UNIVERSITÉ LIBRE DE BRUXELLES

Methods for multivariate and multi-step-ahead time series forecasting

Jacopo De Stefani, Ir.
Prof. Gianluca Bontempi
{jdestefa,gbonte}@ulb.ac.be

Scisports
Amersfoort, Nederlands

Friday 2nd November, 2018

About me

- Jacopo De Stefani, M.Sc^2 T.I.M.E.
- Computer Engineer (ULB, PoliMi)
  - *Software Engineering*
  - *Machine Learning*
- Master Thesis: **Spatial Allocation in Swarm Robotics**
  - Swarm Robotics = Swarm Intelligence + Collective robotics
- Teaching assistant & PhD Student
  - Time series analysis

About the MLG-ULB

Expertise

- Big data anaylsis and forecasting

About the MLG-ULB

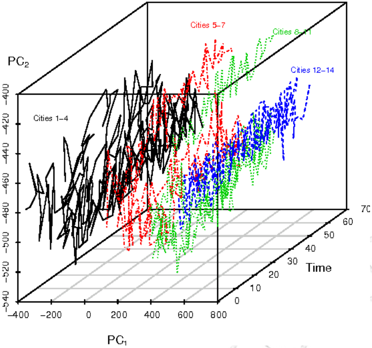
Expertise

- Big data anaylsis and forecasting
- Spatio-temporal data analysis

About the MLG-ULB

Expertise

- Big data anaylsis and forecasting
- Spatio-temporal data analysis
- Time series analysis



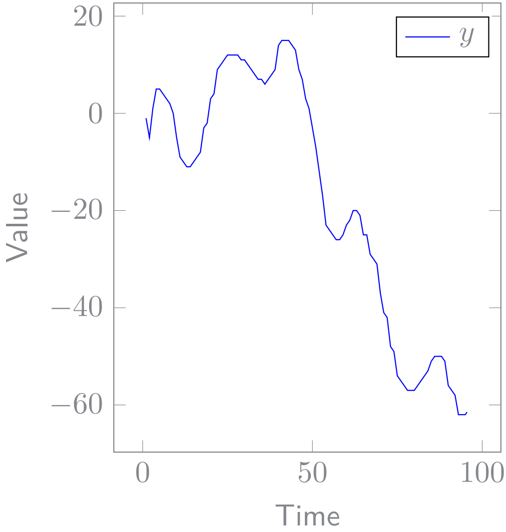
About the MLG-ULB

Expertise

- Big data anaylsis and forecasting
- Spatio-temporal data analysis
- Time series analysis
- Fraud detection

Univariate time series

Univariate time series



t	$\mathbf{y}_t$
0	-1
1	5
2	1
3	5
4	5
5	4
6	3
7	2
⋮	⋮
n	y_n

Time series forecasting - ?

Definition

Given a univariate time series $\{y_1, \cdots, y_T\}$ comprising T observations, forecast the next H observations $\{y_{T+1}, \cdots, y_{T+H}\}$ where H is the forecast horizon.

Hypotheses:

- Autoregressive model $y_t = f(y_{t-1}, \cdots, y_{t-d}) + \varepsilon_t$ with lag order d
- ε is a stochastic iid model with $\mu_\varepsilon = 0$ and $\sigma_\varepsilon^2 = \sigma^2$

Univariate time series embedding

t	y_t
0	-1
1	5
2	1
3	5
4	5
5	4
6	3
7	2
⋮	⋮
n	y_n

$1 - step$

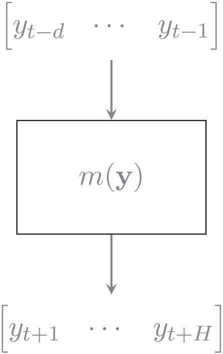
$\mathbf{y}_{t-2}$	$\mathbf{y}_{t-1}$	$\mathbf{y}_t$	$\mathbf{y}_{t+1}$
-1	5	1	5
5	1	5	5
1	5	5	4
5	5	4	3
5	4	3	2
⋮	⋮	⋮	⋮
y_{n-3}	y_{n-2}	y_{n-1}	y_n

$h - step$

$\mathbf{y}_{t-2}$	$\mathbf{y}_{t-1}$	$\mathbf{y}_t$	$\mathbf{y}_{t+h}$
-1	5	1	3
5	1	5	2
⋮	⋮	⋮	⋮
y_{n-7}	y_{n-6}	y_{n-5}	y_n

$$y_t = m(y_{t-1}, \cdots, y_{t-d}) + \varepsilon_t$$

Forecasting strategies - ?



1 TS Input
1 TS Output

Recursive strategy

- Single model recursively employed H times.

Direct strategy

- H models, one for each forecasting horizon.

Rectify strategy

1. Recursive Linear AR(p)
2. Direct KNN on residuals $\sim$ "single strong rectification"

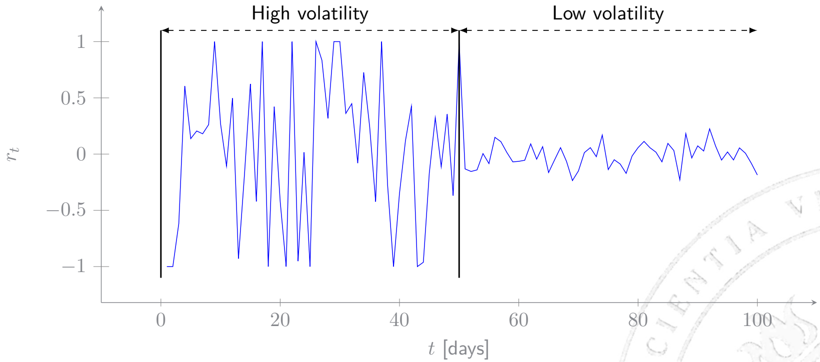
Boost strategy

1. Recursive Linear AR(p)
2. Direct Boosting of bivariate learners $\sim$ "several weak rectifications"

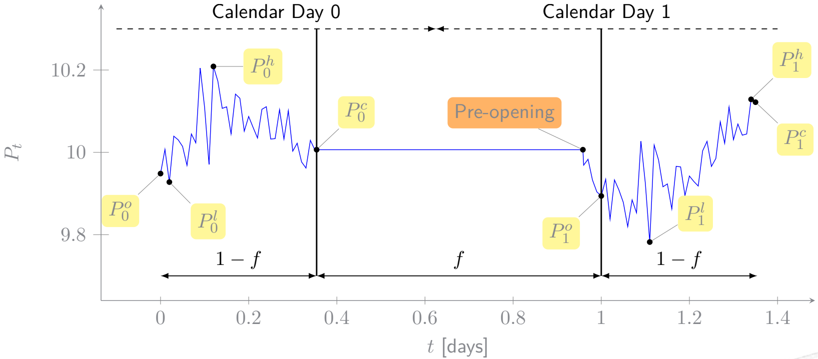
What is volatility?

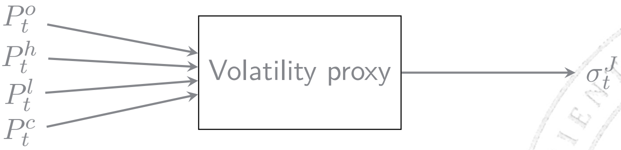
Definition

Volatility is a statistical measure of the dispersion of returns for a given security or market index.



A closer look on data - Volatility measures





Proposed model - MIDAS 2017 ?

$$[\sigma^J_{t-d} \quad \cdots \quad \sigma^J_{t-1}]$$
$$[\sigma^X_{t-d} \quad \cdots \quad \sigma^X_{t-1}]$$
$$\downarrow$$
$$\boxed{m(\sigma^{\mathbf{J}}, \sigma^{\mathbf{X}})}$$
$$\downarrow$$
$$[\hat{\sigma}^J_t \quad \cdots \quad \hat{\sigma}^J_{t+H}]$$

2 TS Input
1 TS Output

Idea:
Combine volatility proxies to improve forecasting performance
Data: Volatility proxies σ^X, σ^J from CAC40:

- $\sigma_{\mathbf{i}}$ **family - ?**
- **GARCH (1,1) model - ?**
- **Sample standard deviation**

Models $m(\sigma^{\mathbf{J}}, \sigma^{\mathbf{X}})$**:**

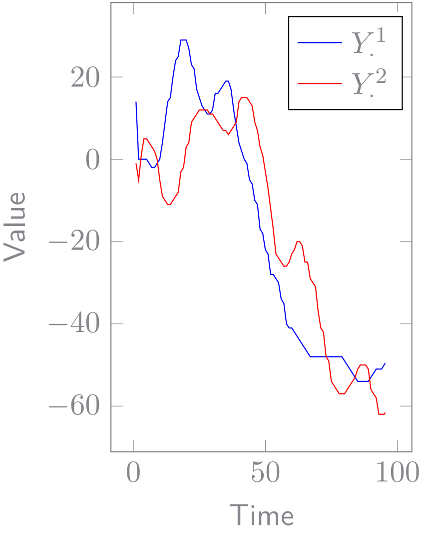
- Feedforward Neural Networks
- k-Nearest Neighbours
- Support Vector Regression

Conclusions - MIDAS 2017 ?

- Correlation clustering among proxies belonging to the same family, i.e. σ_t^i and $\sigma_t^{SD,n}$.
- All ML methods outperform the reference GARCH method, both in the single input and the multiple input configuration.
- Only the addition of an external regressor, and for $h > 8$ bring a statistically significant improvement (paired t-test, pv=0.05).
- No model appear to clearly outperform all the others on every horizons, but generally SVR performs better than ANN and k-NN.

Multivariate time series

What about multivariate time series?



t	$\mathbf{Y}^1_t$	$\mathbf{Y}^2_t$
1	14	-1
2	0	-5
3	0	1
4	0	5
5	0	5
6	-1	4
7	-2	3
⋮	⋮	⋮
n	Y^1_n	Y^2_n

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Multivariate time series embedding

t	$\mathbf{Y}^1_t$	$\mathbf{Y}^2_t$
1	14	-1
2	0	-5
3	0	1
4	0	5
5	0	5
6	-1	4
7	-2	3
8	-2	2
$\vdots$	$\vdots$	$\vdots$
n	$\mathbf{Y}^1_n$	$\mathbf{Y}^2_n$

$1-step$

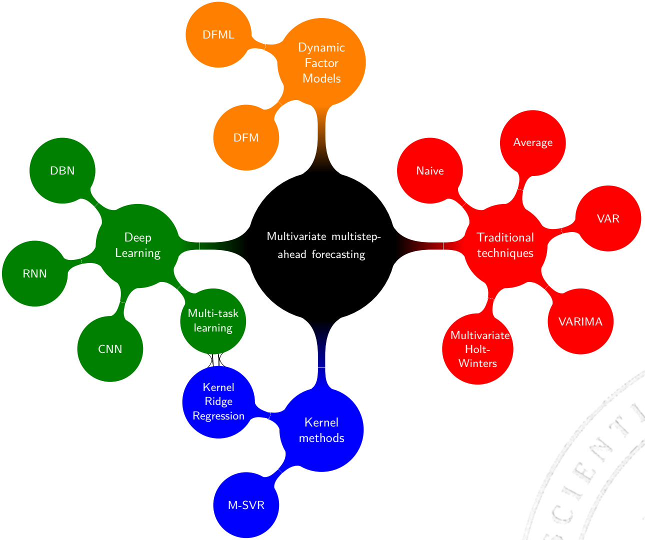
$\mathbf{Y}^1_{t-1}$	$\mathbf{Y}^1_t$	$\mathbf{Y}^2_{t-1}$	$\mathbf{Y}^2_t$	$\mathbf{Y}^1_{t+1}$	$\mathbf{Y}^2_{t+1}$
14	0	-1	-5	0	1
0	0	5	1	0	5
0	0	1	5	0	5
0	0	5	5	-1	4
0	-1	5	4	-2	3
-1	-2	4	3	-2	2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\mathbf{Y}^1_{n-2}$	$\mathbf{Y}^1_{n-1}$	$\mathbf{Y}^2_{n-2}$	$\mathbf{Y}^2_{n-1}$	$\mathbf{Y}^1_n$	$\mathbf{Y}^2_n$

$h-step$

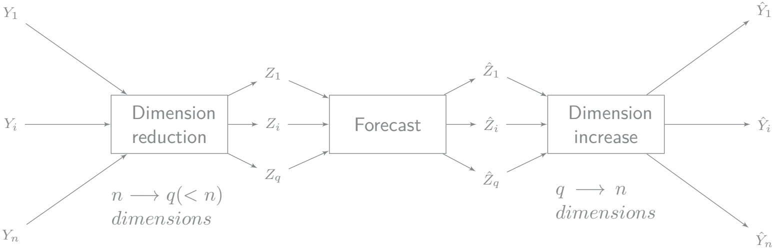
$\mathbf{Y}^1_{t-1}$	$\mathbf{Y}^1_t$	$\mathbf{Y}^2_{t-1}$	$\mathbf{Y}^2_t$	$\mathbf{Y}^1_{t+h}$	$\mathbf{Y}^2_{t+h}$
14	0	-1	-5	-1	4
0	0	5	1	-2	3
0	0	1	5	-2	2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\mathbf{Y}^1_{n-6}$	$\mathbf{Y}^1_{n-5}$	$\mathbf{Y}^2_{n-6}$	$\mathbf{Y}^2_{n-5}$	$\mathbf{Y}^1_n$	$\mathbf{Y}^2_n$

$$y_t = m(y_{t-1}, \cdots, y_{t-d}) + \varepsilon_t \mapsto \mathbf{Y}_t = M(\mathbf{Y}_{t-1}, \cdots, \mathbf{Y}_{t-d}) + \mathbf{W}$$

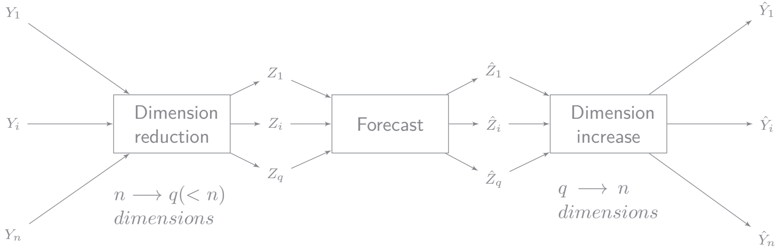
Models for multivariate forecasting - Internal Report



Proposed approach: DFML - ??



Proposed approach: DFML - ??



	Dimensionality reduction	Forecast
DFM	PCA	VAR
$DFML_{PC}$	PCA	Lazy-learning
$DFML_{A}$	Autoencoder	Lazy-learning
$DFML'_{PC}$	Optimized PCA	Optimized Lazy-learning
$DFML_{IPC}$	Online PCA	Lazy-learning

Proposed model - MIDAS 2018 De Stefani et al. [2018]

$$\begin{bmatrix} \sigma^1_{t-1} & \cdots & \sigma^n_{t-1} \\ \vdots & \ddots & \vdots \\ \sigma^1_{t-d} & \cdots & \sigma^n_{t-d} \end{bmatrix}$$

$$\downarrow$$

$$\boxed{m(\sigma^{\mathbf{1}}, \cdots, \sigma^{\mathbf{n}})}$$

$$\downarrow$$

$$\begin{bmatrix} \sigma^1_{t+H} & \cdots & \sigma^n_{t+H} \\ \vdots & \ddots & \vdots \\ \sigma^1_{t+1} & \cdots & \sigma^n_{t+1} \end{bmatrix}$$

n **TS Input**
n **TS Output**

Data:

- Cryptocurrencies (291 series $\times$ 495 points)
- CAC40 (40 series $\times$ 1645 points)

Volatility proxies from:

- **Price based** - $\sigma_{\mathbf{i}}$ **family** - **?**
- **Return based** - Sample standard deviation

Models:

- DFML
- Recurrent Neural Network (w and w/o LSTM)
- Partial Least Squares
- Naive
- Univariate

Conclusions - MIDAS 2018 De Stefani et al. [2018]

- From the forecasting point of view, the cryptocurrency dataset yields lower errors than the CAC40 one.
- A greater smoothing factor w (as in the $\sigma^{SD,w}$ proxies family), consistently reduce the forecasting power of all the considered methods.
- DFML is effective, especially for high-dimensional noisy series (i.e. Cryptocurrencies dataset) and for a longer horizons.
- For the considered datasets, a linear dimensionality reduction technique (PCA) consistently outperforms the non-linear correspondent.

Ideas for brainstorming

- **My goals**:
  - Time dependent data (ideally numerical time series)
  - Large scale data
  - Multivariate and multistep ahead forecasting
- **Points of interest in Scisport**:
  - Insights data
  - Univariate / Multivariate Potential forecasting

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Thank you for your attention! Any questions/comments?

Bibliography I

Appendix

Volatility proxies (1) - ?

- Closing prices

$$\hat{\sigma}_0(t) = \left[\ln\left(\frac{P_{t+1}^{(c)}}{P_t^{(c)}}\right)\right]^2 = r_t^2 \tag{1}$$

- Opening/Closing prices

$$\hat{\sigma}_1(t) = \underbrace{\frac{1}{2f}\cdot\left[\ln\left(\frac{P_{t+1}^{(o)}}{P_t^{(c)}}\right)\right]^2}_{\text{Nightly volatility}} + \underbrace{\frac{1}{2(1-f)}\cdot\left[\ln\left(\frac{P_t^{(c)}}{P_t^{(o)}}\right)\right]^2}_{\text{Intraday volatility}} \tag{2}$$

- OHLC prices

$$\hat{\sigma}_2(t) = \frac{1}{2\ln 4}\cdot\left[\ln\left(\frac{P_t^{(h)}}{P_t^{(l)}}\right)\right]^2 \tag{3}$$

$$\hat{\sigma}_3(t) = \underbrace{\frac{a}{f}\cdot\left[\ln\left(\frac{P_{t+1}^{(o)}}{P_t^{(c)}}\right)\right]^2}_{\text{Nightly volatility}} + \underbrace{\frac{1-a}{1-f}\cdot\hat{\sigma}_2(t)}_{\text{Intraday volatility}} \tag{4}$$

Volatility proxies (2) - ?

- **OHLC prices**

$$u = \ln\left(\frac{P_t^{(h)}}{P_t^{(o)}}\right) \qquad d = \ln\left(\frac{P_t^{(l)}}{P_t^{(o)}}\right) \qquad c = \ln\left(\frac{P_t^{(c)}}{P_t^{(o)}}\right) \tag{5}$$

$$\hat{\sigma}_4(t) = 0.511(u-d)^2 - 0.019[c(u+d) - 2ud] - 0.383c^2 \tag{6}$$

$$\hat{\sigma}_5(t) = 0.511(u-d)^2 - (2\ln 2 - 1)c^2 \tag{7}$$

$$\hat{\sigma}_6(t) = \underbrace{\frac{a}{f} \cdot \log\left(\frac{P_{t+1}^{(o)}}{P_t^{(c)}}\right)^2}_{\text{Nightly volatility}} + \underbrace{\frac{1-a}{1-f} \cdot \hat{\sigma}_4(t)}_{\text{Intraday volatility}} \tag{8}$$

Volatility proxies (3)

- **GARCH (1,1) model - ?**

$$\sigma_t^G = \sqrt{\omega + \sum_{j=1}^{p} \beta_j (\sigma_{t-j}^G)^2 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2}$$

where $\varepsilon_{t-i} \sim \mathcal{N}(0,1)$, with the coefficients $\omega, \alpha_i, \beta_j$ fitted according to **?**.

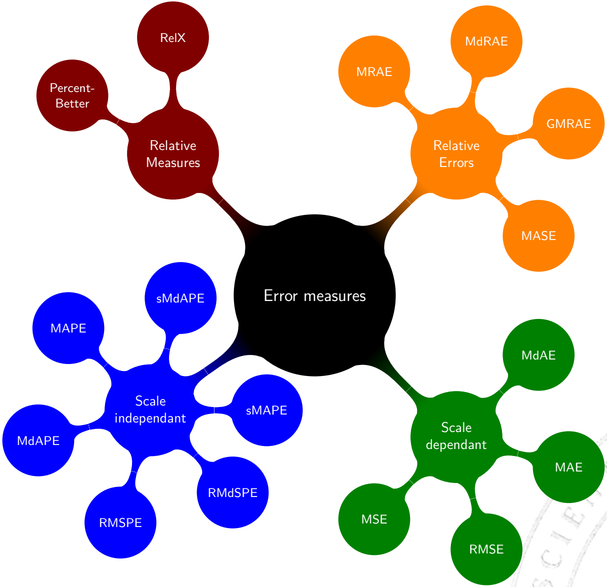
- **Sample standard deviation**

$$\sigma_t^{SD,n} = \sqrt{\frac{1}{n-1} \sum_{i=0}^{n-1} (r_{t-i} - \bar{r})^2}$$

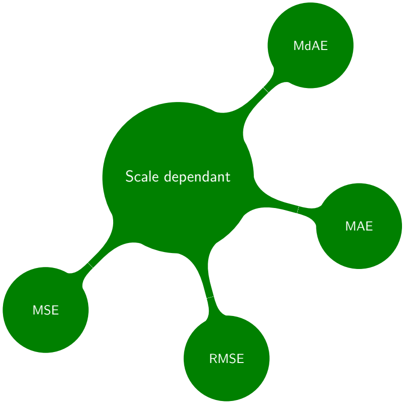
where

$$r_t = \ln\left(\frac{P_t^{(c)}}{P_{t-1}^{(c)}}\right) \qquad \bar{r_n} = \frac{1}{n} \sum_{j=t-n}^{t} r_j$$

? - Error measures



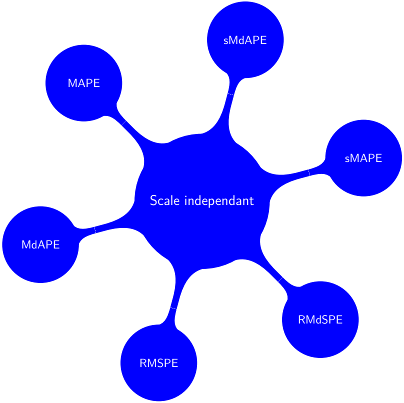
? - Scale dependant



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

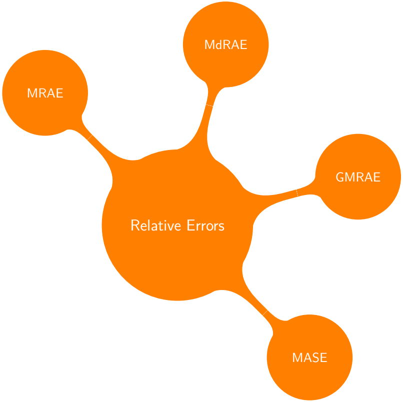
- **MSE** : $\frac{1}{n}\sum_{t=0}^{n}(y_t - \hat{y}_t)^2$
- **RMSE** : $\sqrt{\frac{1}{n}\sum_{t=0}^{n}(y_t - \hat{y}_t)^2}$
- **MAE** : $\frac{1}{n}\sum_{t=0}^{n}|y_t - \hat{y}_t|$
- **MdAE** : $Md_{t\in\{1\cdots n\}}(|y_t - \hat{y}_t|)$

? - Scale independant



- **MAPE** :
  $\frac{1}{n}\sum_{t=0}^{n} \mid 100 \cdot \frac{y_t-\hat{y}_t}{y_t} \mid$
- **MdAPE** :
  $Md_{t\in\{1\cdots n\}}(\mid 100 \cdot \frac{y_t-\hat{y}_t}{y_t} \mid)$
- **RMSPE** :
  $\sqrt{\frac{1}{n}\sum_{t=0}^{n}(100 \cdot \frac{y_t-\hat{y}_t}{y_t})^2}$
- **RMdSPE** :
  $\sqrt{Md_{t\in\{1\cdots n\}}((100 \cdot \frac{y_t-\hat{y}_t}{y_t})^2)}$
- **sMAPE** :
  $\frac{1}{n}\sum_{t=0}^{n} 200 \cdot \frac{|y_t-\hat{y}_t|}{y_t+\hat{y}_t}$
- **sMdAPE** :
  $Md_{t\in\{1\cdots n\}}(200 \cdot \frac{|y_t-\hat{y}_t|}{y_t+\hat{y}_t})$

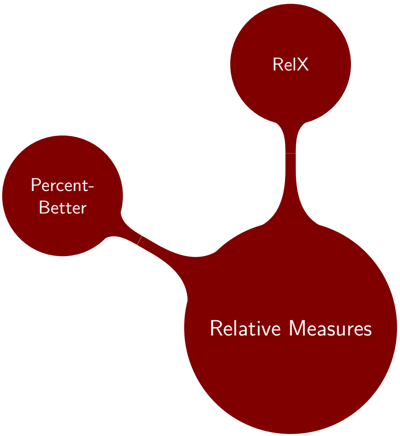
? - Relative errors



$$r_t = \frac{e_t}{e_t^*}$$

- **MRAE** : $\frac{1}{n}\sum_{t=0}^{n} \mid r_t \mid$
- **MdRAE** : $Md_{t\in\{1\cdots n\}}(\mid r_t \mid)$
- **GMRAE** : $\sqrt[n]{\frac{1}{n}\prod t = 0^n \mid r_t \mid}$
- **MASE** : $\frac{1}{T}\sum_{t=1}^{T}\left(\frac{|e_t|}{\frac{1}{T-1}\sum_{i=2}^{T}|Y_i - Y_{i-1}|}\right)$

? - Relative measures



- **RelX** : $\frac{X}{X_{\text{bench}}}$
- **Percent Better** : $PB(X) = 100 \cdot \frac{1}{n} \sum_{\text{forecasts}} I(X < X_b)$

where

- X: Error measure of the analyzed method
- X_b: Error measure of the benchmark