Time-series Analytics

Giacomo Ziffer
Politecnico di Milano

POLITECNICO
MILANO 1863

Time Series Forecasting

Let’s change perspective again

Photo by **T B** from **Pexels**

Thinking …

- If
  - stationary **implies** predictable
- and by removing
  - the trend (a.k.a., the change in mean)
  - the seasonality
- **I obtain a stationary time-series**

Thinking …

- If
    - stationary **implies** predictable
- and by removing
    - the trend (a.k.a., the change in mean)
    - the seasonality
- **I obtain a stationary time-series**
- **Then, I can predict the time-series!**

Then, I can **predict** the time-series!

Definition

Time series forecasting **occurs when you make predictions based on historical time stamped data**.

Fact

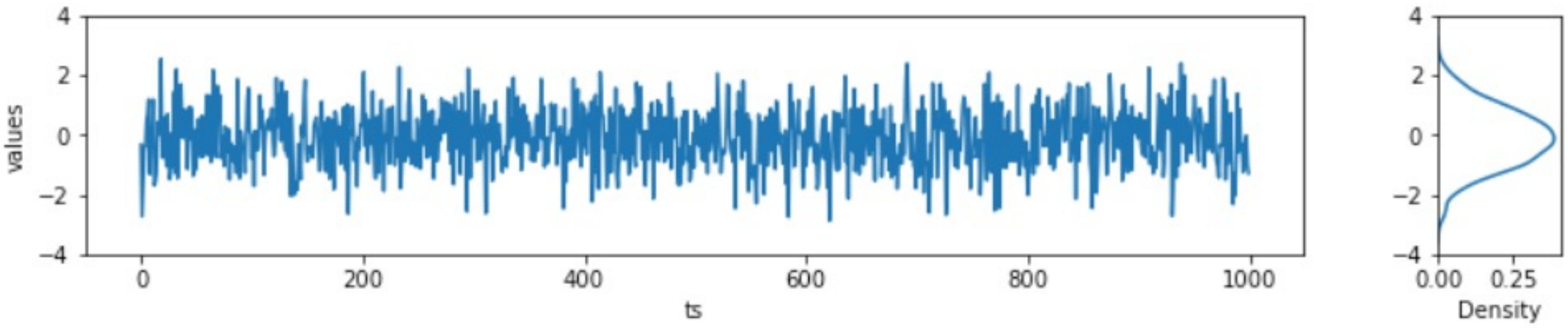
If a time series is stationary,… *… it is predictable*

Recall

Stationary

White noise: the perfect time series

- A sequence of random numbers with zero mean and finite variance



- NOTE: it is perfectly predictable
  - If you predict 0 (the mean), you minimize the error (which is proportional to the variance)

Recall

Thinking ....

- How far can we predict?
  - short-term
    - A **one-step-ahead** is a forecast for the next observation only
  - medium- and long-term (no line of demarcation)
    - A **multi-step-ahead** forecast is for 2,3,…, n steps ahead
- What can we predict?
  - Trend: long-term, the easiest to predict
  - Seasonal repetition: medium-term, does it really repeat identically?
  - Residual: short-term, the harder to predict

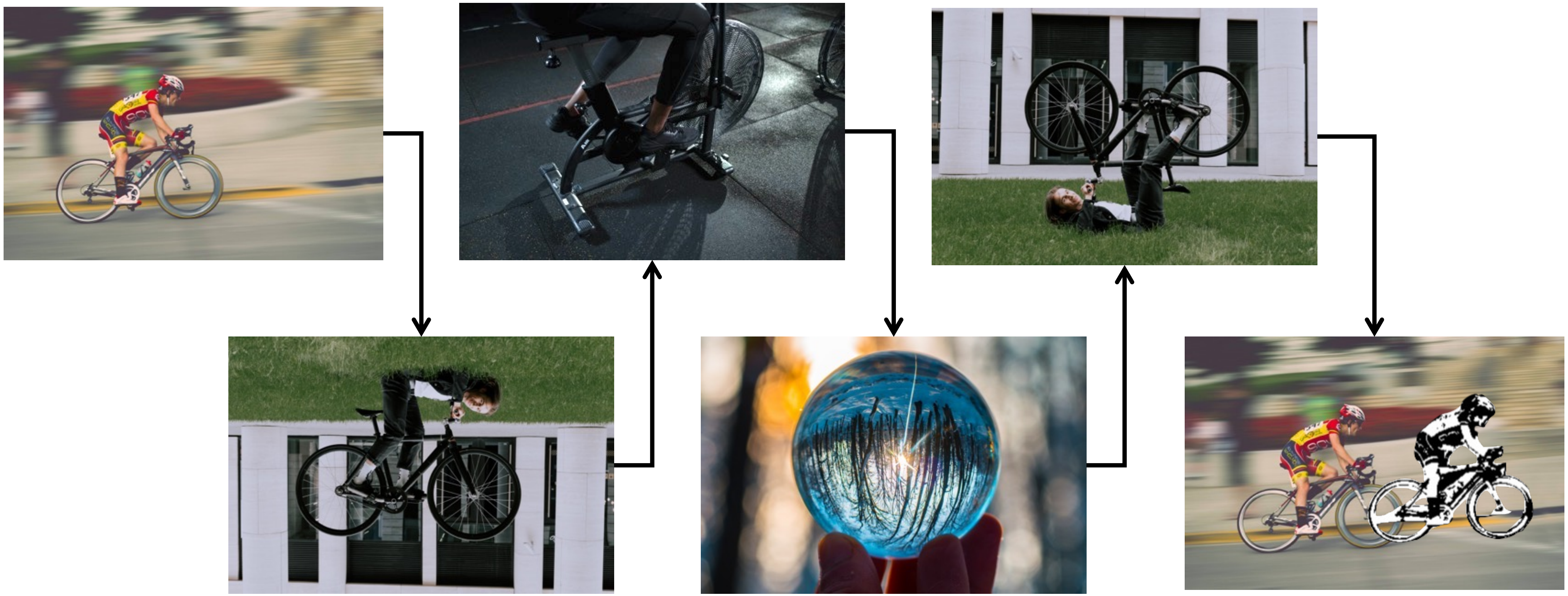
Let’s build the intuition of **forecasting**

The overall forecasting process

original time series

residual component

recomposed time series



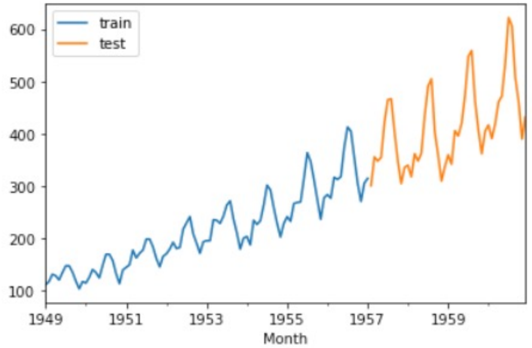
decomposed time series

prediction of the residual

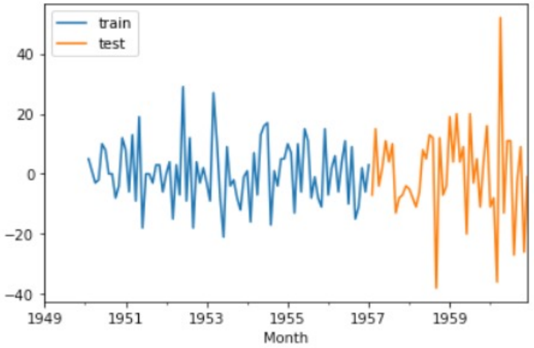
predicted time series

The overall forecasting process

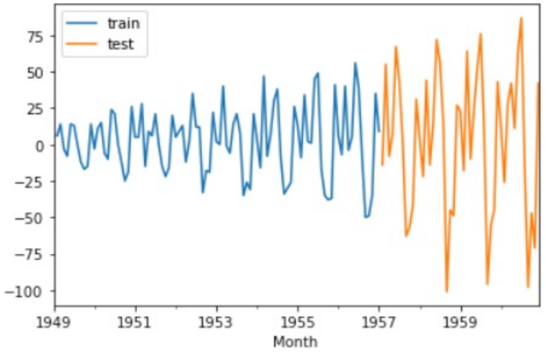
original time series



residual component



recomposed time series



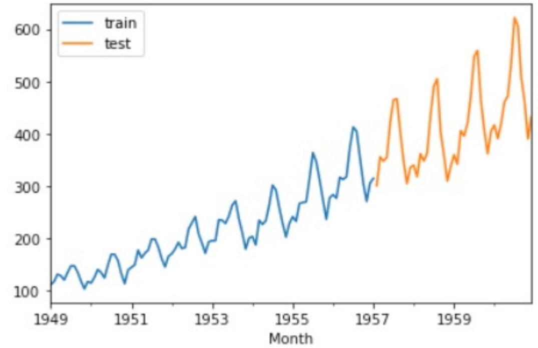
decomposed time series

prediction of the residual

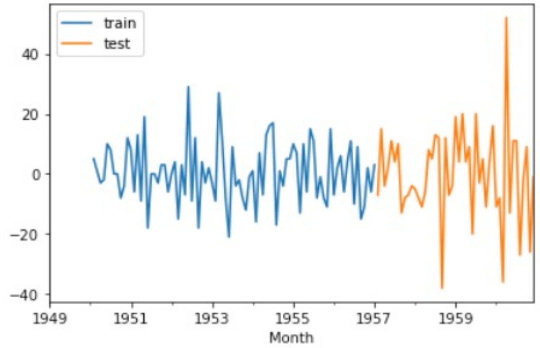
predicted time series

The overall forecasting process

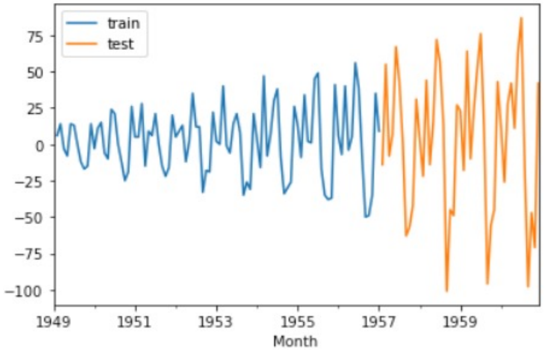
original time series



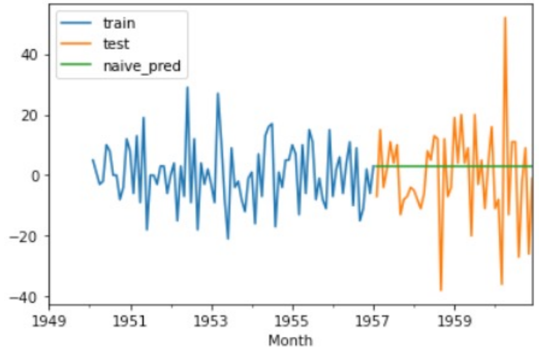
residual component



recomposed time series



decomposed time series



prediction of the residual

predicted time series

Two simplified time series models

Non-seasonal Decomposition Model with Trend

$$X_t = \boxed{m_t} \otimes Y_t$$

Time-series trend residual

components

Decomposition Model **with** Trend and **Seasonal** Components

$$X_t = \boxed{m_t \otimes s_t} \otimes Y_t$$

Time-series trend seasonal residual

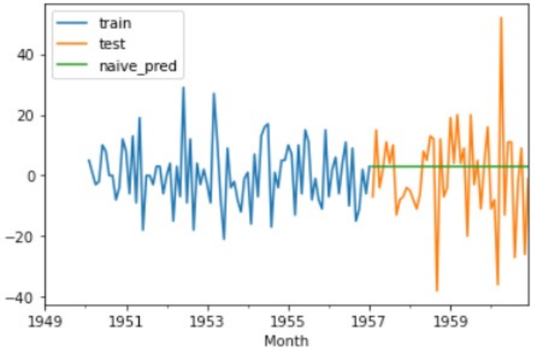
components

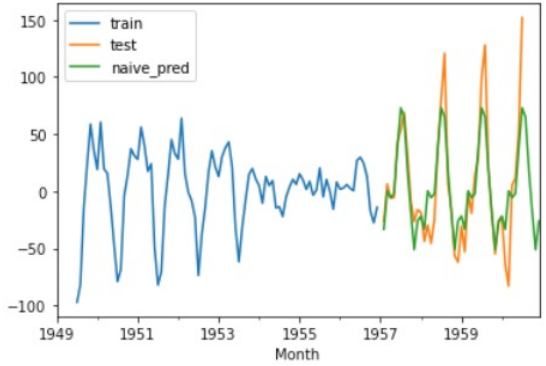
Recall

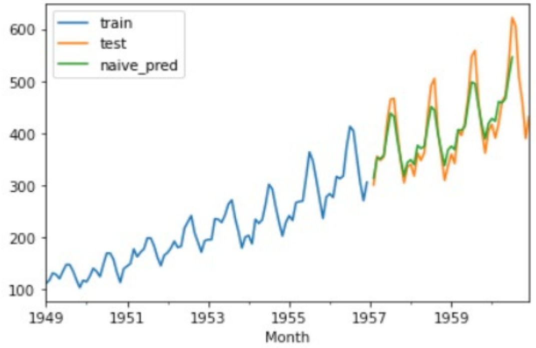
We can recompose the time series

Recomposed Time Series

Prediction of the Residual Component ⊗ **Seasonal** Component ⊗ **Trend** Component







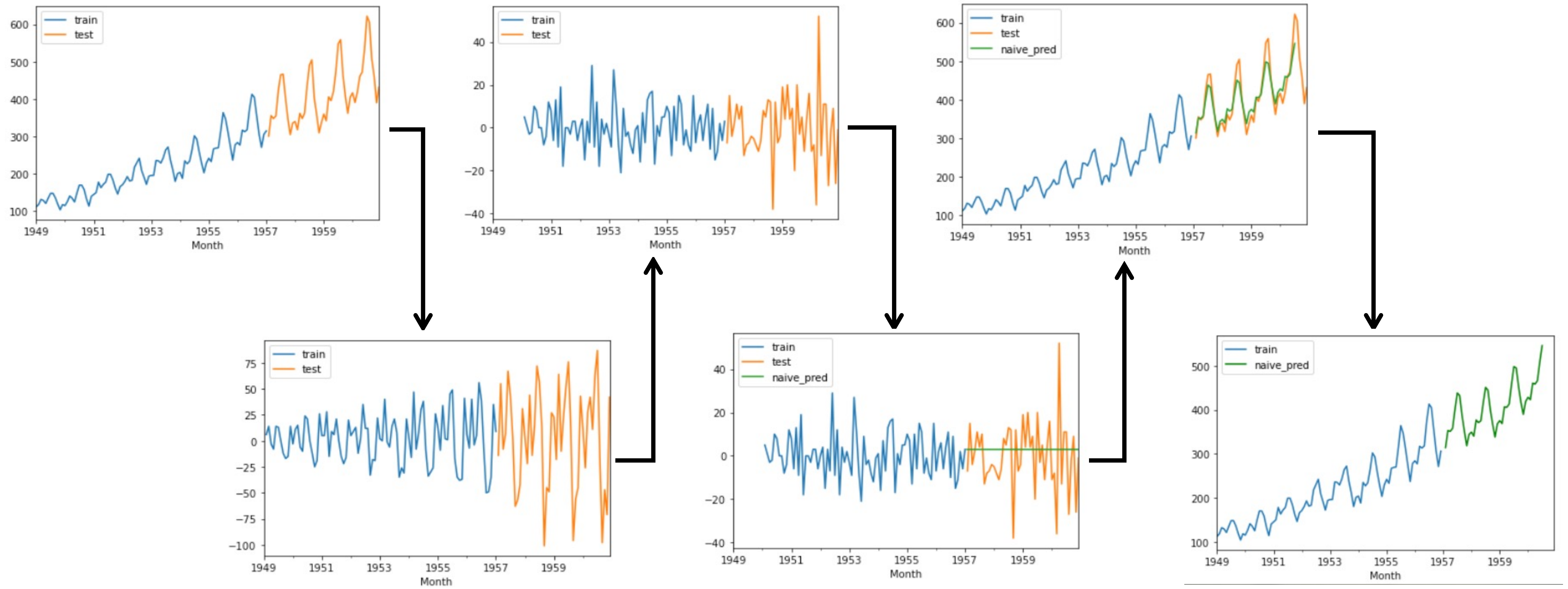
⊗ This charachter is a placeholder for various mathematical operation used to assemble the components

The overall forecasting process

original time series

residual component

recomposed time series



decomposed time series

prediction of the residual

predicted time series

Formally…

Photo by **George Becker** from **Pexels**

How do we measure accuracy?

- mean absolute error (MAE)

$$MAE = \sum_i \frac{|Y_i - \hat{Y}_i|}{n}$$

- mean absolute percent error (MAPE)

$$\mathrm{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \frac{\left|\mathrm{Y}_i - \hat{\mathrm{Y}}_i\right|}{\mathrm{Y}_i}$$

- mean square error (MSE)

$$\mathrm{MSE} = \sum_{i=1}^{n} \frac{\left(\mathrm{Y}_i - \hat{\mathrm{Y}}_i\right)^2}{n}$$

- root mean square error (RMSE)

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}}$$

Let’s focus on predicting stationary time-series

original time series

residual component

recomposed time series

decomposed time series

prediction of the residual

predicted time series

Two simplified time series models

Non-seasonal Decomposition Model with Trend

$$\boxed{X_t = m_t \otimes \boxed{Y_t}}$$

Time-series trend residual

components

Decomposition Model **with** Trend and **Seasonal** Components

$$\boxed{X_t = m_t \otimes s_t \otimes \boxed{Y_t}}$$

Time-series trend seasonal residual

components

Recall

Base line forecasting methods

“Naïve” approach

- The only important value to make predictions is the last one

$$\hat{y}_{t+1} = y_t$$

“Average” approach

- All the previous values are equally important to make predictions

$$\hat{y}_{t+1} = \frac{1}{N}\sum y_i$$

Base line forecasting methods

1. Load the airline time series
2. Make it stationary as in the previous lab
3. Predict using both the “Naïve” approach and the “Average” approach
4. Compute the accuracy metrics for both predictions
5. Which prediction is better?

Thinking …

- When shall we forget?
  - Never, i.e., we use a landmark window
    - the initial estimation date is fixed
    - the additional observations are added one by one to the estimation time span
  - After a “while”, i.e., using sliding window
    - the estimation time period is fixed
    - the start and end dates successively increase by 1

More sophisticated forecasting methods
Last-k average approach

- Only the Last-k previous values are important to make predictions

$$\hat{y}_{t+1} = \frac{1}{k}\sum_{i=0}^{k-1} y_{t-i}$$

- PROBLEM: how do we choose k?

Last-k average forecasting methods

Continuing from previous

1. Predict using “Last-k average” approach
2. Which is the best k?
3. Does the “Last-k average” approach outperform the two previous one for at least a value of k?

Thinking…

- The last-k appears to be in the middle way between the «Naive» approach and «Average» approach
- but it is hard to find a k that clearly separates
  - the data to “forget” (older than k)
  - from those to use
- Is there any other way to consider all the values and still introduce a notion of “forgetting”?
- Are all values equally important?

More sophisticated forecasting methods
Exponential Smoothing

- Middle way between Naive approach and Average approach:
  - Not only the last one (Naive)
  - Not all equally important (Average)

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

Where $0 < \alpha < 1$ is the smoothing level

- Notes:
  - **α close to 1** indicates **fast forgetting** (only the most recent values influence the forecasts)
  - **α close to 0** indicates **slow forgetting** (past observations have a large influence on forecasts)

More sophisticated forecasting methods
Why is it call exponential smoothing?

- Starting with $\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t$ we can substitute for $\hat{y}_t$ and get

$$\begin{aligned}\hat{y}_{t+1} &= \alpha y_t + (1-\alpha)[\alpha y_{t-1} + (1-\alpha)\hat{y}_{t-1}] \\ &= \alpha y_t + \alpha(1-\alpha)y_{t-1} + (1-\alpha)^2\hat{y}_{t-1}\end{aligned}$$

Continuing to substitute until the first element of the time-series leads to

$$\hat{y}_{t+1} = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \ \ldots + \alpha(1-\alpha)^{t-1}y_1 + (1-\alpha)^t l_0$$

where l_0 represents the first fitted values at time 1 (the process has to start somewhere)
Hence, the forecasts are weighted averages of past observations, with
the weights decaying exponentially as the observations get older.

More sophisticated forecasting methods

Exponential Smoothing is a streaming algorithm

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

$$= \alpha y_t + \hat{y}_t - \alpha\hat{y}_t$$

$$= \hat{y}_t + \alpha y_t - \alpha\hat{y}_t$$

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

$$= \hat{y}_t + \alpha(\mathrm{e}_t)$$

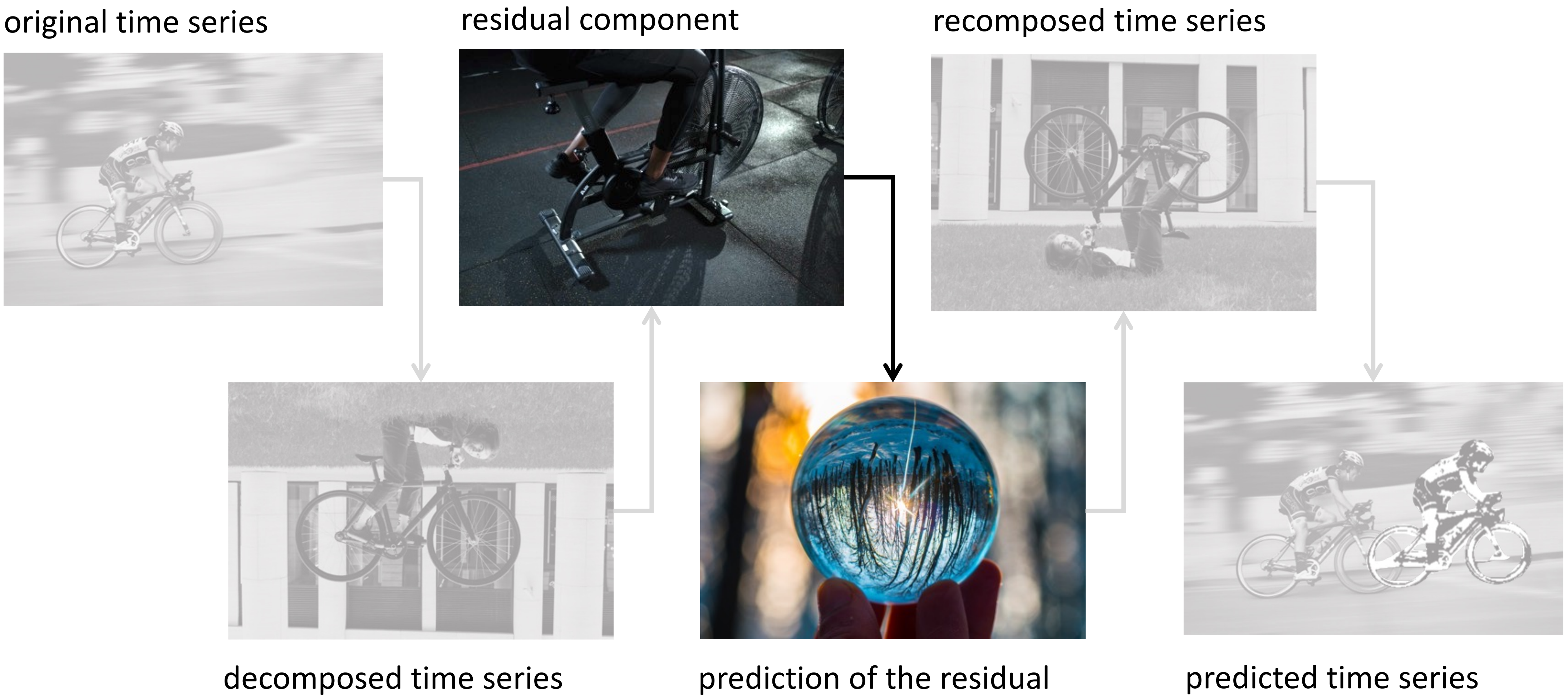
- The **next prediction** is the sum of the **current prediction** and **alpha times** the **current error** (i.e., the error of the current prediction)

Exponential smoothing forecasting methods

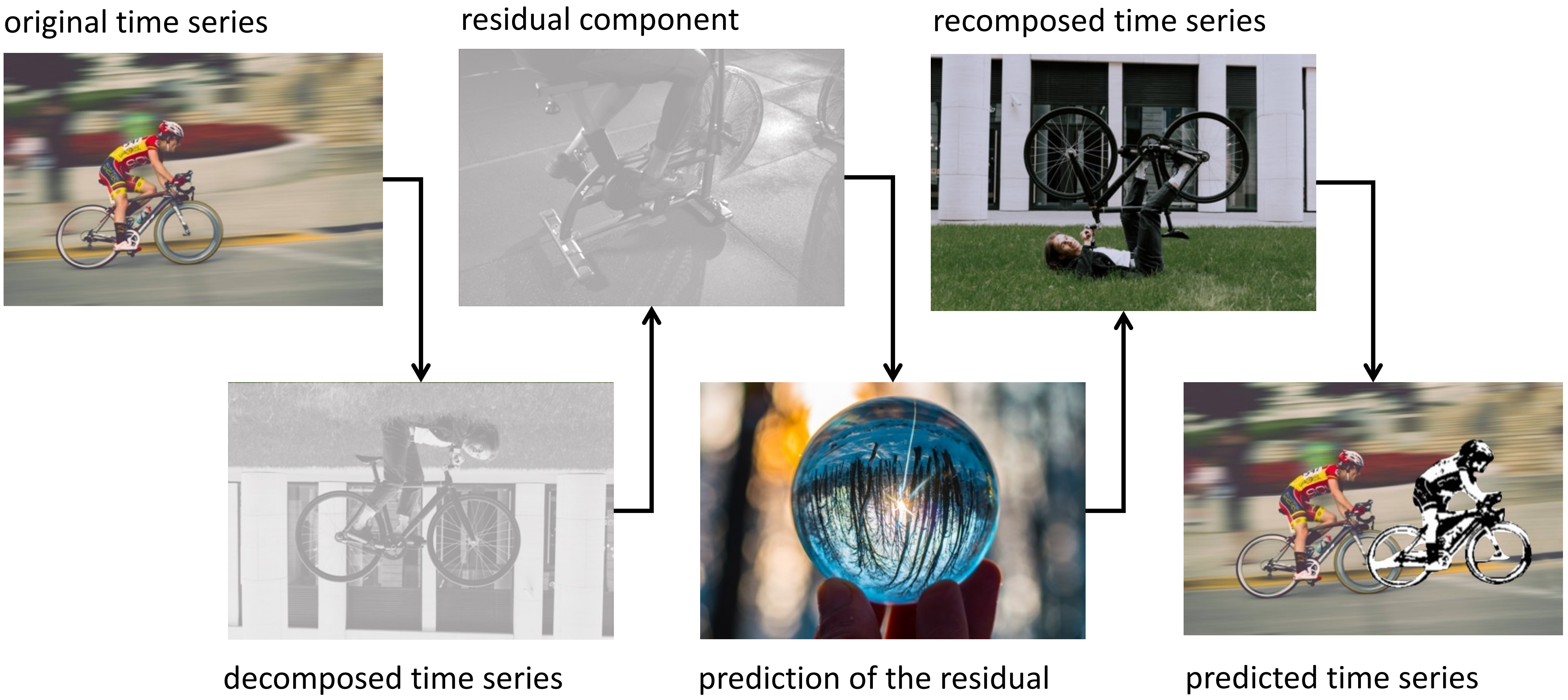
Continuing from previous

1. Predict using “exponential smoothing” approach
2. Which is the best smoothing level?
3. Does the “exponential smoothing” approach outperform the three previous ones for at least a value of the smoothing level?

(-: we can predict a stationary time series :-)

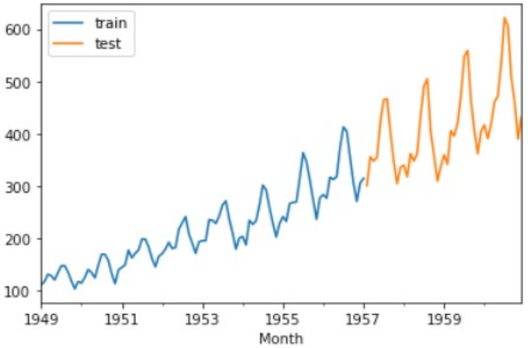


What about directly forecasting a time series with trend and seasonality?

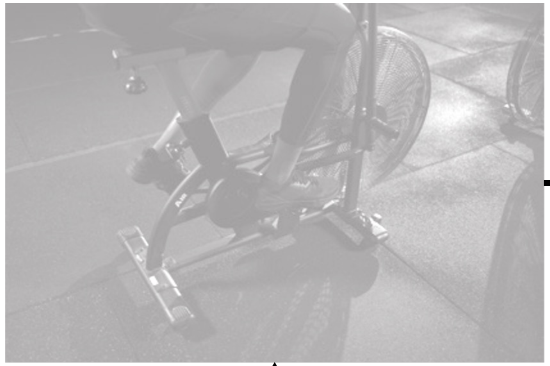


Forecasting a time series with trend using the Holt method

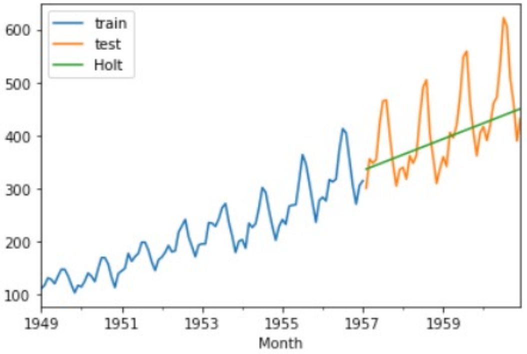
original time series



residual component

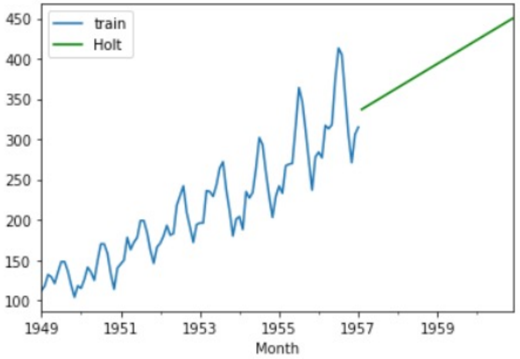


recomposed time series



decomposed time series

prediction of the residual



predicted time series

More sophisticated forecasting methods
Double Exponential Smoothing (a.k.a., Holt's linear method)

- An extension to Exponential Smoothing that **adds support for trends** using an additional smoothing factor (called β) to control the decay of the influence of the change in trend
- The method supports trends that change in different ways:
  - **additive** when the trend is **linear**
    - Forecast equation $\hat{y}_{t+\mathrm{h}} = l_t + hb_t$
    - Level equation $l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1})$
    - Trend equation $b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$
  - **multiplicative** when the trend is exponential
    - Forecast equation $\hat{y}_{t+\mathrm{h}} = l_t (b_t)^h$

More sophisticated forecasting methods
Double Exponential Smoothing (a.k.a., Holt's linear method)

- Zooming a bit more on the equations

The simple exponential smoothing equation

Adding a trend

- Forecast equation $\hat{y}_{t+\mathrm{h}} = l_t + hb_t$
- Level equation $\boxed{l_t = \alpha y_t + (1-\alpha)(l_{t-1}}\boxed{+ b_{t-1})}$
- Trend equation $b_t \;= \beta(\boxed{l_t - l_{t-1}}) + (1-\beta)b_{t-1}$

The trend gets updated based on the most recent error

More sophisticated forecasting methods
Double Exponential Smoothing (a.k.a., Holt’s linear method)

- Conclusions
  - The trend can vary adaptively over time
  - The trend smoothing parameter β controls the speed of adjusting the trend

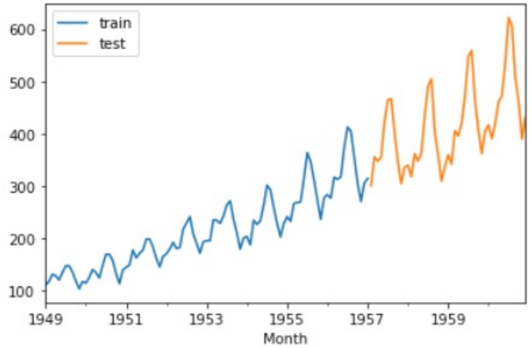
Double Exponential smoothing forecasting methods

Continuing from previous

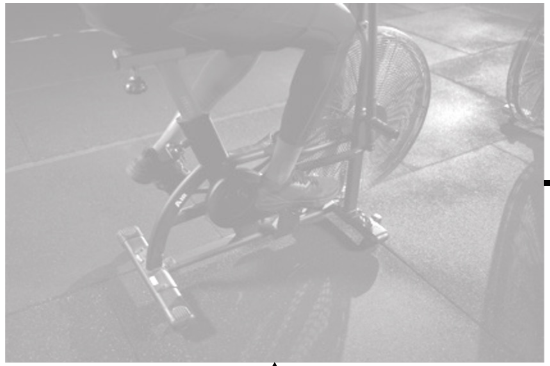
1. Pick the original data without detrending and deseasonazing them
2. Predict the test using the “double exponential smoothing” approach
3. Which are the best smoothing levels?

Forecasting a time series with trend and seasonality using the Holt-Winters method

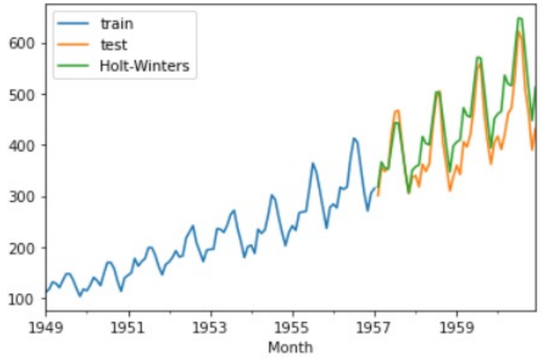
original time series



residual component

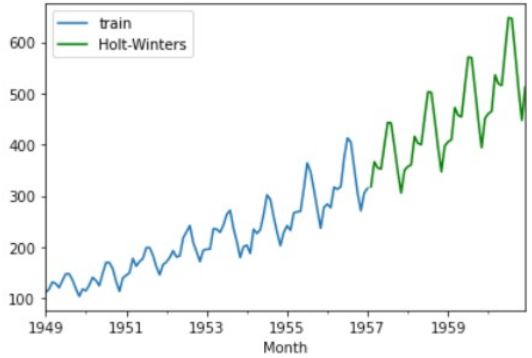


recomposed time series



decomposed time series

prediction of the residual



predicted time series

More sophisticated forecasting methods
Triple Exponential Smoothing (a.k.a. Holt-Winters method)

- An extension to Exponential Smoothing that **adds support for** seasonality using a new parameter *gamma* that controls the influence on the seasonal component
- As with the trend, the seasonality may be modelled as an **additive or multiplicative** process for a linear or exponential change in the seasonality.
- Additive equations
  - Forecast equation $\hat{y}_{t+h} = l_t + hb_t + s_{t+h-d}$
  - Level equation $l_t = \alpha(y_t - s_{t-d}) + (1-\alpha)(l_{t-1} + b_{t-1})$
  - Trend equation $b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$
  - Seasonality equation $s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-d}$

More sophisticated forecasting methods
Triple Exponential Smoothing (a.k.a. Holt-Winters method)

- Zooming a bit more on the equations

Here we deseasonalise!

- Forecast equation $\hat{y}_{t+h} = l_t + hb_t + s_{t+h-d}$
- Level equation $l_t = \alpha(\boxed{y_t - s_{t-d}}) + (1-\alpha)(l_{t-1} + b_{t-1})$
- Trend equation $b_t = \boxed{\beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}}$
- Seasonality equation $s_t = \gamma(\boxed{y_t - l_{t-1}} - b_{t-1}) + (1-\gamma)s_{t-d}$

Here we detrend!

Same as in Double Exponential Smoothing

More sophisticated forecasting methods
Triple Exponential Smoothing (a.k.a. Holt-Winters method)

- Conclusions
  - Both the trend and the seasonality can vary adaptively over time
  - The trend smoothing level β and the seasonality smoothing level γ control the speed of adjusting the trend and the seasonality, respectively
  - The only fixed parameter is the period *d* of the seasonality

Triple Exponential smoothing forecasting methods

Continuing from previous

1. Pick the original data without detrending and deseasonazing them
2. Predict the test using the "triple exponential smoothing" approach
3. Which are the best smoothing levels?
4. Does the "triple exponential smoothing" approach outperform the previous ones?

Time-series Analytics

Giacomo Ziffer
Politecnico di Milano

POLITECNICO
MILANO 1863