tr>
<tr>
<td>A_d (Additive damped)</td>
<td>A_d,N</td>
</tr>
<tr>
<td>M (Multiplicative)</td>
<td>M,N</td>
</tr>
<tr>
<td>M_d (Multiplicative damped)</td>
<td>M_d,N</td>
</tr>
</tbody>
</table>

General notation: E T S : Exponential Smoothing

Error Trend Seasonal

Examples:

A,N,N: Simple exponential smoothing with additive errors
A,A,N: Holt’s linear method with additive errors
M,A,M: Multiplicative Holt-Winters’ method with multiplicative errors
Exponential smoothing methods

General notation: ETS : Exponential Smoothing

Error Trend Seasonal

Examples:
A,N,N: Simple exponential smoothing with additive errors
A,A,N: Holt’s linear method with additive errors
M,A,M: Multiplicative Holt-Winters’ method with multiplicative errors
ETS(A,N,N)

SES with additive errors.

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

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

- Forecast errors: \( \varepsilon_t = y_t - \hat{y}_{t|t-1} \)
- “innovations” or “single source of error” because same error process, \( \varepsilon_t \).
- Observation equation: relationship between observations and states.
- State equation: evolution of the state through time.
ETS(A,N,N)

SES with additive errors.

Forecast equation
\[ \hat{y}_{t+h|t} = l_t \]

Observation equation
\[ y_t = l_{t-1} + \varepsilon_t \]

State equation
\[ l_t = l_{t-1} + \alpha \varepsilon_t \]

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

- Forecast errors: \( \varepsilon_t = y_t - \hat{y}_{t|t-1} \)
- “innovations” or “single source of error” because same error process, \( \varepsilon_t \).

- Observation equation: relationship between observations and states.
- State equation: evolution of the state 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
\[ \ell_t = \ell_{t-1} + \alpha \varepsilon_t \]

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

- Forecast errors: \( \varepsilon_t = y_t - \hat{y}_{t|t-1} \)
- “innovations” or “single source of error” because same error process, \( \varepsilon_t \).
- Observation equation: relationship between observations and states.
- State equation: evolution of the state through time.
ETS(A,N,N)

SES with additive errors.

Forecast equation
\[ \hat{y}_{t+h|t} = l_t \]

Observation equation
\[ y_t = l_{t-1} + \varepsilon_t \]

State equation
\[ l_t = l_{t-1} + \alpha \varepsilon_t \]

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

- Forecast errors: \( \varepsilon_t = y_t - \hat{y}_{t|t-1} \)
- “innovations” or “single source of error” because same error process, \( \varepsilon_t \).
- Observation equation: relationship between observations and states.
- State equation: evolution of the state through time.
ETS(M,N,N)

SES with multiplicative errors.

Forecast equation
\[ \hat{y}_{t+h|t} = l_t \]

Observation equation
\[ y_t = l_{t-1}(1 + \varepsilon_t) \]

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

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

- Relative forecast errors: \( \varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}} \)
- Models with additive and multiplicative errors with the same parameters generate the same point forecasts but different prediction intervals.
ETS(M,N,N)

SES with multiplicative errors.

Forecast equation
\[ \hat{y}_{t+h|t} = l_t \]

Observation equation
\[ y_t = l_{t-1}(1 + \varepsilon_t) \]

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

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

- Relative forecast errors: \( \varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}} \)

- Models with additive and multiplicative errors with the same parameters generate the same point forecasts but different prediction intervals.
### ETS(M,N,N)

**SES with multiplicative errors.**

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

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

- **Relative forecast errors:**  \( \varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}} \)
- **Models with additive and multiplicative errors** with the same parameters generate the same point forecasts but different prediction intervals.
Holt’s linear method with additive errors.

| Forecast equation | \( \hat{y}_{t+h|t} = \ell_t + h b_t \) |
|-------------------|----------------------------------------|
| 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(A,A,A)

Holt-Winters additive method with additive errors.

Forecast equation
\[ \hat{y}_{t+h|t} = \ell_t + h\beta_t + s_{t-m+h^+_m} \]

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: \[ \varepsilon_t = y_t - \hat{y}_{t|t-1} \]
- \[ h^+_m = \lfloor (h - 1) \mod m \rfloor + 1. \]
### Additive error models

<table>
<thead>
<tr>
<th>Trend</th>
<th>N</th>
<th>Seasonal</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N</strong></td>
<td>$y_t = \ell_{t-1} + \varepsilon_t$</td>
<td>$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$</td>
<td>$y_t = \ell_{t-1} s_{t-m} + \varepsilon_t$</td>
</tr>
<tr>
<td></td>
<td>$\ell_t = \ell_{t-1} + \alpha \varepsilon_t$</td>
<td>$\ell_t = \ell_{t-1} + \alpha \varepsilon_t$</td>
<td>$\ell_t = \ell_{t-1} + \alpha \varepsilon_t / s_{t-m}$</td>
</tr>
<tr>
<td></td>
<td>$s_t = s_{t-m} + \gamma \varepsilon_t$</td>
<td>$s_t = s_{t-m} + \gamma \varepsilon_t / \ell_{t-1}$</td>
<td>$s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + b_{t-1})$</td>
</tr>
<tr>
<td><strong>A</strong></td>
<td>$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$</td>
<td>$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$</td>
<td>$y_t = (\ell_{t-1} + b_{t-1}) s_{t-m} + \varepsilon_t$</td>
</tr>
<tr>
<td></td>
<td>$\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$</td>
<td>$\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$</td>
<td>$\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t / s_{t-m}$</td>
</tr>
<tr>
<td></td>
<td>$b_t = b_{t-1} + \beta \varepsilon_t$</td>
<td>$b_t = b_{t-1} + \beta \varepsilon_t$</td>
<td>$b_t = b_{t-1} + \beta \varepsilon_t / s_{t-m}$</td>
</tr>
<tr>
<td></td>
<td>$s_t = s_{t-m} + \gamma \varepsilon_t$</td>
<td>$s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + b_{t-1})$</td>
<td>$s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + b_{t-1})$</td>
</tr>
<tr>
<td><strong>A_d</strong></td>
<td>$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$</td>
<td>$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$</td>
<td>$y_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m} + \varepsilon_t$</td>
</tr>
<tr>
<td></td>
<td>$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$</td>
<td>$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$</td>
<td>$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t / s_{t-m}$</td>
</tr>
<tr>
<td></td>
<td>$b_t = \phi b_{t-1} + \beta \varepsilon_t$</td>
<td>$b_t = \phi b_{t-1} + \beta \varepsilon_t$</td>
<td>$b_t = \phi b_{t-1} + \beta \varepsilon_t / s_{t-m}$</td>
</tr>
<tr>
<td></td>
<td>$s_t = s_{t-m} + \gamma \varepsilon_t$</td>
<td>$s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + \phi b_{t-1})$</td>
<td>$s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + \phi b_{t-1})$</td>
</tr>
<tr>
<td><strong>M</strong></td>
<td>$y_t = \ell_{t-1} b_{t-1} + \varepsilon_t$</td>
<td>$y_t = \ell_{t-1} b_{t-1} + s_{t-m} + \varepsilon_t$</td>
<td>$y_t = \ell_{t-1} b_{t-1} s_{t-m} + \varepsilon_t$</td>
</tr>
<tr>
<td></td>
<td>$\ell_t = \ell_{t-1} b_{t-1} + \alpha \varepsilon_t$</td>
<td>$\ell_t = \ell_{t-1} b_{t-1} + \alpha \varepsilon_t$</td>
<td>$\ell_t = \ell_{t-1} b_{t-1} + \alpha \varepsilon_t / s_{t-m}$</td>
</tr>
<tr>
<td></td>
<td>$b_t = b_{t-1} + \beta \varepsilon_t / \ell_{t-1}$</td>
<td>$b_t = b_{t-1} + \beta \varepsilon_t / \ell_{t-1}$</td>
<td>$b_t = b_{t-1} + \beta \varepsilon_t / (s_{t-m} \ell_{t-1})$</td>
</tr>
<tr>
<td></td>
<td>$s_t = s_{t-m} + \gamma \varepsilon_t$</td>
<td>$s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} b_{t-1})$</td>
<td>$s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} b_{t-1})$</td>
</tr>
<tr>
<td><strong>M_d</strong></td>
<td>$y_t = \ell_{t-1} b_{t-1}^\phi + \varepsilon_t$</td>
<td>$y_t = \ell_{t-1} b_{t-1}^\phi + s_{t-m} + \varepsilon_t$</td>
<td>$y_t = \ell_{t-1} b_{t-1}^\phi s_{t-m} + \varepsilon_t$</td>
</tr>
<tr>
<td></td>
<td>$\ell_t = \ell_{t-1} b_{t-1}^\phi + \alpha \varepsilon_t$</td>
<td>$\ell_t = \ell_{t-1} b_{t-1}^\phi + \alpha \varepsilon_t$</td>
<td>$\ell_t = \ell_{t-1} b_{t-1}^\phi + \alpha \varepsilon_t / s_{t-m}$</td>
</tr>
<tr>
<td></td>
<td>$b_t = b_{t-1}^\phi + \beta \varepsilon_t / \ell_{t-1}$</td>
<td>$b_t = b_{t-1}^\phi + \beta \varepsilon_t / \ell_{t-1}$</td>
<td>$b_t = b_{t-1}^\phi + \beta \varepsilon_t / (s_{t-m} \ell_{t-1})$</td>
</tr>
<tr>
<td></td>
<td>$s_t = s_{t-m} + \gamma \varepsilon_t$</td>
<td>$s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} b_{t-1}^\phi)$</td>
<td>$s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} b_{t-1}^\phi)$</td>
</tr>
</tbody>
</table>
### Multiplicative error models

<table>
<thead>
<tr>
<th>Trend</th>
<th>N</th>
<th>Seasonal</th>
<th>A</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$y_t = \ell_{t-1}(1 + \varepsilon_t)$</td>
<td>$y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$</td>
<td>$y_t = \ell_{t-1}s_{t-m}(1 + \varepsilon_t)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$</td>
<td>$\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$</td>
<td>$\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$</td>
<td>$s_t = s_{t-m}(1 + \gamma\varepsilon_t)$</td>
</tr>
<tr>
<td>A</td>
<td>$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$</td>
<td>$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$</td>
<td>$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$</td>
<td>$\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$</td>
<td>$\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$</td>
<td>$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$</td>
</tr>
<tr>
<td></td>
<td>$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$</td>
<td>$s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$</td>
<td>$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$</td>
<td>$s_t = s_{t-m}(1 + \gamma\varepsilon_t)$</td>
</tr>
<tr>
<td>Ad</td>
<td>$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$</td>
<td>$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$</td>
<td>$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$</td>
<td>$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$</td>
<td>$\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$</td>
<td>$b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$</td>
</tr>
<tr>
<td></td>
<td>$b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$</td>
<td>$s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$</td>
<td>$b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$</td>
<td>$s_t = s_{t-m}(1 + \gamma\varepsilon_t)$</td>
</tr>
<tr>
<td>M</td>
<td>$y_t = \ell_{t-1}b_{t-1}(1 + \varepsilon_t)$</td>
<td>$y_t = (\ell_{t-1}b_{t-1} + s_{t-m})(1 + \varepsilon_t)$</td>
<td>$y_t = \ell_{t-1}b_{t-1}s_{t-m}(1 + \varepsilon_t)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$</td>
<td>$\ell_t = \ell_{t-1}b_{t-1} + \alpha(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$</td>
<td>$\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$</td>
<td>$b_t = b_{t-1} + \beta(\ell_{t-1}b_{t-1})\varepsilon_t$</td>
</tr>
<tr>
<td></td>
<td>$b_t = b_{t-1} + \beta(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$</td>
<td>$s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$</td>
<td>$b_t = b_{t-1} + \beta(\ell_{t-1}b_{t-1})\varepsilon_t$</td>
<td>$s_t = s_{t-m}(1 + \gamma\varepsilon_t)$</td>
</tr>
<tr>
<td>Md</td>
<td>$y_t = \ell_{t-1}b_{t-1}^\phi(1 + \varepsilon_t)$</td>
<td>$y_t = (\ell_{t-1}b_{t-1}^\phi + s_{t-m})(1 + \varepsilon_t)$</td>
<td>$y_t = \ell_{t-1}b_{t-1}^\phi s_{t-m}(1 + \varepsilon_t)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\ell_t = \ell_{t-1}b_{t-1}^\phi(1 + \alpha\varepsilon_t)$</td>
<td>$\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t$</td>
<td>$\ell_t = \ell_{t-1}b_{t-1}^\phi(1 + \alpha\varepsilon_t)$</td>
<td>$b_t = b_{t-1}^\phi + \beta(\ell_{t-1}b_{t-1}^\phi)\varepsilon_t$</td>
</tr>
<tr>
<td></td>
<td>$b_t = b_{t-1}^\phi + \beta(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t$</td>
<td>$s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t$</td>
<td>$b_t = b_{t-1}^\phi + \beta(\ell_{t-1}b_{t-1}^\phi)\varepsilon_t$</td>
<td>$s_t = s_{t-m}(1 + \gamma\varepsilon_t)$</td>
</tr>
</tbody>
</table>
Innovations state space models

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

$$y_t = h(x_{t-1}) + k(x_{t-1})\varepsilon_t$$

Observation equation

$$x_t = f(x_{t-1}) + g(x_{t-1})\varepsilon_t$$

State equation

Additive errors:

$$k(x_{t-1}) = 1. \quad y_t = \hat{y}_{t|t-1} + \varepsilon_t.$$ 

Multiplicative errors:

$$k(x_{t-1}) = \hat{y}_{t|t-1}. \quad y_t = \hat{y}_{t|t-1}(1 + \varepsilon_t).$$

$$\varepsilon_t = (y_t - \hat{y}_{t|t-1})/\hat{y}_{t|t-1} \text{ is relative error.}$$
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: \( \text{ETS}(M,M,A) \), \( \text{ETS}(M,M_d,A) \), \( \text{ETS}(A,N,M) \), \( \text{ETS}(A,A,M) \), \( \text{ETS}(A,A_d,M) \), \( \text{ETS}(A,M,N) \), \( \text{ETS}(A,M,A) \), \( \text{ETS}(A,M,M) \), \( \text{ETS}(A,M_d,N) \), \( \text{ETS}(A,M_d,A) \), and \( \text{ETS}(A,M_d,M) \).
Some of the combinations of (Error, Trend, Seasonal) can lead to numerical difficulties; see equations with division by a state.

These are: $\text{ETS}(M,M,A)$, $\text{ETS}(M,M_d,A)$, $\text{ETS}(A,N,M)$, $\text{ETS}(A,A,M)$, $\text{ETS}(A,A_d,M)$, $\text{ETS}(A,M,N)$, $\text{ETS}(A,M,A)$, $\text{ETS}(A,M,M)$, $\text{ETS}(A,M_d,N)$, $\text{ETS}(A,M_d,A)$, and $\text{ETS}(A,M_d,M)$. 
# Exponential smoothing models

## Additive Error Seasonal Component

<table>
<thead>
<tr>
<th>Trend Component</th>
<th>Seasonal Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>(None)</td>
<td>N, A, M</td>
</tr>
<tr>
<td>(Additive)</td>
<td>A, A, A</td>
</tr>
<tr>
<td>(Additive damped)</td>
<td>A, A, d, A</td>
</tr>
<tr>
<td>(Multiplicative)</td>
<td>A, M, N</td>
</tr>
<tr>
<td>(Multiplicative damped)</td>
<td>A, M, d, A</td>
</tr>
</tbody>
</table>

## Multiplicative Error Seasonal Component

<table>
<thead>
<tr>
<th>Trend Component</th>
<th>Seasonal Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>(None)</td>
<td>M, N, N</td>
</tr>
<tr>
<td>(Additive)</td>
<td>M, A, M</td>
</tr>
<tr>
<td>(Additive damped)</td>
<td>M, A, d, M</td>
</tr>
<tr>
<td>(Multiplicative)</td>
<td>M, M, N</td>
</tr>
<tr>
<td>(Multiplicative damped)</td>
<td>M, M, d, M</td>
</tr>
</tbody>
</table>
Estimation

\[ L^*(\theta, x_0) = T \log \left( \sum_{t=1}^{T} \frac{\varepsilon_t^2}{k^2(x_{t-1})} \right) + 2 \sum_{t=1}^{T} \log |k(x_{t-1})| \]

\[ = -2 \log(\text{Likelihood}) + \text{constant} \]

Minimize wrt \( \theta = (\alpha, \beta, \gamma, \phi) \) and initial states \( x_0 = (l_0, b_0, s_0, s_{-1}, \ldots, s_{-m+1}) \).

Model selection

Select amongst all models using AIC (or similar).
Estimation

\[ L^*(\theta, x_0) = T \log \left( \sum_{t=1}^{T} \frac{\varepsilon_t^2}{k^2(x_{t-1})} \right) + 2 \sum_{t=1}^{T} \log |k(x_{t-1})| \]

\[ = -2 \log(\text{Likelihood}) + \text{constant} \]

- Minimize wrt \( \theta = (\alpha, \beta, \gamma, \phi) \) and initial states \( x_0 = (\ell_0, b_0, s_0, s_{-1}, \ldots, s_{-m+1}) \).

Model selection

- Select amongst all models using AIC (or similar).
Estimation

\[ L^*(\theta, x_0) = T \log \left( \sum_{t=1}^{T} \frac{\varepsilon_t^2}{k^2(x_{t-1})} \right) + 2 \sum_{t=1}^{T} \log |k(x_{t-1})| \]

\[ = -2 \log(\text{Likelihood}) + \text{constant} \]

- Minimize wrt \( \theta = (\alpha, \beta, \gamma, \phi) \) and initial states \( x_0 = (\ell_0, b_0, s_0, s_{-1}, \ldots, s_{-m+1}) \).

Model selection

- Select amongst all models using AIC (or similar).
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 AIC:

\[
AIC = -2 \log(\text{Likelihood}) + 2p
\]

where \( p = \# \) parameters.
- Produce forecasts using best method.
- Obtain prediction intervals using underlying state space model.

Method performed very well in M3 competition.
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 AIC:

\[ \text{AIC} = -2 \log(\text{Likelihood}) + 2p \]

where \( p = \# \) parameters.

- Produce forecasts using best method.
- Obtain prediction intervals using underlying state space model.

Method performed very well in M3 competition.
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 AIC:

\[
AIC = -2 \log(\text{Likelihood}) + 2p
\]

where \( p = \# \) parameters.
- Produce forecasts using best method.
- Obtain prediction intervals using underlying state space model.

Method performed very well in M3 competition.
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 AIC:

\[
AIC = -2 \log(\text{Likelihood}) + 2p
\]

where \( p = \# \) parameters.
- Produce forecasts using best method.
- Obtain prediction intervals using underlying state space model.

Method performed very well in M3 competition.
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 AIC:

\[
AIC = -2 \log(\text{Likelihood}) + 2p
\]

where \( p \) = \# parameters.
- Produce forecasts using best method.
- Obtain prediction intervals using underlying state space model.

Method performed very well in M3 competition.
Forecasting with ETS models

- Point forecasts obtained by iterating equations for \( t = T + 1, \ldots, T + h \), setting \( \varepsilon_t = 0 \) for \( t > T \).
- Not the same as \( E(y_{t+h}|y_1, \ldots, y_t) \) unless trend and seasonality are both additive.
- Point forecasts for ETS(A,x,y) are identical to ETS(M,x,y) if the parameters are the same.
- Prediction intervals will differ between models with additive and multiplicative methods.
- Exact PI available for many models.
- Otherwise, simulate future sample paths, conditional on last estimate of states, and obtain PI from percentiles of simulated paths.
Point forecasts obtained by iterating equations for \( t = T + 1, \ldots, T + h \), setting \( \varepsilon_t = 0 \) for \( t > T \).

Not the same as \( \mathbb{E}(y_{t+h}|y_1, \ldots, y_t) \) unless trend and seasonality are both additive.

Point forecasts for ETS(A,x,y) are identical to ETS(M,x,y) if the parameters are the same.

Prediction intervals will differ between models with additive and multiplicative methods.

Exact PI available for many models.

Otherwise, simulate future sample paths, conditional on last estimate of states, and obtain PI from percentiles of simulated paths.
Forecasting with ETS models

- Point forecasts obtained by iterating equations for $t = T + 1, \ldots, T + h$, setting $\varepsilon_t = 0$ for $t > T$.
- Not the same as $E(y_{t+h}|y_1, \ldots, y_t)$ unless trend and seasonality are both additive.
- Point forecasts for ETS(A,x,y) are identical to ETS(M,x,y) if the parameters are the same.
- Prediction intervals will differ between models with additive and multiplicative methods.
- Exact PI available for many models.
- Otherwise, simulate future sample paths, conditional on last estimate of states, and obtain PI from percentiles of simulated paths.
Forecasting with ETS models

- Point forecasts obtained by iterating equations for $t = T + 1, \ldots, T + h$, setting $\varepsilon_t = 0$ for $t > T$.
- Not the same as $E(y_{t+h} \mid y_1, \ldots, y_t)$ unless trend and seasonality are both additive.
- Point forecasts for ETS(A,x,y) are identical to ETS(M,x,y) if the parameters are the same.
- Prediction intervals will differ between models with additive and multiplicative methods.
- Exact PI available for many models.
- Otherwise, simulate future sample paths, conditional on last estimate of states, and obtain PI from percentiles of simulated paths.
Forecasting with ETS models

- Point forecasts obtained by iterating equations for \( t = T + 1, \ldots, T + h \), setting \( \varepsilon_t = 0 \) for \( t > T \).
- Not the same as \( \mathbb{E}(y_{t+h} | y_1, \ldots, y_t) \) unless trend and seasonality are both additive.
- Point forecasts for ETS(A,x,y) are identical to ETS(M,x,y) if the parameters are the same.
- Prediction intervals will differ between models with additive and multiplicative methods.
- Exact PI available for many models.
- Otherwise, simulate future sample paths, conditional on last estimate of states, and obtain PI from percentiles of simulated paths.
Point forecasts obtained by iterating equations for $t = T + 1, \ldots, T + h$, setting $\varepsilon_t = 0$ for $t > T$.

Not the same as $E(y_{t+h}|y_1, \ldots, y_t)$ unless trend and seasonality are both additive.

Point forecasts for ETS(A,x,y) are identical to ETS(M,x,y) if the parameters are the same.

Prediction intervals will differ between models with additive and multiplicative methods.

Exact PI available for many models.

Otherwise, simulate future sample paths, conditional on last estimate of states, and obtain PI from percentiles of simulated paths.
Exponential smoothing

library(forecast)
fit <- ets(h02)
fcast <- forecast(fit)
plot(fcast)
Exponential smoothing

```r
> fit
ETS(M,Md,M)

Smoothing parameters:
  alpha = 0.3318
  beta  = 4e-04
  gamma = 1e-04
  phi   = 0.9695

Initial states:
  l = 0.4003
  b = 1.0233
  s = 0.8575 0.8183 0.7559 0.7627 0.6873 1.2884
     1.3456 1.1867 1.1653 1.1033 1.0398 0.9893

sigma: 0.0651

AIC   AICc   BIC
-121.97999 -118.68967 -65.57195
```

Functional time series with applications in demography
2. Automatic time series forecasting
Functional time series

Exponential smoothing

ARIMA modelling

Forecasting functional time series

References
Conventional ARIMA forecasting

- calculate forecasts from the best fitting ARIMA model
- Not necessarily the best forecasting ARIMA model.
- Model identification subjective and complex.
Conventional ARIMA forecasting

- calculate forecasts from the best fitting ARIMA model
- Not necessarily the best forecasting ARIMA model.
- Model identification subjective and complex.
Conventional ARIMA forecasting

- calculate forecasts from the best fitting ARIMA model
- Not necessarily the best forecasting ARIMA model.
- Model identification subjective and complex.
Non-seasonal ARIMA model

\[ y_t \sim ARIMA(p, d, q) \]

\[ y'_t = (1 - B)^d y_t \]

\[ y'_t = c + \sum_{j=1}^{p} \phi_j y'_{t-j} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \]

\( p \) AR parameters: \( \phi = (\phi_1, \ldots, \phi_p) \)

\( q \) MA parameters: \( \theta = (\theta_1, \ldots, \theta_q) \)

\( d \) is the differencing order

\[ \phi(B)(1 - B)^d y_t = c + \theta(B) \varepsilon_t \]

Need to select appropriate orders \( p, q, d \), and whether to include \( c \).
How does `auto.arima()` work?

A non-seasonal ARIMA process

\[ \phi(B)(1 - B)^d y_t = c + \theta(B) \varepsilon_t \]

Need to select appropriate orders \( p, q, d \), and whether to include \( c \).
How does auto.arima() work?

A non-seasonal ARIMA process

$$\phi(B)(1 - B)^d y_t = c + \theta(B)\varepsilon_t$$

Need to select appropriate orders $p, q, d$, and whether to include $c$.

Hyndman & Khandakar (JSS, 2008) algorithm:

- Select no. differences $d$ via KPSS unit root test.
- Select $p, q, c$ by minimising AIC.
- Use stepwise search to traverse model space, starting with a simple model and considering nearby variants.
How does auto.arima() work?

$$\text{AICc} = -2 \log(L) + 2(p + q + k + 1) \left( \frac{T}{T-p-q-k-2} \right).$$

where $L$ is the maximised likelihood fitted to the differenced data, $k = 1$ if $c \neq 0$ and $k = 0$ otherwise.

**Step 1:** Select current model (with smallest AIC) from:
- ARIMA(2, d, 2)
- ARIMA(0, d, 0)
- ARIMA(1, d, 0)
- ARIMA(0, d, 1)

**Step 2:** Consider variations of current model:
- vary one of $p$, $q$, from current model by $\pm 1$
- $p$, $q$ both vary from current model by $\pm 1$
- Include/exclude $c$ from current model

Model with lowest AIC becomes current model. Repeat Step 2 until no lower AIC can be found.
How does auto.arima() work?

\[
AICc = -2 \log(L) + 2(p + q + k + 1) \left( \frac{T}{T - p - q - k - 2} \right).
\]

where \( L \) is the maximised likelihood fitted to the *differenced* data, \( k = 1 \) if \( c \neq 0 \) and \( k = 0 \) otherwise.

**Step 1:** Select current model (with smallest AIC) from:
- ARIMA(2, d, 2)
- ARIMA(0, d, 0)
- ARIMA(1, d, 0)
- ARIMA(0, d, 1)

**Step 2:** Consider variations of current model:
- vary one of \( p, q, \) from current model by \( \pm 1 \)
- \( p, q \) both vary from current model by \( \pm 1 \)
- Include/exclude \( c \) from current model

Model with lowest AIC becomes current model.

Repeat Step 2 until no lower AIC can be found.
How does auto.arima() work?

\[
\text{AICc} = -2 \log(L) + 2(p + q + k + 1) \left( \frac{T}{T - p - q - k - 2} \right)
\]

where \( L \) is the maximised likelihood fitted to the differenced data, \( k = 1 \) if \( c \neq 0 \) and \( k = 0 \) otherwise.

**Step 1:** Select current model (with smallest AIC) from:
- ARIMA(2, d, 2)
- ARIMA(0, d, 0)
- ARIMA(1, d, 0)
- ARIMA(0, d, 1)

**Step 2:** Consider variations of current model:
- vary one of \( p, q \), from current model by \( \pm 1 \)
- \( p, q \) both vary from current model by \( \pm 1 \)
- Include/exclude \( c \) from current model

Model with lowest AIC becomes current model.

Repeat Step 2 until no lower AIC can be found.
How does auto.arima() work?

\[
\text{AICc} = -2 \log(L) + 2(p + q + k + 1) \left( \frac{T}{T-p-q-k-2} \right).
\]

where \( L \) is the maximised likelihood fitted to the differenced data, \( k = 1 \) if \( c \neq 0 \) and \( k = 0 \) otherwise.

**Step 1:** Select current model (with smallest AIC) from:
- ARIMA(2, d, 2)
- ARIMA(0, d, 0)
- ARIMA(1, d, 0)
- ARIMA(0, d, 1)

**Step 2:** Consider variations of current model:
- vary one of \( p, q \), from current model by \( \pm 1 \)
- \( p, q \) both vary from current model by \( \pm 1 \)
- Include/exclude \( c \) from current model

Model with lowest AIC becomes current model.

Repeat Step 2 until no lower AIC can be found.
How does auto.arima() work?

A seasonal ARIMA process

\[
\Phi(B^m) \phi(B) (1 - B)^d (1 - B^m)^D y_t = c + \Theta(B^m) \theta(B) \varepsilon_t
\]

Need to select appropriate orders \(p, q, d, P, Q, D\), and whether to include \(c\).

Hyndman & Khandakar (JSS, 2008) algorithm:

- Select no. differences \(d\) via KPSS unit root test.
- Select \(D\) using OCSB unit root test.
- Select \(p, q, P, Q, c\) by minimising AIC.
- Use stepwise search to traverse model space, starting with a simple model and considering nearby variants.
Automatic seasonal ARIMA

Functional time series with applications in demography

2. Automatic time series forecasting

<table>
<thead>
<tr>
<th>Year</th>
<th>Total scripts (millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>0.4</td>
</tr>
<tr>
<td>2000</td>
<td>0.6</td>
</tr>
<tr>
<td>2005</td>
<td>0.8</td>
</tr>
<tr>
<td>2010</td>
<td>1.0</td>
</tr>
</tbody>
</table>
**Auto ARIMA**

```r
fit <- auto.arima(h02)
fcast <- forecast(fit)
plot(fcast)
```

**Automatic ARIMA forecasts**

<table>
<thead>
<tr>
<th>Year</th>
<th>Total scripts (millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>0.4</td>
</tr>
<tr>
<td>2000</td>
<td>0.6</td>
</tr>
<tr>
<td>2005</td>
<td>0.8</td>
</tr>
<tr>
<td>2010</td>
<td>1.0</td>
</tr>
</tbody>
</table>

**Functional time series with applications in demography**

2. Automatic time series forecasting
Auto ARIMA

```
> fit
Series: h02
ARIMA(3,1,3)(0,1,1)[12]

Coefficients:
   ar1  ar2  ar3   ma1   ma2   ma3
   -0.3648 -0.0636  0.3568  -0.4850  0.0479  -0.353
   s.e.  0.2198  0.3293  0.1268  0.2227  0.2755  0.212
   sma1
   -0.5931  s.e.  0.0651

sigma^2 estimated as 0.002706:  log likelihood=290.25
AIC=-564.5    AICc=-563.71    BIC=-538.48
```
1. Functional time series
2. Exponential smoothing
3. ARIMA modelling
4. Forecasting functional time series
5. References
Functional principal components

Main effects

Interaction

French male mortality

ETS(A,A,N)
ETS(A,N,N)
ETS(A,N,N)
French male mortality

ARIMA(1,1,1)+c
ARIMA(1,1,1)
ARIMA(0,1,1)
Functional principal components

Main effects

Interaction

Australian fertility

Functional time series with applications in demography 2. Automatic time series forecasting
Functional principal components

Australian fertility

ETS(A,Ad,N)  ETS(A,Ad,N)  ETS(A,Ad,N)

Functional time series with applications in demography  2. Automatic time series forecasting
Functional principal components

Main effects

Interaction

Australian fertility

ARIMA(1,1,1)+c
ARIMA(5,2,2)
ARIMA(2,1,0)

Functional time series with applications in demography

2. Automatic time series forecasting
Outline

1 Functional time series
2 Exponential smoothing
3 ARIMA modelling
4 Forecasting functional time series
5 References

www.exponentialsmoothing.net


Hyndman (2014). *forecast: Forecasting functions for time series and linear models*.  
cran.r-project.org/package=forecast