Simple Exponential Smoothing (SES)

Prof. Juan Garbayo

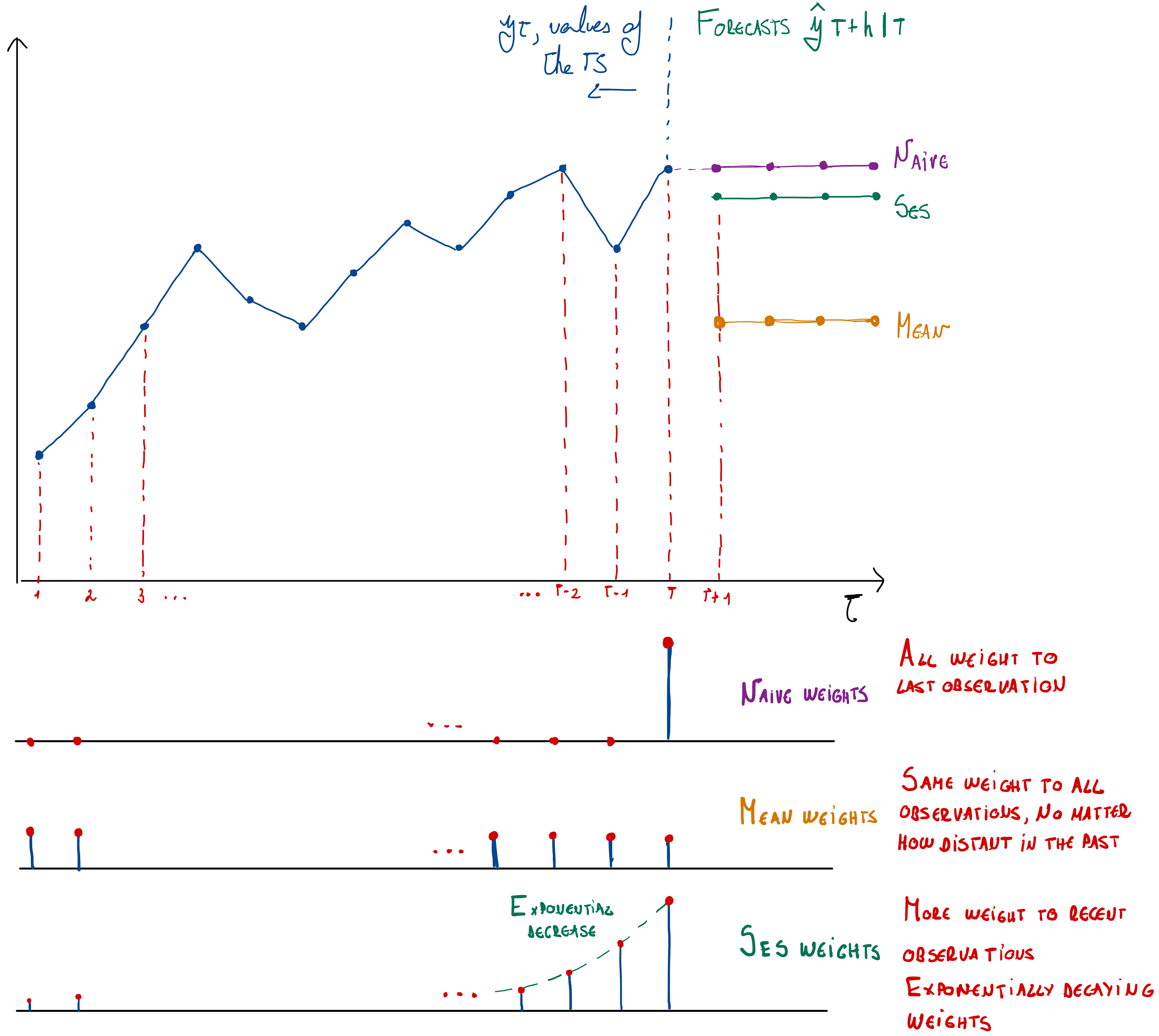
Time Series Analysis

UNIVERSITY

Contents

1. SES vs Naive vs Mean
2. Geometric progression – basic properties
3. SES - weighted average form
4. SES - components form - the level
5. Filled values and flat forecasts
6. Exponential Moving Average vs Simple Moving Average

ie UNIVERSITY



y_t, values of the TS
Forecasts $\hat{y}_{T+h|T}$
Naive
Ses
Mean
1
2
3 ...
... T-2
T-1
T
T+1
t
Naive weights
All weight to last observation
Mean weights
Same weight to all observations, no matter how distant in the past
Exponential decrease
Ses weights
More weight to recent observations
Exponentially decaying weights

Geometric sequence (1/3)

$r = \frac{a_n}{a_{n-1}}$ → ratio (given) → Definition

a_1 → given

$a_2 = r \cdot a_1$

$a_3 = r^2 a_1$

$\vdots$

$a_n = r^{n-1} \cdot a_1$

If $r < 1 \Rightarrow \lim_{n \to \infty} a_n = 0$

A specific example:

$a_1 = \alpha \qquad \alpha \in (0, 1]$

$r = (1 - \alpha) \Rightarrow (1 - \alpha) \in [0, 1)$

$a_n = \alpha \underbrace{(1 - \alpha)}_{< 1}{}^{n-1}$ → a_n decreases exponentially with ↑ n

$\lim_{n \to \infty} a_n = 0$

IE UNIVERSITY

Geometric sequence (2/3)

Sum of the first n terms of a geometric sequence

$$S_n = a_1 + \overbrace{a_1 \cdot r}^{a_2} + \overbrace{a_1 \cdot r^2}^{a_3} + \dots + \overbrace{a_1 \cdot r^{n-1}}^{a_n}$$

$$r \cdot S_n = a_1 \cdot r + a_1 \cdot r^2 + a_1 \cdot r^3 + \dots + a_1 \cdot r^n$$

$$S_n - r \cdot S_n = S_n(1-r) = a_1 - a_1 \cdot r^n = a_1(1-r^n)$$

For $r < 1$, S_n converges

Solving for S_n $\rightarrow$ $S_n = \dfrac{a_1(1-r^n)}{(1-r)}$ $\rightarrow$ If $r < 1 \Rightarrow \lim\limits_{n \to \infty} S_n = \dfrac{a_1}{1-r}$

Specific case: $\left.\begin{array}{l} a_1 = \alpha \in (0,1] \\ r = (1-\alpha) < 1 \end{array}\right\} \rightarrow S_n = \dfrac{\alpha\left(1-(1-\alpha)^n\right)}{1-(1-\alpha)} = 1-(1-\alpha)^n$

$$\left[S_n = \left(1-(1-\alpha)^n\right) ; \lim_{n \to \infty} S_n = 1 \right]$$

ie UNIVERSITY

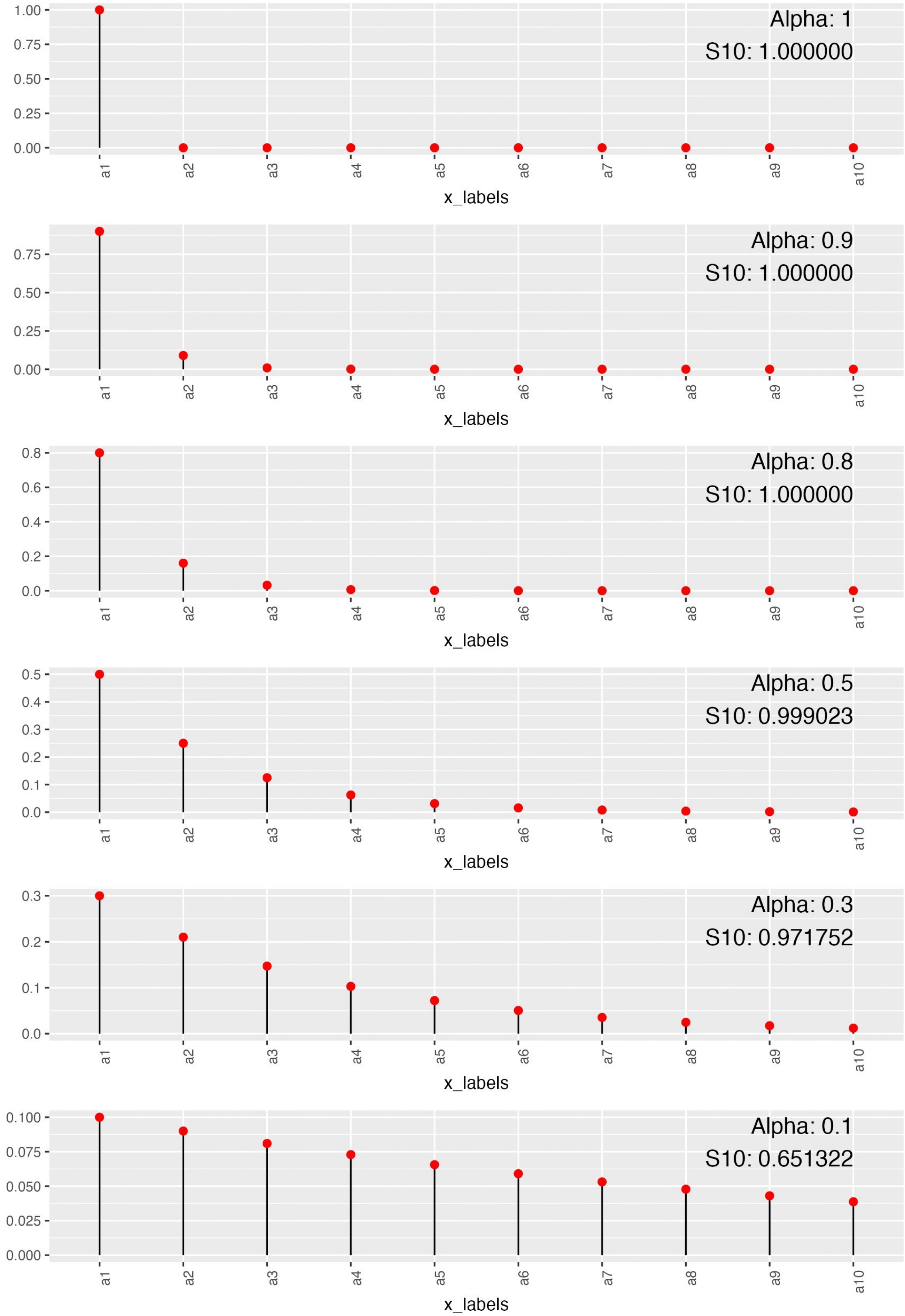
Geometric sequence (3/3)

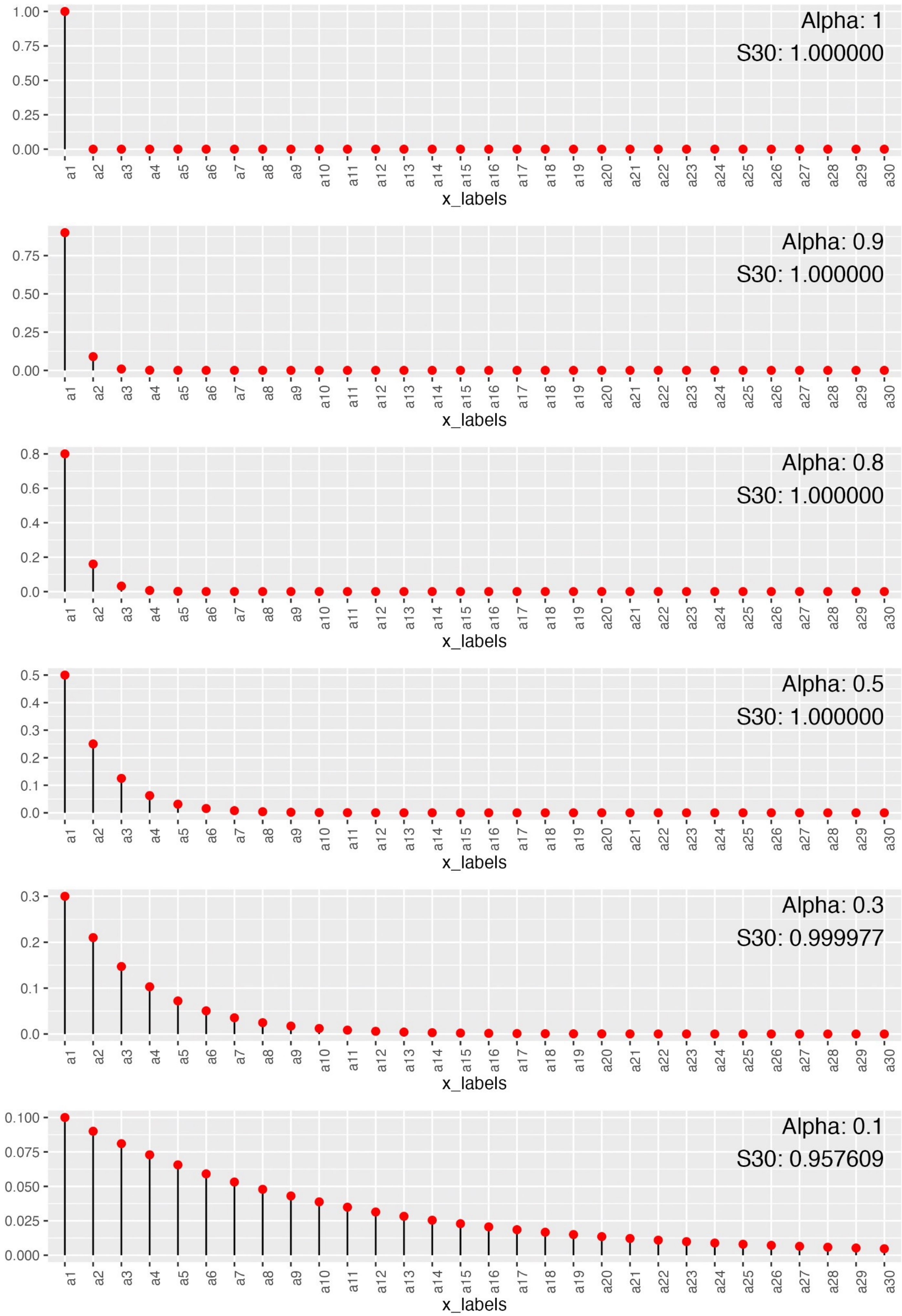
We have shown that, for $\left.\begin{array}{l} a_1 = \alpha \in (0, 1] \\ r = (1-\alpha) \in [0, 1) \end{array}\right\}$

EXPONENTIAL DECAY

- $a_n = \alpha (\underbrace{1-\alpha}_{<1})^{n-1}$; $\lim_{n\to\infty} a_n = 0$ → DISTANT OBSERVATIONS HAVE NEGLIGIBLE WEIGHTS
- $S_n = 1-(1-\alpha)^n$; $\lim_{n\to\infty} S_n = 1$ → a_i CAN BE INTERPRETED AS THE PERCENT WEIGHT ASSIGNED TO y_i

ie UNIVERSITY


Alpha: 1
S10: 1.000000
Alpha: 0.9
S10: 1.000000
Alpha: 0.8
S10: 1.000000
Alpha: 0.5
S10: 0.999023
Alpha: 0.3
S10: 0.971752
Alpha: 0.1
S10: 0.651322
x_labels
a1 a2 a3 a4 a5 a6 a7 a8 a9 a10


Alpha: 1
S30: 1.000000
Alpha: 0.9
S30: 1.000000
Alpha: 0.8
S30: 1.000000
Alpha: 0.5
S30: 1.000000
Alpha: 0.3
S30: 0.999977
Alpha: 0.1
S30: 0.957609
x_labels

Weighted Average Form

- Given these properties, we may produce a forecast with exp decreasing weights:

$$\hat{y}_{T+1|T} = \underbrace{\alpha}_{a_1} y_T + \underbrace{\alpha(1-\alpha)}_{a_2}{}^{1} y_{T-1} + \underbrace{\alpha(1-\alpha)}_{a_3}{}^{2} y_{T-2} + \dots + \underbrace{\alpha(1-\alpha)}_{a_T}{}^{T-1} y_1$$

$\hookrightarrow 1 = T-(T-1)$

$a_n = \alpha(1-\alpha)^{n-1} \leftarrow \alpha(1-\alpha)^{T-1}$

- We may express this in a compact manner using the sum operator:

$$\hat{y}_{T+1|T} = \sum_{i=1}^{T} a_i \, y_{T-(i-1)} = \sum_{i=1}^{T} \alpha(1-\alpha)^{i-1} y_{T-(i-1)} =$$

$$= \sum_{j=0}^{T-1} \alpha(1-\alpha)^{j} y_{T-j}$$

Re-indexing: $j = i-1$

Weighted average form

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

By definition of the SES model

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

Flat forecasts

ie UNIVERSITY

Fitted values using W. Av. Form

Fitted values are **defined** as $\hat{y}_{t+1|t}$

$$\hat{y}_{T|T-1} = \alpha\, y_{T-1} + \alpha(1-\alpha)\, y_{T-2} + \alpha(1-\alpha)^2\, y_{T-3} + \dots + \alpha(1-\alpha)^{T-2}\, y_1$$

$\llcorner\, 1 = T-(T-1)$

$$\hat{y}_{T-1|T-2} = \alpha\, y_{T-2} + \alpha(1-\alpha)\, y_{T-3} + \alpha(1-\alpha)^2\, y_{T-4} + \dots + \alpha(1-\alpha)^{T-3}\, y_1$$

$\hookrightarrow 1 = T-(T-1)$

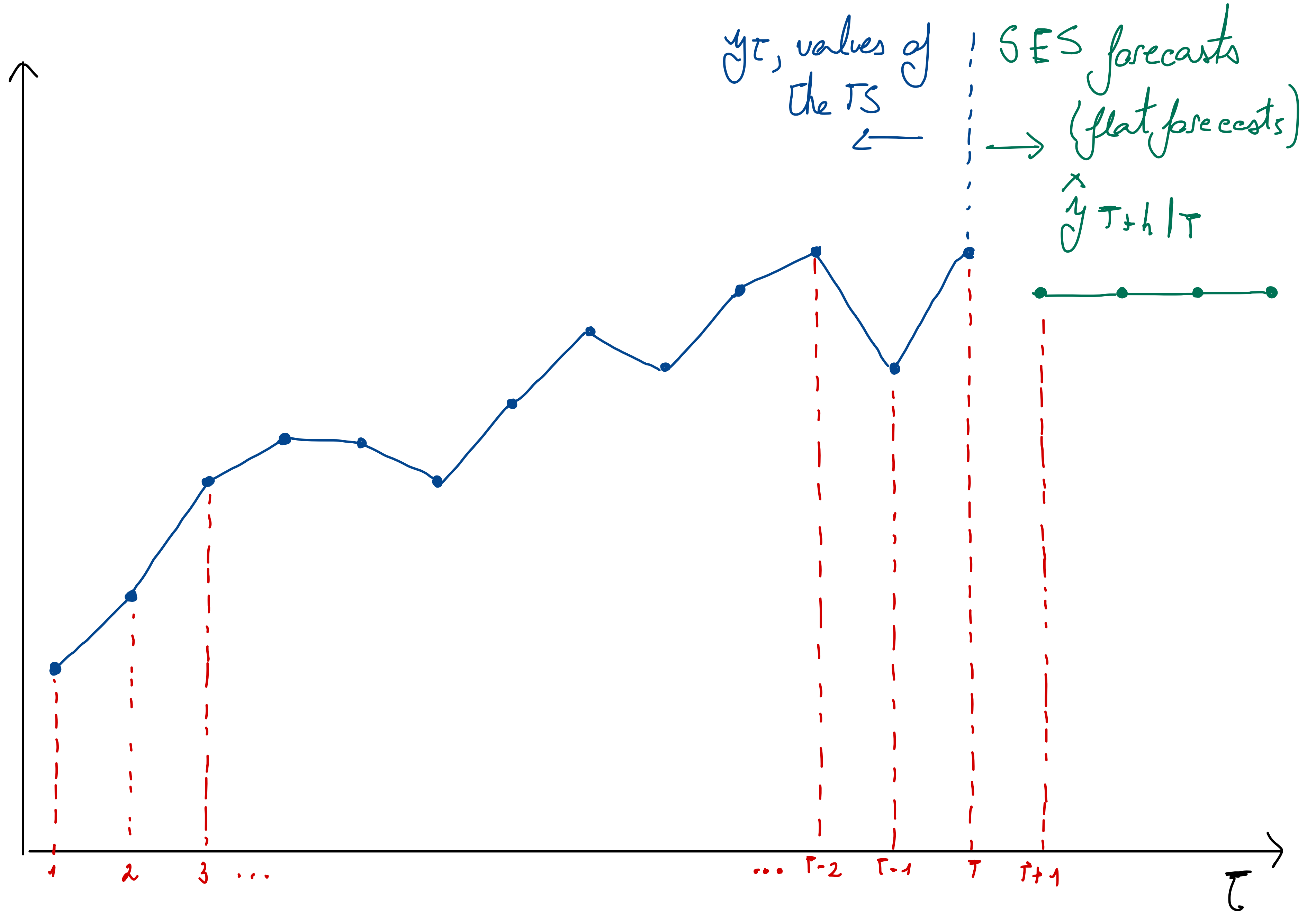
$\vdots$

$$\hat{y}_{3|2} = \alpha\, y_2 + \alpha(1-\alpha)\, y_1$$

$$\hat{y}_{2|1} = \alpha\, y_1$$

$$\hat{y}_{1|0} = ?$$

$\llcorner\, \alpha \cdot y_0 ??$



Problem:

- → Computationally quite inefficient → The amount of multiplications and sums may be substantially reduced
- → No fitted value at $t = 1$
- → Only one parameter α to adjust the fit of our model

IE University

Recursive equation for the fitted values

Weighted average form

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

We are going to use an alternative procedure to compute the fitted values that will solve the problems outlined and enable us to interpret exponential smoothing in terms of "components". This will be most clear in trended and seasonal exponential smoothing.

$$\hat{y}_{T+1|T} = \underset{\uparrow\, a_1}{\alpha y_T} + \overbrace{\underset{\uparrow\, a_2}{\alpha(1-\alpha)^{1} y_{T-1}} + \alpha(1-\alpha)^2 y_{T-2} + \dots + \underset{\uparrow\, a_T}{\alpha(1-\alpha)^{T-1} y_1}}^{(*)}$$

$1 = T-(T-1)$

$$\hat{y}_{T|T-1} = \alpha y_{T-1} + \alpha(1-\alpha)^{1} y_{T-2} + \alpha(1-\alpha)^2 y_{T-3} + \dots + \alpha(1-\alpha)^{T-2} y_1$$

$1 = T-(T-1)$

$$(1-\alpha)\hat{y}_{T|T-1} = \underbrace{\alpha(1-\alpha) y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \alpha(1-\alpha)^3 y_{T-3} + \dots + \alpha(1-\alpha)^{T-1} y_1}_{(*)}$$

$$\hat{y}_{T+1|T} = \alpha \cdot y_T + (*) = \alpha \cdot y_T + (1-\alpha)\hat{y}_{T|T-1} \Longrightarrow \boxed{\hat{y}_{T+1|T} = \alpha y_T + (1-\alpha)\hat{y}_{T|T-1}}$$

Recursive equation for the fitted values

No new info. Direct consequence of the weighted average form

ie UNIVERSITY

Recursive computation of fitted values

C.V.: $\tau = T+1 \Rightarrow T = \tau - 1$

Note that $\alpha + (1-\alpha) = 1$

$$\boxed{\hat{y}_{T+1|T} = \alpha y_T + (1-\alpha)\hat{y}_{T|T-1}} \Longleftrightarrow \boxed{\hat{y}_{\tau|\tau-1} = \alpha y_{\tau-1} + (1-\alpha)\hat{y}_{\tau-1|\tau-2}}$$

Equation at $\tau = T+1$

Equation at generic instant τ

The fitted value at τ is a weighted average of:

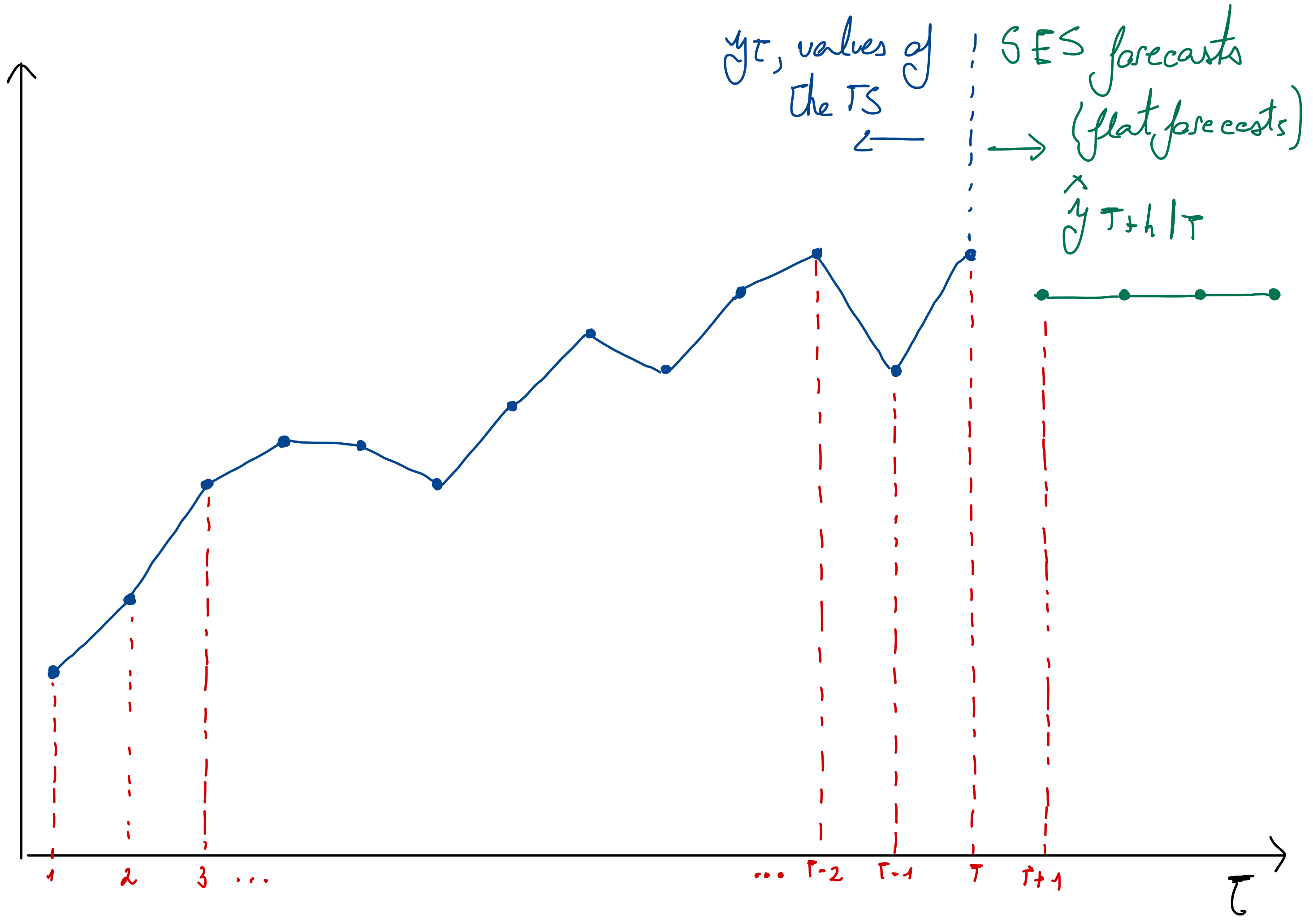
→ The observation at $\tau-1$ ($y_{\tau-1}$)

→ The fitted value at $\tau-1$ ($\hat{y}_{\tau-1|\tau-2}$)

τ	$\hat{y}_{\tau\|\tau-1}$
T	$\hat{y}_{T\|T-1} = \alpha y_{T-1} + (1-\alpha)\hat{y}_{T-1\|T-2}$
$T-1$	$\hat{y}_{T-1\|T-2} = \alpha y_{T-2} + (1-\alpha)\hat{y}_{T-2\|T-3}$
$T-2$	$\hat{y}_{T-2\|T-3} = \alpha y_{T-3} + (1-\alpha)\hat{y}_{T-3\|T-4}$
$\vdots$	$\vdots$
3	$\hat{y}_{3\|2} = \alpha y_2 + (1-\alpha)\hat{y}_{2\|1}$
2	$\hat{y}_{2\|1} = \alpha y_1 + (1-\alpha)\hat{y}_{1\|0}$
1	$\hat{y}_{1\|0} = \ell_0$

↳ Initial value for the level

↳ New parameter



Fitted Values as updating aproximation

Note that $\alpha + (1-\alpha) = 1$

$$\boxed{\hat{y}_{\tau|\tau-1} = \alpha\, y_{\tau-1} + (1-\alpha)\,\hat{y}_{\tau-1|\tau-2}} \oplus \boxed{\hat{y}_{1|0} = \ell_0}$$

Equation at generic instant τ

The fitted value at τ $\hat{y}_{\tau|\tau-1}$ is a weighted average of:

→ The observation at $\tau-1$ $\quad y_{\tau-1}$

→ The fitted value at $\tau-1$ $\quad \hat{y}_{\tau-1|\tau-2}$

→ Getting a new observation at τ y_τ allows us to update our approximation for the next point in time $\tau+1$

$$\underbrace{\hat{y}_{\tau+1|\tau}}_{\text{New approx}} = \alpha \cdot \underbrace{y_\tau}_{\substack{\text{New info}\\\text{(observation)}}} + (1-\alpha)\underbrace{\hat{y}_{\tau|\tau-1}}_{\substack{\text{Previous}\\\text{approx}}}$$

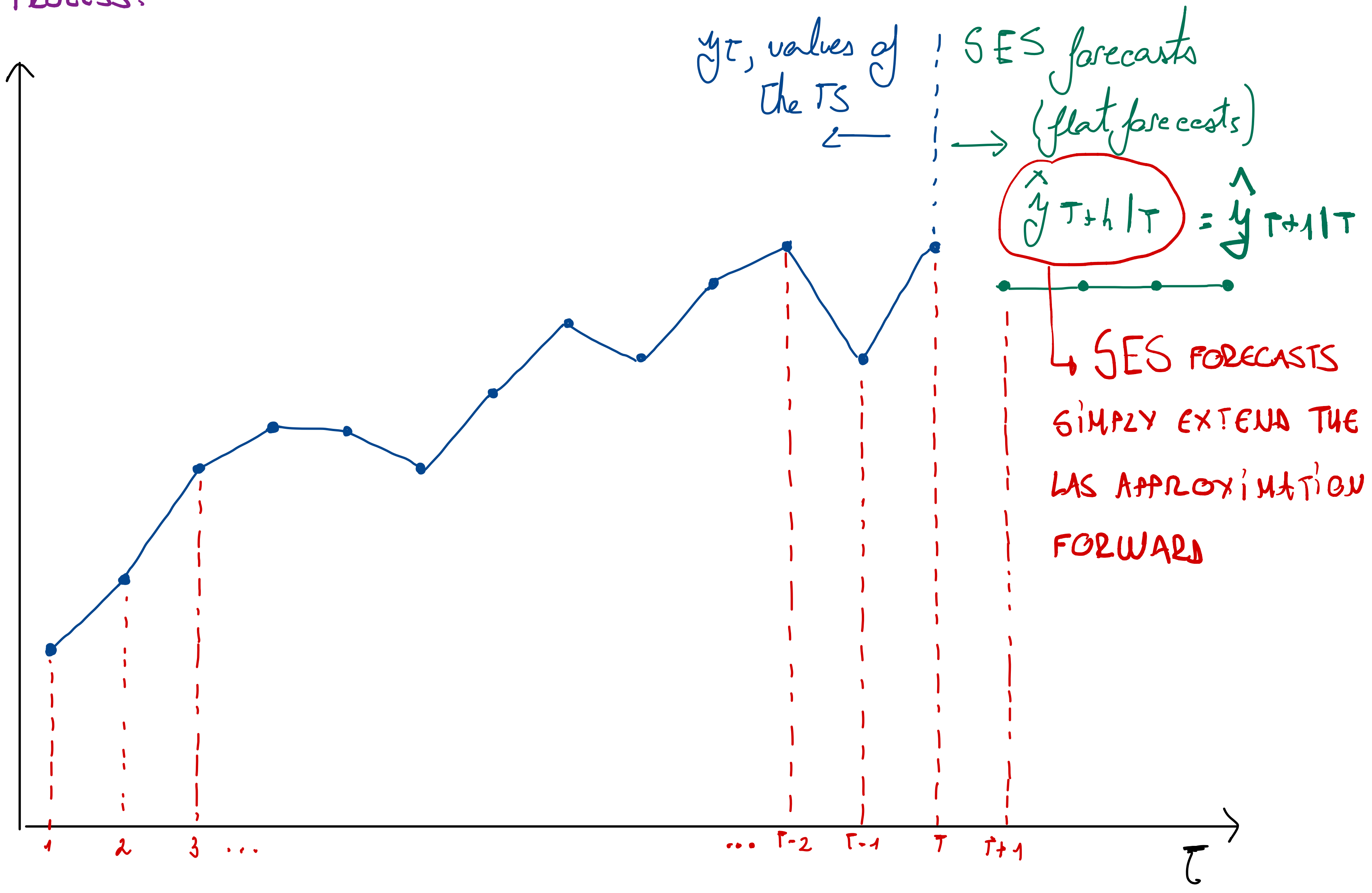
τ	$\hat{y}_{\tau\vert\tau-1}$
1	$\hat{y}_{1\vert 0} = \ell_0$
2	$\hat{y}_{2\vert 1} = \alpha\, y_1 + (1-\alpha)\,\hat{y}_{1\vert 0}$
3	$\hat{y}_{3\vert 2} = \alpha \cdot y_2 + (1-\alpha)\,\hat{y}_{2\vert 1}$
⋮	⋮
$T-2$	$\hat{y}_{T-2\vert T-1} = \alpha \cdot y_{T-1} + (1-\alpha)\,\hat{y}_{T-3\vert T-4}$
$T-1$	$\hat{y}_{T-1\vert T-2} = \alpha \cdot y_{T-2} + (1-\alpha)\,\hat{y}_{T-2\vert T-1}$
T	$\hat{y}_{T\vert T-1} = \alpha \cdot y_{T-1} + (1-\alpha)\,\hat{y}_{T-1\vert T-2}$

Initial level. Guess. Parameter of process.

Observation: new info

Previous Approximation

Updated Approximation



ie UNIVERSITY

FITTING THE MODEL AND GETTING FORECASTS

Given α and l_0, the equations below provide the fitted values at each point:

Initial level. Guess. Parameter of process.

	t	$\hat{y}_{t\mid t-1}$
TRAINING	1	$\hat{y}_{1\mid 0} = l_0$
	2	$\hat{y}_{2\mid 1} = \alpha\, y_1 + (1-\alpha)\,\hat{y}_{1\mid 0}$
	3	$\hat{y}_{3\mid 2} = \alpha \cdot y_2 + (1-\alpha)\,\hat{y}_{2\mid 1}$
	$\vdots$	$\vdots$
	$T-2$	$\hat{y}_{T-2\mid T-1} = \alpha \cdot y_{T-1} + (1-\alpha)\,\hat{y}_{T-3\mid T-4}$
	$T-1$	$\hat{y}_{T-1\mid T-2} = \alpha \cdot y_{T-2} + (1-\alpha)\,\hat{y}_{T-2\mid T-1}$
	T	$\hat{y}_{T\mid T-1} = \alpha \cdot y_{T-1} + (1-\alpha)\,\hat{y}_{T-1\mid T-2}$
TEST	$T+1$	$\hat{y}_{T+1\mid T} = \alpha \cdot y_T + (1-\alpha)\,\hat{y}_{T\mid T-1}$
		$= \hat{y}_{T+h\mid T}$ (SES has flat forecasts)

Observation: new info

Updated approximation

Previous approximation

Excel File

1. Determine initial values

l_0, α

2. Apply the component form of the equations to calculate the fitted values.
That is, at each point in time t compute:

- $\hat{y}_{1|0} = l_0$
- $\hat{y}_{2|1} = \alpha\, y_1 + (1-\alpha)\,\hat{y}_{1|0}$
- $\hat{y}_{3|2} = \alpha\, y_2 + (1-\alpha)\,\hat{y}_{2|1}$
- $\vdots$
- $\hat{y}_{T|T-1} = \alpha\, y_{T-1} + (1-\alpha)\,\hat{y}_{T-1|T}$

3. Compute the Sum of Square Residuals

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

4. Minimality of SSE satisfied?

NO

YES

Model has been fitted.

Flat forecasts:

$$\hat{y}_{T+h|T} = \hat{y}_{T+1|T} = \alpha\, y_T + (1-\alpha)\,\hat{y}_{T|T-1}$$

Parameters: α and ℓ_0

→ α: determines the weight assigned to the new info $(\alpha \cdot y_t)$ and the previous approximation $((1-\alpha)\,\hat{y}_{t|t-1})$

Smoothing parameter

→ Equivalent to: determines how the weight assigned to past observations decays

Reminder
$\alpha \in (0,1] \rightarrow 1-\alpha \in [0,1)$
$\alpha + (1-\alpha) = 1 \rightarrow$ Weighted average

→ ℓ_0: influences the value of the fitted exponential smoothing

→ Its influence becomes weaker the more distant from $t=1$ the process gets:

t	
1	$\hat{y}_{1\|0} = \ell_0$
2	$\hat{y}_{2\|1} = \alpha y_1 + (1-\alpha)\hat{y}_{1\|0} = \alpha y_1 + (1-\alpha)\ell_0$
3	$\hat{y}_{3\|2} = \alpha \cdot y_2 + (1-\alpha)\hat{y}_{2\|1} = \alpha y_2 + (1-\alpha)[\alpha \cdot y_1 + (1-\alpha)\ell_0] =$
	$= \alpha y_2 + \alpha(1-\alpha) \cdot y_1 + \underbrace{(1-\alpha)^2}_{<1}\ell_0$
⋮	⋮
Generic point t	$\hat{y}_{t\|t-1} = \alpha y_{t-1} + \alpha(1-\alpha) y_{t-2} + \alpha(1-\alpha)^2 y_{t-3} + \dots + \alpha(1-\alpha)^{t-2} y_1 + \boxed{(1-\alpha)^{t-1} \cdot \ell_0}$

$\hat{y}_{1|0} = \ell_0$

$1 = t-(t-1)$

$(1-\alpha) < 1$

Influence of ℓ_0 decays exponentially with increasing t

Weighted Average Form

Definition of the level ℓ_τ

We define the level at τ: $\boxed{\ell_\tau = \hat{y}_{\tau+1|\tau}}$ $\Rightarrow \ell_0 = \hat{y}_{1|0}$

T	$\hat{y}_{t|t-1}$
T	$\hat{y}_{T|T-1} = \alpha\, y_{T-1} + (1-\alpha) \overbrace{\hat{y}_{T-1|T-2}}^{\ell_{\tau-2}}$
$T-1$	$\underbrace{\hat{y}_{T-1|T-2}}_{\ell_{\tau-2}} = \alpha\, y_{T-2} + (1-\alpha) \underbrace{\hat{y}_{T-2|T-3}}_{\ell_{\tau-1}}$
$T-2$	
$\vdots$	$\vdots \quad \vdots \quad \vdots$
3	$\overbrace{\hat{y}_{3|2}}^{\ell_2} = \alpha\, y_2 + (1-\alpha) \overbrace{\hat{y}_{2|1}}^{\ell_1}$
2	$\hat{y}_{2|1} = \alpha\, y_1 + (1-\alpha)\, \hat{y}_{1|0}$
1	$\hat{y}_{1|0} = \ell_0$

$\ell_{\tau-1}$ $\qquad$ ℓ_1 $\qquad$ ℓ_0

ℓ_0 → Initial value for the level

In summary, with this definition of ℓ_τ

$$\hat{y}_{\tau+1|\tau} = \alpha \cdot y_\tau + (1-\alpha)\, \hat{y}_{\tau|\tau-1}$$

$$\boxed{\ell_\tau = \alpha \cdot y_\tau + (1-\alpha)\, \ell_{\tau-1}}$$

SMOOTHING EQUATION

(ALTERNATIVE FORM OF RECURSIVE EQUATION FOR FITTED VALUES)

ie UNIVERSITY

COMPONENT FORM OF THE SES EQUATIONS

FLAT FORECASTS   DEFINITION OF LEVEL $\ell_T = \hat{y}_{T+1|T}$

FORECAST EQUATION $\hat{y}_{T+h|T} = \hat{y}_{T+1|T} = \ell_T$

$\ell_T = \hat{y}_{T+1|T}$ (DEFINITION OF LEVEL)

SMOOTHING EQUATION: $\underbrace{\hat{y}_{T+1|T}}_{\ell_T} = \alpha \cdot y_T + (1-\alpha)\underbrace{\hat{y}_{T|T-1}}_{\ell_{T-1}} \longleftrightarrow \ell_T = \alpha \cdot y_T + (1-\alpha)\,\ell_{T-1}$

SES · EQUATIONS (2 EQUIVALENT FORMS)

WEIGHTED AVERAGE FORM	COMPONENT FORM
$\hat{y}_{T+1\|T} = \sum_{j=0}^{T-1} \alpha(1-\alpha)^j y_{T-j}$	FORECAST. EQ: $\hat{y}_{T+h\|T} = \hat{y}_{T+1\|T} = \ell_T$
	SMOOTHING. EQ: $\ell_T = \alpha \cdot y_T + (1-\alpha)\,\ell_{T-1}$

ie UNIVERSITY

Exponential Moving Average and ℓ_t

Investopedia:

Formula for Exponential Moving Average (EMA)

Usually ~ 2

$$EMA_{\text{Today}} = \left(\text{Value}_{\text{Today}} * \overbrace{\left(\frac{\text{Smoothing}}{1 + \text{Days}}\right)}^{\alpha}\right) + EMA_{\text{Yesterday}} * \left(1 - \underbrace{\left(\frac{\text{Smoothing}}{1 + \text{Days}}\right)}_{\alpha}\right)$$

e.g: 20 days → EMA_{20}

where:

EMA = Exponential moving average

$$EMA_{TODAY} = \alpha \cdot y_{TODAY} + (1-\alpha)\, EMA_{YESTERDAY}$$

$$EMA_t = \alpha \cdot y_t + (1-\alpha)\, EMA_{t-1}$$

Component Form of SES

Forecast. Eq: $\hat{y}_{T+h|T} = \hat{y}_{T+1|T} = \ell_T$

Smoothing. Eq: $\ell_t = \alpha \cdot y_t + (1-\alpha)\,\ell_{t-1}$

The level is the EMA of the time series

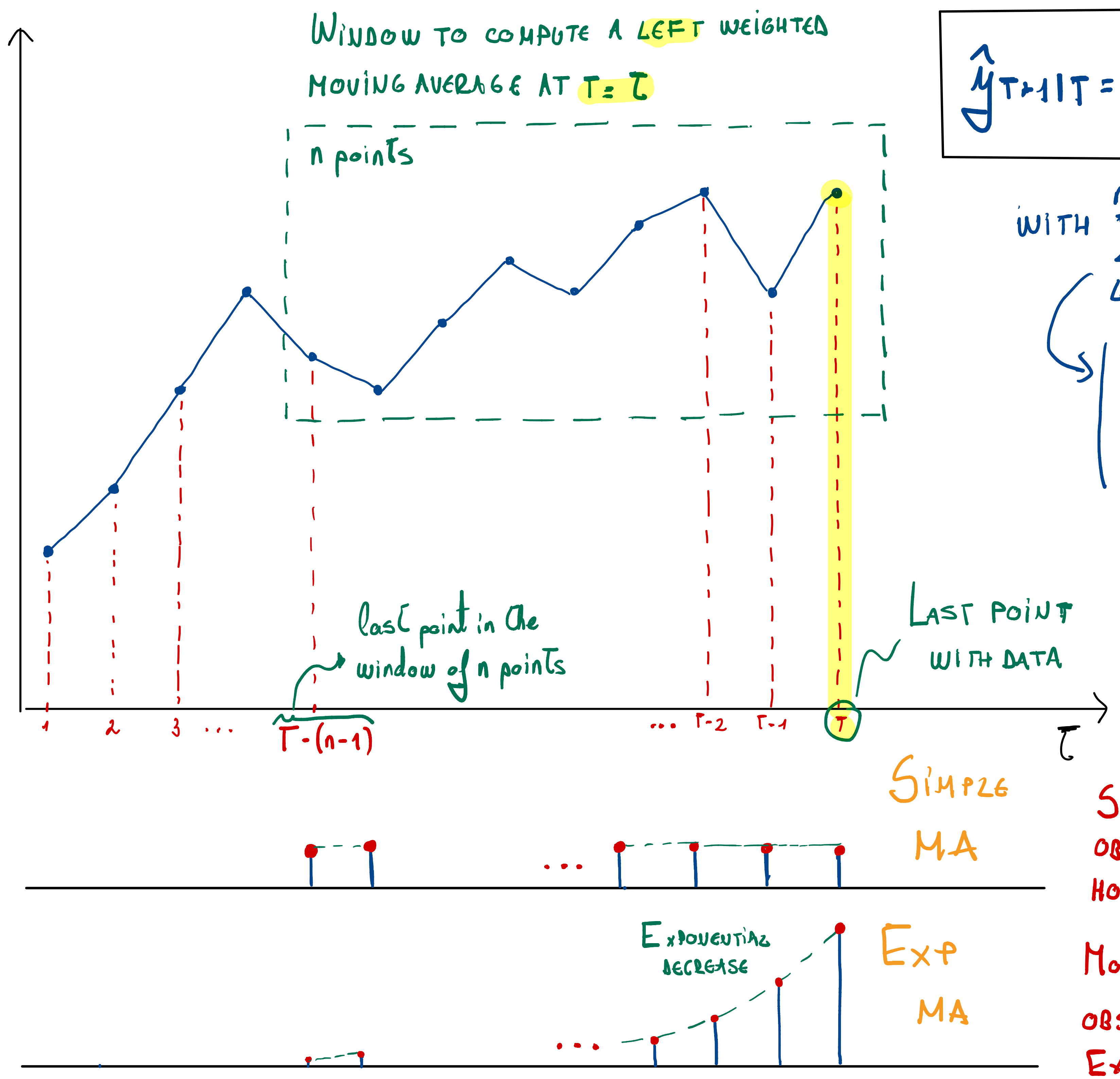
EMA_0 required (ℓ_0)

Simple Moving Average (next slide)

Calculating the EMA

Calculating the EMA requires one more observation than the SMA. Suppose that you want to use 20 days as the number of observations for the EMA. Then, you must wait until the 20th day to obtain the SMA. On the 21st day, you can then use the SMA from the previous day as the first EMA for yesterday.

Exponential vs Simple Moving Average



Left weighted moving average of n points

$$\boxed{\hat{y}_{T+1|T} = \sum_{i=0}^{n-1} w_i \cdot y_{T-i}}$$

with $\sum_{i=0}^{n-1} w_i = 1$

Each w_i is the percent contribution of y_i

Same weight to all observations, no matter how distant in the past

More weight to recent observations

Exponentially decaying weights

- **Simple (Left) Moving Average**

$$w_i = w_j = \frac{\sum_{i=0}^{n-1} y_i}{n}$$

- **Exponential Moving Average**

$$w_i = \alpha(1-\alpha)^i$$

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

Reminder: weighted AVG form of SES

Note: if all points from 1 to T are included in the window ($n = T$)

SMA → Mean Model

EMA → Simple Exponential Smoothing.

ie UNIVERSITY