

ETC3550/ETC5550

Applied forecasting

Week 7: Seasonal exponential
smoothing



ETS models

General notation E T S : ExponenTial S moothing



Error T r e n d S e a s o n

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

ETS models

General notation E T S : ExponenTial S moothing



The diagram shows three arrows pointing upwards from the words 'Error', 'Trend', and 'Season' to the letters 'E', 'T', and 'S' respectively in the 'ETS' part of the notation above.

Error T r e n d S e a s o n

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

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

ETS models

General notation E T S : ExponenTial Smoothing



The diagram consists of three arrows pointing upwards from the words 'Error', 'Trend', and 'Season' to the letters 'E', 'T', and 'S' respectively in the notation 'ETS : ExponenTial Smoothing'. The arrows are thin and black.

Error Trend Season

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

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

Seasonality: None ("N"), additive ("A") or multiplicative ("M")

ETS(A,N,A): No trend, additive seasonal model

State equations

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

$$S_t = S_{t-m} + \gamma \varepsilon_t$$

Observation equation

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

Forecast equation

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

- Forecast errors: $\varepsilon_t = y_t - \hat{y}_{t|t-1} \sim \text{NID}(0, \sigma^2)$
- k is integer part of $(h-1)/m$.

ETS(A,A,A): Holt-Winters additive method

State equations

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

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

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

Observation equation

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

Forecast equation

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

- Forecast errors: $\varepsilon_t = y_t - \hat{y}_{t|t-1} \sim \text{NID}(0, \sigma^2)$
- k is integer part of $(h-1)/m$.

ETS(M,A,M): Holt-Winters multiplicative method

State equations

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

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

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

Observation equation

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

Forecast equation

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

- Relative forecast errors: $\varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}} \sim \text{NID}(0, \sigma^2)$
- k is integer part of $(h - 1)/m$.

All ETS models

Additive Error

Trend Component

Seasonal Component

| | | N (None) | A (Additive) | M (Multiplicative) |
|----------------|-------------------|---------------------|---------------------|-----------------------|
| N | (None) | A,N,N | A,N,A | A,N,M |
| A | (Additive) | A,A,N | A,A,A | A,A,M |
| A _d | (Additive damped) | A,A _d ,N | A,A _d ,A | A,A _d ,M |

All ETS models

Additive Error

| Trend Component | | Seasonal Component | | |
|-----------------|-------------------|---------------------|---------------------|-----------------------|
| | | N (None) | A (Additive) | M (Multiplicative) |
| N | (None) | A,N,N | A,N,A | A,N,M |
| A | (Additive) | A,A,N | A,A,A | A,A,M |
| A _d | (Additive damped) | A,A _d ,N | A,A _d ,A | A,A _d ,M |

Multiplicative Error

| Trend Component | | Seasonal Component | | |
|-----------------|-------------------|---------------------|---------------------|-----------------------|
| | | N (None) | A (Additive) | M (Multiplicative) |
| N | (None) | M,N,N | M,N,A | M,N,M |
| A | (Additive) | M,A,N | M,A,A | M,A,M |
| A _d | (Additive damped) | M,A _d ,N | M,A _d ,A | M,A _d ,M |

Additive error models

| Trend | Seasonal | | |
|----------------|--|--|--|
| | N | 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

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

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

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

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

Parameter restrictions

Traditional region

- $0 < \alpha < 1$
- $0 < \beta < \alpha$
- $0 < \gamma < 1 - \alpha$
- $0.8 < \phi < 0.98$ — 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 *traditional* region
- e.g., ETS(A,N,N): *traditional* $0 < \alpha < 1$
admissible $0 < \alpha < 2$

Parameter restrictions

table default:
intersection of both regions

Traditional region

- $0 < \alpha < 1$
- $0 < \beta < \alpha$
- $0 < \gamma < 1 - \alpha$
- $0.8 < \phi < 0.98$ — 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 *traditional* region
- e.g., ETS(A,N,N): *traditional* $0 < \alpha < 1$
admissible $0 < \alpha < 2$

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

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

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

Corrected AIC

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

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

Model selection

Akaike's Information Criterion

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

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

Corrected AIC

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

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

Bayesian Information Criterion

$$BIC = AIC + k[\log(T) - 2].$$

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: $ETS(A,N,M)$, $ETS(A,A,M)$, $ETS(A,A_d,M)$.

ETS 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 |

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:

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