ETS Models

Handout

ETS: ExponenTial Smoothing

- Because we try to take more of the recent values into account for the current estimate. This then “smooths” out the future values.

ETS: Error, Trend, Seasonality

- More appropriate way of representing

Notations:

- ℓ_t - Level (or the smoothed value) of the series at time t
  - α - Corresponding weight parameter for level component. $(0 \leq \alpha \leq 1)$
- b_t - Trend of the series at time t
  - β^* - Corresponding weight parameter for trend component $(0 \leq \beta^* \leq 1)$
- s_t - Season of the series at time t
  - γ - Corresponding weight parameter for seasonal component $(0 \leq \gamma \leq 1 - \alpha)$
    - $\gamma = 0$ - No change in seasonal component
    - $\gamma = 1$ - Seasonal Naive
    - Usually $\gamma \leq 0.2$
- ϕ - Damp factor $(0 < \phi < 1)$
- $\hat{y}_{t+1|t} = \alpha y_t + (1 - \alpha)\hat{y}_{t|t-1}$ - One step ahead forecasts (which are essentially the fitted values).
- $\hat{y}_{T+h|T} = \ell_T, h = 2, 3, \cdots$ - h step ahead forecasts

Method vs Models

▼ **Methods**

We are able to generate forecasts without any further assumptions, however we will not be able to generate error terms or prediction intervals

- Algorithms that return point forecasts

▼ **Models**

For models, there is the idea of an error term

With an error term, we can then determine prediction intervals

- Generate same point forecasts as methods but can also generate forecast distributions
- A stochastic (or random) data generating process that can generate an entire forecast distribution
- Allows for “proper” model selection

State Space Models

We will try to get our models to be in this state space model equation so that it will have nice properties for us to work with

Observation/Measurement Equation

$$y_t = w' x_{t-1} + e_t$$

w' - Vector

e_t - Innovation Residuals

- e_t is a WN process

Transition Equation

Note that we will not observe the x_t but it will help to drive what the next y_t will be

$$x_t = F x_{t-1} + g e_t$$

F - Matrix

g - Vector

e_t - Innovation Residuals

- e_t is a WN process

General Form

- This could be the form if we have equations that a non linear

$$y_t = w(x_{t-1}) + r(x_{t-1}) e_t$$
$$x_t = f(x_{t-1}) + g(x_{t-1}) e_t$$

Note that these becomes functions now

$w(.), r(.)$ - Takes in a vector and returns a scalar

$f(.), g(.)$ - Takes in a vector and returns a vector

▼ Exponential Smoothing Methods

Note that there will be 2 components: **Forecast Equation & Smoothing Equation**

Forecast Equation: This tells us how we generate our forecasts

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

Smoothing Equation:

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

For our smoothing equations, we try to get the smoothed values of the components by taking a weighted average between the current estimate of that component and the previous estimate of the component.

Note: For methods, we do not look at errors. Therefore, we are only looking at the type of trend and seasonality.

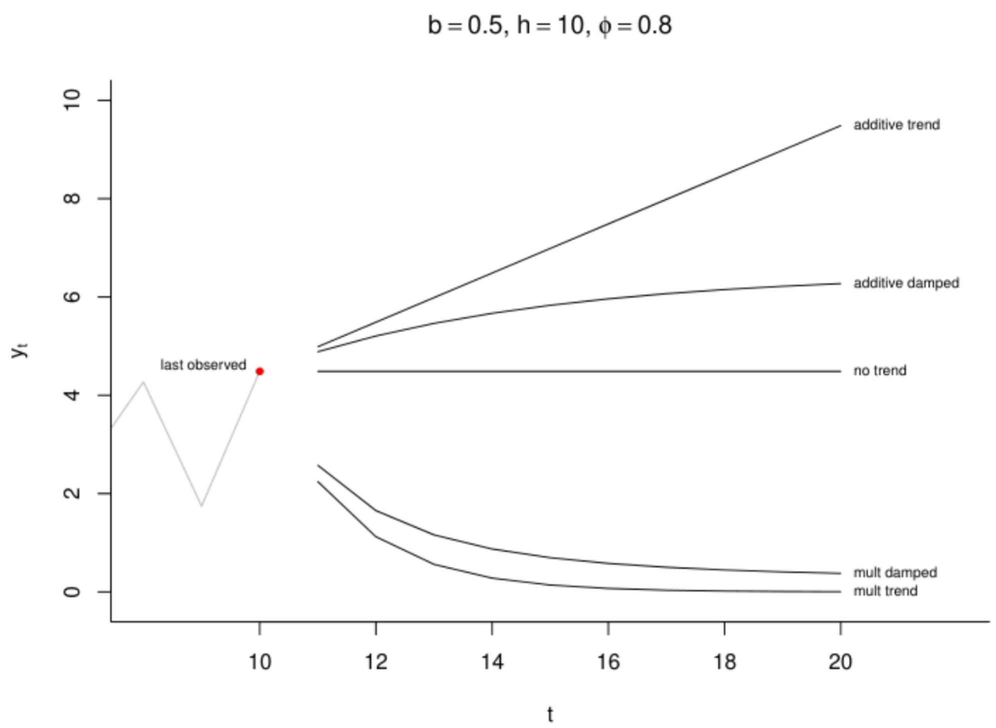
▼ Trend Types

Note that with this, we are allowing for the level to change over time

Id	Trend type	Formula
1	None	$T_h = l$
2	Additive	$T_h = l + bh$
3	Additive damped	$T_h = l + (\phi + \phi^2 + \cdots + \phi^h)b$
4	Multiplicative	$T_h = lb^h$
5	Multiplicative damped	$T_h = lb^{(\phi+\phi^2+\cdots+\phi^h)}$

- T_h - Forecast of trend h time periods ahead
- $0 < \phi < 1$ - Damping parameter
- l - Level that we are currently at
- b - Growth term

▼ Visualisation



▼ Seasonality Types

Note that we have 3 different types of seasonality

1. No Seasonality
2. Additive Seasonality
3. Multiplicative Seasonality

All combinations of Trend and Seasonality

Trend	Seasonal None	Seasonal Add	Seasonal Mult
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
M (Mult.)	M,N	M,A	M,M
M_d (Mult. damped)	M_d,N	M_d,A	M_d,M

- (N, N) - Simple Exponential Smoothing
- (A, N) - Holt's Linear Method
- (M, N) - Exponential Trend Method
- (A_d, N) - Additive damped trend Method

- (M_d, N) - Multiplicative damped trend Method
- (A, A) - Additive Holt-Winters Method
- (A, M) - Multiplicative Holt-Winters Method

Trend	Seasonal N	Seasonal A	Seasonal M
N	$\hat{y}_{t+h\|t} = \ell_t$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t/\ell_{t-1}) + (1-\gamma)s_{t-m}$
A	$\hat{y}_{t+h\|t} = \ell_t + hb_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + hb_t)s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + b_{t-1})) + (1-\gamma)s_{t-m}$
$\mathbf{A_d}$	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + \phi_h b_t)s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + \phi b_{t-1})) + (1-\gamma)s_{t-m}$

Formulas for recursive calculations and point forecasts. In each case,
ℓ_t - series level at time t
b_t - slope at time t
s_t - seasonal component of the series at time t
m - number of seasons in a year
$\alpha, \beta^*, \gamma, \phi$ - smoothing parameters
$\phi_h = \phi + \phi^2 + \cdots + \phi^h$
k - integer part of $(h-1)/m$

▼ Simple Exponential Smoothing (N, N) Method

- No Trend and No Seasonality
- Note that this is just a method since we are looking at the forecast without any error terms

▼ Definition 1

We just add a dampened version of the forecast error from the previous forecast

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

- $0 < \alpha < 1$ - Dampening factor of the forecast error
- $\hat{y}_t$ - Most recent forecast
- $y_t - \hat{y}_t$ - Most recent forecast error

▼ Definition 2

Weighted average of the most recent observation and the most recent forecast. This results in a weighted average of all past observations. **Noting** that we need to initialize the value of $\hat{y}_1$ over here.

We can see that the more recent the observation, the more we take that into account.

$$\begin{aligned}\hat{y}_{t+1} &= \alpha y_t + (1-\alpha)\hat{y}_t \\ &= \sum_{k=0}^{t-1} \alpha(1-\alpha)^k y_t + (1-\alpha)^t \hat{y}_1\end{aligned}$$

Note that the bottom formula is derived through substituting $\hat{y}_t$ into the formula and doing it recursively until we get $\hat{y}_1$

▼ **Weighted Average Formula**

$$\frac{\sum w_i y_i}{\sum w_i}$$

▼ **Definition 3**

There are 2 things that we are concerned with

1. Evolution of the components
2. Forecast Equation

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

Note that for the SES method, the one-step ahead forecast is just the current level since there is no trend or seasonality.

▼ Holt's Linear Method (A, N) Method

- **Applicable**: When we observe a trend in the data
- **Assumption**: At each point, there is a linear trend b_t and level l_t from which the trend starts

Updating Equations

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

- $\ell_t - \ell_{t-1}$ - Estimate of most recent growth since it is the difference between the current level and previous level

- Note that when $\beta^* = 0$ we can see that the growth does not change over time. Since this makes use of the most recent growth only
- l_t - Level at time t . 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 - Trend at time t. Weighted average of $(\ell_t - \ell_{t-1})$ and b_{t-1} , current and previous estimate of the slope.
- $\hat{y}_{t+h|t}$ - h step ahead forecast of y. This is given by the level at time t, l_t and the increase in the trend term at time t, b_t

Forecast Function

- Note that the forecast equation is no longer just the level, l_t since there is a trend component and it becomes a linear function of h instead.

▼ **Parameters we need to choose**

$\alpha, \beta^*, \ell_0, b_0$

▼ Additive Damped Model (A_d, N) Method

Note that if $\phi = 1$ then it is identical to Holt's linear method

Updating Equations

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

- l_t - Level at time t is a weighted average of y_t and the most recent forecast of $\hat{y}_t$
- b_t - Trend at time t is a weighted average of the damped trend from b_{t-1} and the estimated trend at time $t-1$ given by $l_t - l_{t-1}$
- $\hat{y}_{t+h|t}$ - The h step ahead forecast is given by the estimate of the level at time t and adding up the dampened effect of all the b_t

Forecast Function:

- Note that since the trend component is damped, the forecast equation will converge to the following (this is just taking the infinity series of a geometric progression)

  $(\phi + \phi^2 + \cdots) = \frac{\phi}{1-\phi}$

$$\lim_{h\to\infty} \hat{y}_{t+h|t} = l_t + \frac{\phi}{1-\phi} b_t$$

- **Short Term** → Forecasts have a trend
- **Long Term** → The forecasts become constant (because the dampening will make it such that the trend no longer adds much at the end)

▼ Holt-Winters Seasonal Method $(A, A/M)$

This allows for introduction of the seasonality component

▼ Additive Seasonality (A, A)

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

where $h_m^+ = \lfloor (h-1)\%m \rfloor + 1$

- For ℓ_t - We note that we first seasonally adjust for the latest observation and add it with the latest one step ahead forecast $(\ell_{t-1} + b_{t-1})$. Note that we will take a weighted average.
- For b_t - We first look at the latest trend $(\ell_t - \ell_{t-1})$, by taking the current level removing the previous level, we can get the trend at the current time point since it is the trend that brings it from the previous level to the current level. We then add it with the previous trend. Note that we will take a weighted average.
- For s_t - We first try to get the seasonal component for the current time point, t by taking $(y_t - \ell_{t-1} - b_{t-1})$ and taking a weighted average with the previous seasonal component for the **same season**.
- **Note**: s_t in absolute terms - within each year $\sum_i s_i \approx 0$
  - Seasonal component will fluctuate around 0

▼ Other formulation of s_t

$$s_t = \gamma^*(y_t - \ell_t) + (1-\gamma^*)s_{t-m}$$

- Note that we can formulate s_t as the above where the only difference is ℓ_t where we are using the current level estimate. However, we want to write it in terms of past data instead so that it is easier to write in our state space equation

We can substitute ℓ_t using the **smoothing equations**

$$
\begin{aligned}
s_t &= \gamma^*(y_t - \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})) + (1-\gamma^*)(s_{t-m}) \\
&= \gamma^*(1-\alpha)y_t + \gamma^*(1-\alpha)(\ell_{t-1} + b_{t-1}) + \gamma^*\alpha(s_{t-m}) + (1-\gamma^*)(s_{t-m}) \\
&= \gamma^*(1-\alpha)(y_t - \ell_{t-1} - b_{t-1}) + [1 - \gamma^*(1-\alpha)]s_{t-m}
\end{aligned}
$$

- We can then set $\gamma = (1-\alpha)\gamma^*$ to get the current smoothing equation that we have
- Note that since $0 \leq \gamma^* \leq 1,$ this translates to $0 \leq \gamma \leq (1-\alpha)$

▼ Multiplicative Seasonality (A, M)

Updating Equations

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

where $h_m^+ = \lfloor (h-1)\% m \rfloor + 1$

- Now instead of subtracting, we divide through the seasonal components
- s_t in relative terms - within each year $\sum_i s_i \approx m$
    - Seasonal component will fluctuate around 1

▼ Code Output

Estimated coefficients

```
tidy(fit) |>
  spread(.model, estimate)
```

```
## # A tibble: 9 x 3
##   term      additive multiplicative
##   <chr>        <dbl>          <dbl>
## 1 alpha     0.236          0.186
## 2 b[0]    -37.4          -33.4
## 3 beta      0.0298         0.0248
## 4 gamma     0.000100       0.000100
## 5 l[0]   9899.          9853.
## 6 s[-1]  -684.   Q3        0.926
## 7 s[-2]  -290.   Q2        0.970
## 8 s[-3]  1512.   Q1        1.16
## 9 s[0]   -538.   Q4        0.943
```

Do note which season corresponds to which s[i]. s[0] corresponds to the last data that we have observed

Holt-Winters Damped Method (A_d, M)

- One of the most accurate forecasting method for seasonal data. From empirical data, it seems like this works the best with seasonal data

$$
\begin{aligned}
l_t &= \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(l_{t-1} + \phi b_{t-1}) \\
b_t &= \beta^*(l_t - l_{t-1}) + (1-\beta^*)\phi b_{t-1} \\
s_t &= \gamma \frac{y_t}{l_{t-1} + \phi b_{t-1}} + (1-\gamma)s_{t-m} \\
\hat{y}_{t+h|t} &= (l_t + (\phi + \phi^2 + \cdots + \phi^h)b_t)s_{t-m+h_m^+}
\end{aligned}
$$

where $h_m^+ = \lfloor (h-1)\%m \rfloor + 1$

For this we have a multiplicative seasonal component and a damped trend

Optimizing Smoothing Parameters

- To choose the best parameters (for e.g. α and ℓ_0 for Simple Exponential Smoothing), we will choose it by minimising the **SSE**:

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

- **Note**: There is no closed form solution for this and therefore, we need to make use of the numerical optimisations.

State Space Models

We want distribution and error terms for us to get prediction intervals

We will add an error term in the methods that we have derived above to get a model

▼ **Notation**

ETS(-, -,-)

- **First term** - Error Term
- **Second Term** - Trend Term
- **Third Term** - Seasonality Term

▼ **How we want to represent our model**

$$\begin{aligned} \text{Measurement Equation:} \quad & y_t = w^T x_{t-1} + e_t \\ \text{State Equation:} \quad & x_t = F x_{t-1} + g e_t \end{aligned}$$

where $e_t = \epsilon_t \sim \mathrm{NID}(0, \sigma^2)$

Measurement Equation: Relationship between observations and states

State Equation(s): Evolution of the state(s) through time

- Note that we do not observe the states but only the the actual observations which is represented through the measurement equation.

Summary of State Space Model Representations

ADDITIVE ERROR MODELS

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

Things to Note

- Models with the same parameters generate the same point forecasts (i.e. If the **Trend** and **Seasonality** components are the same). However, their prediction intervals will be different if they have additive errors & multiplicative errors respectively.

▼ ETS(A, N, N)

SES with **additive errors**

Component Form

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

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

- This is essential the difference between what we see (y_t) and what we **predicted** we will see $(\hat{y}_{t|t-1})$
- We will introduce an error term to allow us to generate a model

<u>Error Correction Form</u>

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

- To get the y_t equation we just rearrange the forecast error equation
- To get the ℓ_t equation, we just rearrange the smoothing equation to group the y_t and ℓ_{t-1} together instead and substitute y_t in here to introduce the error term

We specify a probability distribution for e_t. Assume that

$$e_t = \epsilon_t \sim \mathrm{NID}(0, \sigma^2)$$

- Note that by specifying the distribution of the error terms, our point forecasts have stochastic properties now.

▼ ETS(A, A, N)

We added in the error term for the Holt's Linear Method and gets us a Model instead

Assume the following (Note that we are taking the difference between what we saw (y_t) and what we predicted we will see $(\ell_{t-1} - b_{t-1})$

$$\epsilon_t = y_t - \ell_{t-1} - b_{t-1} \sim \mathrm{NID}(0, \sigma^2)$$

▼ Error Correction Equations

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

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

▼ Derivation of the equations

We first note this is the smoothing and forecast equations of the (A, N) method

$$
\begin{aligned}
\text{Forecast Equation} \qquad & \hat{y}_{t+1|t} = \ell_t + b_t \\
\text{Component Updating} \qquad & \ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
& b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}
\end{aligned}
$$

Note that for our state space model: $x_t = \begin{bmatrix} \ell_t \\ b_t \end{bmatrix}$

Rearranging from the prediction error, we can get the **measurement equation**

$$
\begin{aligned}
y_t &= \ell_{t-1} + b_{t-1} + \epsilon_t \\
y_t &= \begin{bmatrix} 1 & 1 \end{bmatrix} x_t + \epsilon_t
\end{aligned}
$$

Getting the State/Transition Equations

For the **level**: We make use of the **component updating** equation and making use of the **forecast equation**

$$
\begin{aligned}
\ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
&= \ell_{t-1} + b_{t-1} + \alpha(y_t - \ell_{t-1} - b_{t-1}) \\
&= \ell_{t-1} + b_{t-1} + \alpha(y_t - \hat{y}_{t|t-1}) \\
&= \ell_{t-1} + b_{t-1} + \alpha\epsilon_t \\
&= \begin{bmatrix} 1 & 1 \end{bmatrix} x_{t-1} + \alpha\epsilon_t
\end{aligned}
$$

For the **level**: We make use of the **component updating** equation and making use of the state equation of ℓ_t

$$
\begin{aligned}
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
&= b_{t-1} + \beta^*(\ell_t - \ell_{t-1} - b_{t-1}) \\
&= b_{t-1} + \beta^*(\ell_{t-1} + b_{t-1} + \alpha e_t - \ell_{t-1} - b_{t-1}) \\
&= b_{t-1} + \beta^*\alpha\epsilon_t \\
&= \begin{bmatrix} 0 & 1 \end{bmatrix} x_{t-1} + \beta\epsilon_t
\end{aligned}
$$

State Space Model

$$
\begin{aligned}
\text{Measurement Equation} \qquad & y_t = \begin{bmatrix} 1 & 1 \end{bmatrix} x_t + \epsilon_t \\
\text{State/Transition Equation} \qquad & x_t = \begin{bmatrix} \ell_t \\ b_t \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x_{t-1} + \begin{bmatrix} \alpha \\ \beta \end{bmatrix} e_t
\end{aligned}
$$

Forecast Mean: We can see that the forecast mean will be the same as the point forecast of the Holt-Winters Method

$$
\begin{aligned}
\mu_{t+h|t} &= E(y_{t+h}|x_t) \\
\mu_{t|t-1} &= E(y_t|x_{t-1}) \\
&= E(\begin{bmatrix}1 & 1\end{bmatrix} x_{t-1}) \\
&= \ell_{t-1} + b_{t-1} \\
&= \text{Forecast of Holt-Winters Method}
\end{aligned}
$$

▼ **Forecast Equation**

$$\hat{y}_{t+1} = \ell_t + b_t$$

▼ ETS(M, A, N)

- Note that here we are multiplying the error term instead
- With multiplicative error terms, we will have a non-linear state space

We assume the following for the forecast error

$$\epsilon_t = \frac{y_t - (\ell_{t-1} + b_{t-1})}{(\ell_{t-1} + b_{t-1})} \sim \text{NID}(0, \sigma^2)$$

▼ **Error Correction Equations**

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

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

▼ ETS(A, A, A)

Holt-Winters additive method with additive errors

$$
\begin{aligned}
\text{Forecast Equation} \quad & \hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)} \\
\text{Observation Equation} \quad & y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \epsilon_t \\
\text{State Equation} \quad & \ell_t = \ell_{t-1} + b_{t-1} + \alpha\epsilon_t \\
& b_t = b_{t-1} + \beta\epsilon_t \\
& s_t = s_{t-m} + \gamma\epsilon_t
\end{aligned}
$$

Forecast Errors: $\epsilon_t = y_t - \hat{y}_{t|t-1}$

k - Integer part of $(h-1)/m$

▼ ETS(M, A, M)

Holt-Winters multiplicate method with multiplicative errors

$$\begin{aligned}
\text{Forecast Equation} \quad \hat{y}_{t+h|t} &= (\ell_t + hb_t)s_{t+h-m(k+1)} \\
\text{Observation Equation} \quad y_t &= (\ell_{t-1} + b_{t-1})s_{t-m}(1+\epsilon_t) \\
\text{State Equation} \quad \ell_t &= (\ell_{t-1} + b_{t-1})(1+\alpha\epsilon_t) \\
b_t &= b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\epsilon_t \\
s_t &= s_{t-m}(1+\gamma\epsilon_t)
\end{aligned}$$

Forecast Errors: $\epsilon_t = (y_t - \hat{y}_{t|t-1})/\hat{y}_{t|t-1}$

k - Integer part of $(h-1)/m$

▼ Residuals and Innovations

▼ Residuals

$$y_t - \hat{y}_t$$

▼ **Innovations**: When we make assumptions about the error terms, we are always talking about the innovations

For multiplicative models: This is a relative error. Note that there could be numerical issues with this because of the denominator

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

For additive models: Residuals is always the same as innovations which is just the difference between the observed and predicted

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

▼ Linear Innovations State Space Model

$$\begin{aligned}
p(\mathbf{y}|x_0) &= \prod_{t=1}^{n} p(y_t|y_1, y_2, \ldots, y_{t-1}, x_0) \\
&= \prod_{t=1}^{n} p(y_t|x_{t-1}) \\
&= \prod_{t=1}^{n} p(e_t)
\end{aligned}$$

Note that for the p on the second line it is the pdf on y

Note that for the p on the last line it is the pdf on e

$$p(y|x_0) = (\frac{1}{\sqrt{2\pi\sigma^2}})^n(-\frac{1}{2\sigma^2}\sum_{t=1}^{n} e_t^2)$$

▼ ETS(A, A, A)

- Note that there is an error in the F matrix in the notes

▼ Non Linear Innovations State Space Model

▼ Estimation in State Space Model

▼ Maximum Likelihood Estimation

▼ Prediction Intervals

Can only be generated using the models

1. Prediction intervals differ between models with additive and multiplicative errors
2. Exact formulae for some models
3. 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. (This is only if we do not have closed form)

$$\text{Prediction Interval: } \hat{y}_{T+h|T} \pm c\sigma_h$$

c - Depends on the coverage probability and σ_h - Forecast standard deviation

▼ Point Forecasts

Traditional Point Forecasts: Iterate the equations for $t = T+1, T+2, \cdots, T+h$ and set all $\epsilon_t = 0$ for $t > T$

1. Point Forecasts for ETS(A, *, *) are identical to ETS(M, *, *)
    a. This means that if the seasonality (**S**) and trend (**T**) are the same, they will generate the same point forecast since error terms do not really matter for point forecasts
2. Not the same as $E(y_{t+h}|x_t)$ unless seasonality (**S**) is additive

▼ **Example: Difference between point forecasts and observed values**

Model: ETS(A, A, N)

The main difference arises with the inclusion of the error terms

- $t = T + 1$
    - **Observed Value**: $y_{T+1} = \ell_T + b_T + \epsilon_{T+1}$
    - **Point Forecast**: $\hat{y}_{T+1|T} = \ell_T + b_T$
- $t = T + 2$
    - **Observed Value**: $y_{T+2} = \ell_{T+1} + b_{T+1} + \epsilon_{T+2} = (\ell_t + b_T + \alpha\epsilon_{T+1}) + (b_T + \beta\epsilon_{T+1}) + \epsilon_{T+2}$
        - We sub in the state equations here
    - **Point Forecast**: $\hat{y}_{T+2|T} = \ell_T + 2b_T$

▼ Prediction Distribution

Distribution of future values, given: the model, its estimated parameters and x_t

$$y_{t+h|t} \equiv y_{t+h|x_t}$$

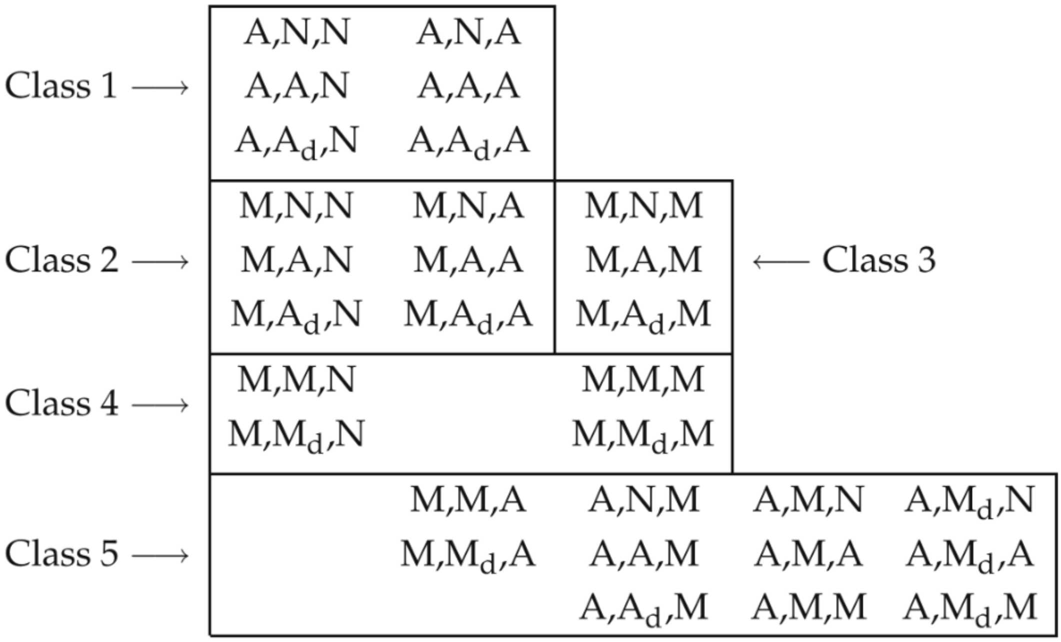
Forecast Mean

$$\mu_{t+h|t} = E(y_{t+h}|x_t)$$

Forecast Variance

$$v_{t+h|t} = Var(y_{t+h|x_t})$$

Classes of Models



Classes of Models

Class 1 - Analytical Expressions are available and easy to derive.

- These are models without any **multiplicative** terms for all the components

Class 2, 3 - Analytical Expressions are available but involve making a few further assumptions

- Class 2 are models with at most **multiplicative error** terms
- Class 3 are models with **multiplicative error and seasonal** terms

Class 4, 5 - Analytical expressions are not available. Use simulation to obtain prediction intervals for those models

▼ Forecast Variances:

Table 8.8: Forecast variance expressions for each additive state space model, where σ^2 is the residual variance, m is the seasonal period, and k is the integer part of $(h-1)/m$ (i.e., the number of complete years in the forecast period prior to time $T+h$).

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

- Note that we have closed form solutions for the forecast variances so long as there are **NO multiplicative Errors, Trend, Seasonality**.

▼ General Derivation for Class 1 Models

Note that for Class 1 Models, they can all be written in the following manner

$$
\begin{aligned}
y_t &= w' x_{t-1} + e_t \\
x_t &= F x_{t-1} + g e_t
\end{aligned}
$$

w - vector, g - vector, F - matrix

- Note that x_t is just a vector of all the state variables

General Result

$$
\begin{aligned}
\hat{y}_{t+h|t} &= w' F^{h-1} x_t \\
v_{t+h|t} &= \begin{cases} \sigma^2, & h=1 \\ \sigma^2[1+\sum_{j=1}^{h-1} c_j^2], & h \geq 2 \end{cases}
\end{aligned}
$$

where $c_j = w' F^{j-1} g$

▼ Example: ETS(A, N, A)

$$x_t = \begin{bmatrix} l_t \\ s_t \\ s_{t-1} \\ \vdots \\ s_{t-(m-1)} \end{bmatrix}, \qquad w = \begin{bmatrix} 1 \\ 0_{m-1} \\ 1 \end{bmatrix}$$

$$F = \begin{bmatrix} 1 & 0'_{m-1} & 0 \\ 0 & 0'_{m-1} & 1 \\ 0_{m-1} & I_{m-1} & 0_{m-1} \end{bmatrix}, \qquad g = \begin{bmatrix} \alpha \\ \beta \\ 0_{m-1} \end{bmatrix}$$

where 0_{m-1} is a vector of $m-1$ zeros

Forecast Equation

From the measurement equation

$$\begin{aligned} \hat{y}_{t+h|t} &= E(y_{t+h}|x_t) \\ &= E(w'x_{t+h-1} + e_t|x_t) \\ &= w'E(x_{t+h-1}|x_t) \end{aligned}$$

We can make use of the state equation and substitute continuously

$$\begin{aligned} E(x_{t+h-1}|x_t) &= E(Fx_{t+h-2} + ge_{t+h-1}|x_t) \\ &= FE(x_{t+h-2}|x_t) \\ &= \cdots \\ &= F^{h-1}x_t \end{aligned}$$

Substituting back to the measurement equation

$$\hat{y}_{t+h|t} = w'F^{h-1}x_t$$

Forecast Variance

$$v_{t+h|t} = w'Var(x_{t+h-1}|x_t)w + \sigma^2$$

▼ Intervals by Simulation

Mainly used for models in **Class 4 and 5** to obtain prediction intervals

Simulating sample paths from the models, conditional on the most recent state x_t

▼ **Steps**

Suppose the required forecast horizon is h. Then for $i = 1, \cdots, M$

1. Generate obseravtions $y^i_{T+1}, y^i_{T+2}, \cdots, y^i_{T+h}$, starting with x_T from the fitted model
2. Each e_{T+k} value is obtained from a random number generator assuming a Gaussian or other appropriate distribution.
3. After generating all the observations for M (where $M = 5000$ usually), then we can take the mean of simulated values at each h as the point forecast
    a. E.g. for $h = 1$, we take the mean for the following M values: $\{y^1_{T+1}, y^2_{T+1}, \cdots, y^M_{T+1}\}$
4. We can make of the quantiles of those values to get the prediction intervals. Using the same M values, we can find the 0.25 and 0.975 quantiles to generate a 95% confidence interval