Exponential Smoothing Methods — Chapter 4

Smoothers : techniques to separate the signal and the noise as much as possible.

A smoother acts as a filter to obtain an 'estimate' for the signal. See figure 4.1

We have seen some smoothers :

- Moving Average
- Centered M.A's
- Hanning filter
- Moving medians

I. Consider the 'constant' process :

$$y_t = \mu + \varepsilon_t \; ; \qquad \varepsilon_t \sim (0, \sigma_\varepsilon^2)$$

[illegible]

We can 'smooth' this by replacing the current observation with the best estimate for $\mu \Rightarrow \hat{\mu}$

We know the LS estimate of μ is :

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} y_t \qquad \left(\underset{\mu}{\text{Min}} \ SSE = \sum_{t=1}^{T} (y_t - \mu)^2 \right)$$

See figure 4.2. and figure 4.3

Why does 4.3 not work? The smoother does not react quickly enough to changes in the process.

We could use a simple moving average, because it allows us to attach less weight to earlier observations, making the smoother "faster-to-react" to changes.

$$M_T = \frac{1}{N}\left[y_T + y_{T-1} + \cdots + y_{T+N-1}\right] = \frac{1}{N}\sum_{t=T-N+1}^{N} y_t .$$

If span N is small, the smoother reacts faster. However, recall that

$$Var(M_T) = \frac{\sigma^2}{N}.$$

So, as span ↓, ~~proce~~ smoother reacts faster, but is more "jittery" (variance is larger).

Are observations in M_T correlated? Yes! Successive MAs contain the same N-1 observations

∴ ACF of M·A that are k-lags apart is:

$$\rho_k = \begin{cases} 1 - \frac{|k|}{N} & ; \ k < N \\ 0 & , \ k \geq N \end{cases}$$

II. First-Order Exponential Smoothing:

Let $|\theta| < 1$ be a discount factor.
Then, to discount past observations in a geometrically decreasing fashion, we can create an exponentially weighted smoother as follows:

$$y_T + \theta y_{T-1} + \theta^2 y_{T-2} + \cdots + \theta^{T-1} y_1 = \sum_{t=0}^{T-1} \theta^t y_{T-t}$$

Note that weights do not add up to 1.

$$\sum_{t=0}^{T-1} \theta^t = \frac{1-\theta^t}{1-\theta} .$$

So, to adjust the smoother, multiply by $\frac{1-\theta}{1-\theta^t}$.

If $T \to \infty$, $\sum_{t=0}^{T-1} \theta^t = \frac{1-\theta^t}{1-\theta} \to \frac{1}{1-\theta}$.

So, multiply smoother by $1-\theta$.

$\therefore$ First order exponential smoother is:

$$\tilde{y}_T = (1-\theta) \sum_{t=0}^{T-1} \theta^t y_{T-t}$$

$$= (1-\theta)\left[y_T + \theta y_{T-1} + \theta^2 y_{T-2} + \cdots + \theta^{T-1} y_1\right)$$

$$= (1-\theta) y_T + (1-\theta)\left[\theta y_{T-1} + \theta^2 y_{T-2} + \cdots + \theta^{T-1} y_1\right)$$

$$\Rightarrow \quad \tilde{y}_T = (1-\theta)y_T + (1-\theta)\theta\left[y_{T-1} + \theta^1 y_{T-2} + \cdots + \theta^{T-2} y_1\right]$$

$$= (1-\theta)y_T + \theta\underbrace{\left\{(1-\theta)\left[y_{T-1} + \theta^1 y_{T-2} + \cdots + \theta^{T-2} y_1\right]\right\}}$$

$$\therefore \quad \boxed{\tilde{y}_T = (1-\theta)y_T + \theta\,\tilde{y}_{T-1}}$$

This is a linear combination of the current observation (y_T) and the smoothed observation at the previous time unit.

$\Rightarrow$ Linear combination of the current observation (y_T) and the discounted sum of all previous observations.

Setting $\lambda = 1-\theta$, can rewrite the first-order exponential smoother as:

$$\tilde{y}_T = \lambda y_T + (1-\lambda)\tilde{y}_{T-1}, \quad \text{where}$$

λ = discount factor = weight put on the last observation, and

$(1-\lambda)$ = weight put on the smoothed value of the previous observations.

Questions: How to choose λ?
What about $\tilde{y}_0$?

4. <u>The initial value of, $\tilde{y}_0$</u>

Recall $\quad \tilde{y}_T = \lambda y_T + (1-\lambda)\tilde{y}_{T-1}$

So,
$$\tilde{y}_1 = \lambda y_1 + (1-\lambda)\tilde{y}_0$$
$$\begin{aligned}\tilde{y}_2 &= \lambda y_2 + (1-\lambda)\tilde{y}_0 \\ &= \lambda y_2 + (1-\lambda)[\lambda y_1 + (1-\lambda)\tilde{y}_0] \\ &= \lambda(y_2 + (1-\lambda)y_1) + (1-\lambda)^2\tilde{y}_0 .\end{aligned}$$

$$\begin{aligned}\tilde{y}_3 &= \lambda y_3 + (1-\lambda)\tilde{y}_2 \\ &= \lambda(y_3 + (1-\lambda)y_2 + (1-\lambda)^2 y_1) + (1-\lambda)^3\tilde{y}_0\end{aligned}$$

$$\vdots$$

$$\therefore \quad \tilde{y}_T = \lambda(y_T + (1-\lambda)y_{T-1} + \ldots + (1-\lambda)^{T-1}y_1) + (1-\lambda)^T\tilde{y}_0 .$$

Note: If T is large, $(1-\lambda)^T \to 0$
$\therefore$ $\tilde{y}_0$ contributes little to $\tilde{y}_T$.

<u>Possibilities</u>:

a) If process is locally constant in the beginning, Take average of a subset of available data, and $\bar{y}$, and set $\tilde{y}_0 = \bar{y}$.

b) If process begins to change early, set $\tilde{y}_0 = y_1$.

B. <u>The value of λ</u> : If $\lambda = 1 \Rightarrow$ Unsmoothed version of original time series, because
$\tilde{y}_T = y_T$

If $\lambda = 0$, $\tilde{y}_T = \tilde{y}_0$ = Constant!

$\therefore$ Variance of the simple exponential smoother varies between zero (when $\lambda = 0$) and the variance of the original time series (when $\lambda = 1$)

If y_i's are independent and have constant variance,

$$\mathrm{Var}(\tilde{y}_T) = \mathrm{Var}\Big(\lambda\big(y_T + (1-\lambda) y_{T-1} + \dots + (1-\lambda)^{T-1} y_1\big) + (1-\lambda)^T \tilde{y}_0 \Big)$$

$$= \mathrm{var}\Big(\lambda \sum_{t=0}^{\infty} (1-\lambda)^t y_{T-t} + (1-\lambda)^T \tilde{y}_0 \Big)$$

$$= \lambda^2 \, \mathrm{var} \sum_{t=0}^{\infty} (1-\lambda)^t y_{T-t} + 0$$

$$= \lambda^2 \sum_{t=0}^{\infty} (1-\lambda)^{2t} \, \mathrm{var}(y_{T-t})$$

$$= \lambda^2 \cdot \mathrm{var}(y_T) \sum_{t=0}^{\infty} (1-\lambda)^{2t}$$

$$= \lambda^2 \cdot \mathrm{var}(y_T) \cdot \frac{1}{1-(1-\lambda)^2}$$ [Sum of infinite geom. series is: $S_\infty = \frac{a_1}{1-r}$

$$= \frac{\lambda}{2-\lambda} \cdot \mathrm{var}(y_T)$$

Usually, values of λ between 0.1 and 0.4 are recommended.

<u>Measures of accuracy</u>:

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_t - \tilde{y}_{t-1}}{y_t}\right| * 100 \; ; \; (y_t \neq 0)$$

$$MAD = \frac{1}{T}\sum_{t=1}^{T}\left|y_t - \tilde{y}_{t-1}\right|$$

$$MSD = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - \tilde{y}_{t-1}\right)^2$$

III. <u>Modelling Time Series Data</u>

Let $y_t = f(t;\beta) + \varepsilon_t$, where

β = vector of unknown parameters, and
$\varepsilon_t \sim (0, \sigma^2_\varepsilon)$ = uncorrelated errors.

For example, the constant-only model is:

$$y_t = \beta_0 + \varepsilon_t$$

To see how the simple exponential smoother can be used for model estimation, consider $SSE = \sum_{t=1}^{T} (y_t - \beta_0)^2$.

We can consider a modified version of the SSE which assigns geometrically decreasing weights:

$$SSE^* = \sum_{t=0}^{T-1} \theta^t (y_{T-t} - \beta_0)^2 \quad ; \quad |\theta| < 1 .$$

Minimizing SSE^* w.r.t. β_0,

$$\frac{d}{d\beta_0} SSE^* = -2 \sum_{t=0}^{T-1} \theta^t (y_{T-t} - \hat{\beta}_0) = 0$$

$$\Rightarrow \quad \hat{\beta}_0 \sum_{t=0}^{T-1} \theta^t = \sum_{t=0}^{T-1} \theta^t y_{T-t}$$

Recall $\sum_{t=0}^{T-1} \theta^t = \frac{1-\theta^t}{1-\theta}$, and for large T

$$\sum_{t=0}^{\infty} \theta^t = \frac{1}{1-\theta}$$

$$\therefore \quad \hat{\beta}_0 = \frac{1-\theta}{1-\theta^t} \sum_{t=0}^{T-1} \theta^t y_{T-t}$$, and for large T

$$\hat{\beta}_0 = 1-\theta \sum_{t=0}^{\infty} \theta^t y_{T-t} .$$

Notice Here that

$$\hat{\beta}_0 \text{ is } = \tilde{y}_T \; !$$

$\therefore$ Exponential Smoother (for constant-only model) is like a WLS !!

IV. Second-Order Exponential Smoothing

~~Consider the linear trend model:~~

$$\cancel{y_t = \beta_0 + \beta_1 t + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma_\varepsilon^2)}$$

Recall:

$$\begin{aligned}
\tilde{y}_T &= \lambda y_T + (1-\lambda)\tilde{y}_{T-1} \\
&= \lambda\left(y_T + (1-\lambda) y_{T-1} + \cdots + (1-\lambda)^{T-1} y_1\right) + (1-\lambda)^T \tilde{y}_0 \\
&= \lambda \sum_{t=0}^{\infty} (1-\lambda)^t y_{T-t}
\end{aligned}$$

~~For linear trend models,~~

$$\begin{aligned}
E(\tilde{y}_T) &= E\left(\lambda \sum_{t=0}^{\infty} (1-\lambda)^t y_{T-t}\right) \\
&= \lambda \sum_{t=0}^{\infty} (1-\lambda)^t E(y_{T-t})
\end{aligned}$$

For model with linear trend:

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t \quad ; \quad \varepsilon_t \sim (0, \sigma_\varepsilon^2)$$

$$\begin{aligned}
\therefore \quad E(\tilde{y}_T) &= \lambda \sum_{t=0}^{\infty} (1-\lambda)^t E(y_{T-t}) \\
&= \lambda \sum_{t=0}^{\infty} (1-\lambda)^t \left[\beta_0 + \beta_1 (T-t)\right] \\
&= \lambda \sum_{t=0}^{\infty} (1-\lambda)^t (\beta_0 + \beta_1 T) - \lambda \sum_{t=0}^{\infty} (1-\lambda)^t (\beta_1 t) \\
&= (\beta_0 + \beta_1 T)\, \lambda \sum_{t=0}^{\infty} (1-\lambda)^t - \lambda \beta_1 \sum_{t=0}^{\infty} (1-\lambda)^t t
\end{aligned}$$

Now, $\sum_{t=0}^{\infty} (1-\lambda)^t = \frac{1}{1-(1-\lambda)} = \frac{1}{\lambda}$

and $\sum_{t=0}^{\infty} (1-\lambda)^t \cdot t = \frac{1-\lambda}{\lambda^2}$

$$\therefore \quad E(\tilde{y}_T) = (\beta_0 + \beta_1 T) \cdot \lambda \cdot \frac{1}{\lambda} - \lambda \beta_1 \cdot \frac{(1-\lambda)}{\lambda^2}$$

$$= (\beta_0 + \beta_1 T) - \frac{(1-\lambda)}{\lambda} \cdot \beta_1$$

$$\therefore \quad E(\tilde{y}_T) = E(y_T) - \frac{(1-\lambda)}{\lambda} \cdot \beta_1$$

$\Rightarrow$ Simple exponential smoother is a <u>biased</u> estimator for the linear trend model.

Amount of bias $= E(\tilde{y}_T) - E(y_T) = -\left(\frac{1-\lambda}{\lambda}\right)\beta_1$

i) If $\lambda \to 1$, Bias $\to 0$

ii) <u>But</u>, large value of λ fails to smooth out the constant pattern in the data!

So, what do we do?

Apply simple exponential smoothing to the smoothed series! That is, use second-order exponential smoother!

Let $\tilde{y}_T^{(1)}$ = first-order smoothed exponentials,

and $\tilde{y}_T^{(2)}$ = second-order smoothed exponential.

Then,

$$\tilde{y}_T^{(2)} = \lambda\, \tilde{y}_T^{(1)} + (1-\lambda)\, \tilde{y}_{T-1}^{(2)} .$$

If the 1st-order exponential is biased, so is the second order! In fact

$$E\left(\tilde{y}_T^{(2)}\right) = E\left(\tilde{y}_T^{(1)}\right) - \frac{1-\lambda}{\lambda}\beta_1$$

And, with several substitutions, we can show that

a) $\cancel{\tilde{y}_T}$ $\hat{\beta}_{1,T} = \frac{\lambda}{1-\lambda}\left(\tilde{y}_T^{(1)} - \tilde{y}_T^{(2)}\right)$, and

(b.0) $\hat{\beta}_{0,T} = \tilde{y}_T^{(1)} - T\,\hat{\beta}_{1,T} + \frac{1-\lambda}{\lambda}\hat{\beta}_{1,T}$

b) $\Rightarrow$ $\hat{\beta}_{0,T} = \left(2 - T\cdot\frac{\lambda}{1-\lambda}\right)\tilde{y}_T^{(1)} - \left(1 - T\frac{\lambda}{1-\lambda}\right)\tilde{y}_T^{(2)}$

$\therefore$ We have a <u>predictor</u> for y_T as

$$\hat{y}_T = \hat{\beta}_{0,T} + \hat{\beta}_{1,T}\cdot T$$

$$\Rightarrow \quad \hat{y}_T = 2\tilde{y}_T^{(1)} - \tilde{y}_T^{(2)}$$

Exercise: Show that $\hat{y}_T$ is an <u>unbiased</u> predictor of y_T.

See R-Code for Examples 4.1 and 4.2.

V. Higher-Order Exponential Smoothing

For $y_t = \beta_0 + \varepsilon_t$, $\varepsilon_t \sim (0, \sigma_\varepsilon^2)$ $\rightarrow$ 1st-Order Exp. Smoother.

For $y_t = \beta_0 + \beta_1 t + \varepsilon_t$, $\varepsilon_t \sim (0, \sigma_\varepsilon^2)$ $\rightarrow$ 2nd-Order Exp. Smooth

For n'th degree polynomial model of the form

$$y_t = \beta_0 + \beta_1 t + \frac{\beta_2}{2!} t^2 + \cdots \frac{\beta_n}{n!} t^n + \varepsilon_t ,$$

$$\varepsilon_t \sim (0, \sigma_\varepsilon^2).$$

we can use the $(n+1)$'th-order exponential smoother: to estimate the model parameters.

$$\tilde{y}_T^{(1)} = \lambda y_T + (1-\lambda) \tilde{y}_{T-1}^{(1)}$$

$$\tilde{y}_T^{(2)} = \lambda \tilde{y}_T^{(1)} + (1-\lambda) \tilde{y}_{T-1}^{(2)}$$

$$\vdots$$

$$\tilde{y}_T^{(n)} = \lambda \tilde{y}_T^{(n-1)} + (1-\lambda) \tilde{y}_{T-1}^{(n)}$$

However, even with a quadratic trend model (2nd-degree polynomial), calculations are difficult analytically. So, we would prefer to use an ARIMA model, which is the subject of the next chapter.

VI. Forecasting

Let the τ-step-ahead forecast made at time T be

$$\hat{y}_{T+\tau}(T)$$

A. Constant Process: $y_t = \beta_0 + \varepsilon_t \;, \quad \varepsilon_t \sim (0, \sigma^2_\varepsilon)$.

In this case, forecast for future observation is simply the current value of the exponential smoother:

$$\Rightarrow \quad \hat{y}_{T+\tau}(T) = \tilde{y}_T = \lambda y_T + (1-\lambda)\tilde{y}_{T-1}$$

When y_{T+1} becomes available, we can UPDATE our forecast : $\tilde{y}_{T+1} = \lambda y_{T+1} + (1-\lambda)\tilde{y}_T$

Therefore,

$$\hat{y}_{T+1+\tau}(T+1) = \lambda y_{T+1} + (1-\lambda)\tilde{y}_T$$
$$= \lambda y_{T+1} + (1-\lambda)\hat{y}_{T+\tau}(T)$$

E.g: when $\tau = 1$,

$$\hat{y}_{T+2}(T+1) = \lambda y_{T+1} + (1-\lambda)\hat{y}_{T+1}(T)$$
$$= \lambda y_{T+1} + \hat{y}_{T+1}(T) - \lambda \hat{y}_{T+1}(T)$$

Rearranging, we have

$$= \hat{y}_{T+1}(T) + \lambda \underbrace{\left(y_{T+1} - \hat{y}_{T+1}(T)\right)}$$

= Previous forecast for current observation + $\lambda *$ Forecast Error made in makin[g] ~~current ob.~~ forecast of curre[nt] observation

$$\therefore \quad \hat{y}_{T+2}(T+1) = \hat{y}_{T+1}(T) + \lambda * e_{T+1}(1).$$

One-step ahead forecast error!

∴ Forecast for next observation =
Previous forecast for current observation
+ Fraction of forecast error in forecasting current observation!

So ... λ (discount factor) determines HOW FAST our forecast reacts to the forecast error.

Large $\lambda \Rightarrow$ Fast Reaction to FE. BUT
Large λ also $\Rightarrow$ forecast reacts fast to random fluctuations!

So ... how to choose λ?
Choose the λ that ...
... Minimises sum of squared forecast errors!!

$$SSE(\lambda) = \sum_{t=1}^{T} e_t^2(1) \cdot \left(\begin{array}{l} \text{Calculate for various} \\ \text{values of } \lambda \end{array} \right)$$

We need to always supply intervals for our forecasts

A $(100-\alpha)\%$ P.I. for any lead time τ is:

$$\boxed{\tilde{y}_T \pm Z_{\alpha/2}} \cdot \hat{\sigma}_\varepsilon \text{, where}$$

$\tilde{y}_T$ = 1st-order Exponential smoother;
$Z_{\frac{\alpha}{2}}$ = ~~100~~ The relevant value from Std. Normal Dist
and $\hat{\sigma}_e$ = standard error of the forecast errors.

Question: This gives constant P.I's. NOT GOOD!

Example 4.4 : Note the $\hat{y}_{T+\tau}(T)$ and P.I.'s for the forecasts : All Constant!

B. Example 4.5 : First-Order Exponential Smoother for a model with LINEAR TREND.

Problem : SSE ↓ as λ ↑

⇒ Data are autocorrelated!

See ACF.

∴ ARIMA may be better. Chapter 6!

OR, try 2nd-Order Exponential Smoothing.

And try ① 1-step-, 2-step-, ..., 12-step-ahead forecasts; OR

② 1-step-ahead forecasts

The latter performs better.

C. Standard Error of forecasts, $\hat{\sigma}_e$

Assuming the model is correct (and constant in time), define 1-step-ahead FE as:

$$e_T(1) = y_T - \hat{y}_T(T-1).$$

Then,

$$\hat{\sigma}_e^2 = \frac{1}{T}\sum_{t=1}^{T} e_t^2(1) = \frac{1}{T}\sum_{t=1}^{T}\left[y_t - \hat{y}_t(t-1)\right]^2.$$

And $\sigma_e^2(\tau)$ = τ-step-ahead forecast variance

$$= \frac{1}{T-\tau+1}\sum_{t=\tau}^{T} e_t^2(\tau)$$

Not same as $e_\tau^2(1)$

D. Adaptive Updating of λ, the discount factor

Trigg and Leach Method:

$$\hat{y}_T = \lambda_T \cdot y_T + (1-\lambda_T)\,\tilde{y}_{T-1}.$$

Let Mean Absolute Deviation, Δ, be defined as:

$$\Delta = E\left(|e - E(e)|\right).$$ Then, the estimate of MAD is:

$$\hat{\Delta}_T = \delta|e_T(1)| + (1-\delta)\,\hat{\Delta}_{T-1}$$

Also define the Smoothed Error as:

$$Q_T = \delta\, e_T(1) + (1-\delta)\, Q_{T-1}.$$

Finally, let the "tracking signal" be defined as:

$$Q_T / \hat{\Delta}_T.$$

(*) Then, setting $\boxed{\lambda_T = \left|\dfrac{Q_T}{\hat{\Delta}_T}\right|}$ will allow for automatic updating of the discount factor.

Example 4.6 : See R Code.

E. Model Assessment: Plot ACF of forecast errors. If sample ACF exceeds ± 2 s.e., violates assumption of uncorrelated errors!

VII. Exponential Smoothing for Seasonal Data

A. Additively seasonal model

Let $\boxed{y_t = L_t + S_t + \varepsilon_t}$, where

$L_t = \beta_0 + \beta_1 t$ = Level + Trend Component
= Permanent Component.

S_t = Seasonal Component, with
$S_t = S_{t+s} = S_{t+2s} = \cdots$ for $t = 1, 2, \ldots, s-1$, where
s = length of season, or period of cycles;
and

ε_t = Random component,
$\sim (0, \sigma_\varepsilon^2)$.

Also, $\sum_{t=1}^{s} S_t = 0 \Rightarrow$ Seasonal Adjustments add to zero during one season.

~~To Steps~~ in forecasting future observations; we use first-order exponential smoothing:

Assuming that the current observation, y_T is obtained, we would like to make τ-step-ahead forecasts.

The principal idea is to update estimates of: $\hat{L}_T$, $\hat{\beta}_{1,T}$, and $\hat{S}_T$, so that τ-step-ahead forecast of y is: $\boxed{\hat{y}_{T+\tau}(T) = \hat{L}_T + \hat{\beta}_{1,T} \cdot \tau + \hat{S}_T(\tau - s)}$

Here, $\hat{L}_T = \underbrace{\lambda_1 (y_T - \hat{S}_{T-s})}_{\lambda_1 * \text{ Current value of } L_T} + (1-\lambda_1)\underbrace{(\hat{L}_{T-1} + \hat{\beta}_{1,T-1})}_{\text{Forecast of } L_T \text{ based on estimates at } T-1}$.

Similarly,

$$\hat{\beta}_{1,T} = \lambda_2 \underbrace{(\hat{L}_T - \hat{L}_{T-1})}_{\text{Current value of } \beta_1} + (1-\lambda_2)\underbrace{\hat{\beta}_{1,T-1}}_{\text{Forecast of } \beta_1 \text{ at time } T-1}.$$

and,

$$\hat{S}_T = \lambda_3 (y_T - \hat{L}_T) + (1-\lambda_3)\hat{S}_{T-s}.$$

Note that $0 < \lambda_1, \lambda_2, \lambda_3 < 1$.

Once again, we need initial values of the exponential smoothers:

all at time $t=0$
$$\begin{cases} \hat{\beta}_{0,0_T} = \hat{L}_0 = \hat{\beta}_0 \\ \hat{\beta}_{1,0_T} = \hat{\beta}_1 \\ \hat{S}_{j-s} = \hat{y}_j \quad \text{for } 1 \le j \le s-1 \\ \hat{S}_0 = -\sum_{j=1}^{s-1} \hat{y}_j \end{cases}$$

Example 4.7 : See R-Code.

B. Multiplicative Seasonal Model

What if the seasonality is proportional to the average level of the seasonal time series.

$$y_t = L_t \cdot S_t + \varepsilon_t$$, where L_t, S_t, and ε_t are as before.

Once again, we can use three exponential smoothers to obtain parameter estimates in the eqn.

Example 4.8 See R-Code.

VIII. Biosurveillance Data : Please read pp 286-299 carefully!

DIY

IX. Exponential Smoothing and ARIMA models

Recall : $\tilde{y}_T = \lambda y_T + (1-\lambda)\tilde{y}_{T-1}$

forecast error : $e_T = y_T - \hat{y}_{T-1}$, and

$\therefore \; e_{T-1} = y_{T-1} - \hat{y}_{T-2}$

$(1-\lambda) * e_{T-1} = (1-\lambda) * y_{T-1} - (1-\lambda) * \hat{y}_{T-2}$

~~Subtract from e_T~~

$\Rightarrow (1-\lambda)\, e_{T-1} = y_{T-1} - \lambda y_{T-1} - (1-\lambda)\hat{y}_{T-2}$

$\Rightarrow$ $\cancel{e_T - (1-\lambda) e_{T-1} = y_T - y}$

Subtract $(1-\lambda) e_{T-1}$ from e_T :

$$\Rightarrow e_T - (1-\lambda)e_{T-1} = (y_T - \hat{y}_{T-1}) - (1-\lambda)[y_{T-1} - \hat{y}_{T-2}]$$

$$\Rightarrow e_T - (1-\lambda)e_{T-1} = y_T - \hat{y}_{T-1} - y_{T-1} + \underbrace{\lambda y_{T-1} + (1-\lambda)\hat{y}_{T-2}}_{= \hat{y}_{T-1}}$$

$$\Rightarrow e_T - (1-\lambda)e_{T-1} = y_T - \hat{y}_{T-1} - y_{T-1} + \hat{y}_{T-1}$$
$$= y_T - y_{T-1}$$

$$\therefore \quad y_T - y_{T-1} = e_T - (1-\lambda)e_{T-1}$$
$$= e_T - \theta e_{T-1}, \text{ where } \theta = 1-\lambda.$$

$$\therefore \quad (1-B)y_T = (1-\theta B)e_T, \text{ where}$$
$B =$ Backshift Operator.

This is an <u>I</u>ntegrated <u>M</u>oving <u>A</u>verage model of order $(1, 1)$

$\uparrow$ I $\quad \uparrow$ MA

We will do ARIMA models in detail next.