Machine Learning for Multi-step Ahead Forecasting of Volatility Proxies

Jacopo De Stefani, Ir. - jdestefa@ulb.ac.be
Prof. Gianluca Bontempi - gbonte@ulb.ac.be
Olivier Caelen, PhD - olivier.caelen@worldline.com
Dalila Hattab, PhD - dalila.hattab@equensworldline.com

INESC-ID - THOR: algoriTHms and cOmputer progRamming seminar

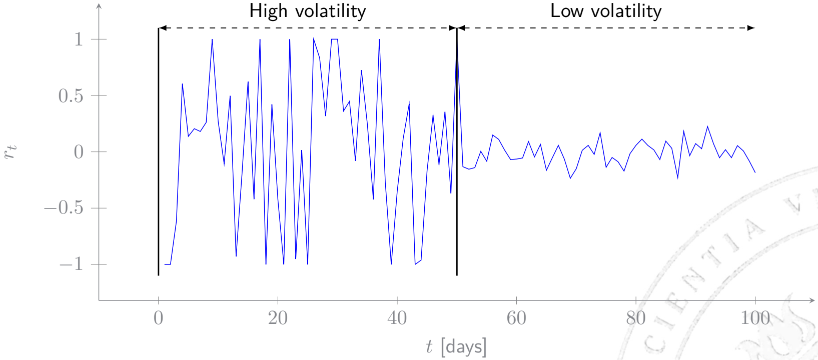
Lisbon, Thursday 2nd November, 2017

Problem overview

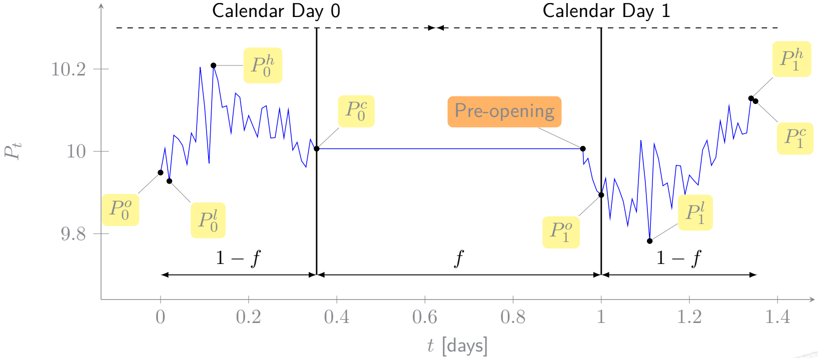
What is volatility?

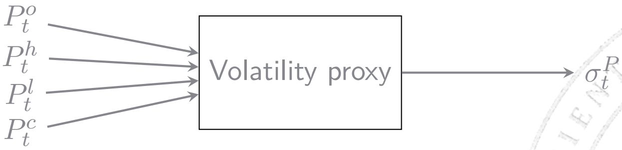
Definition

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

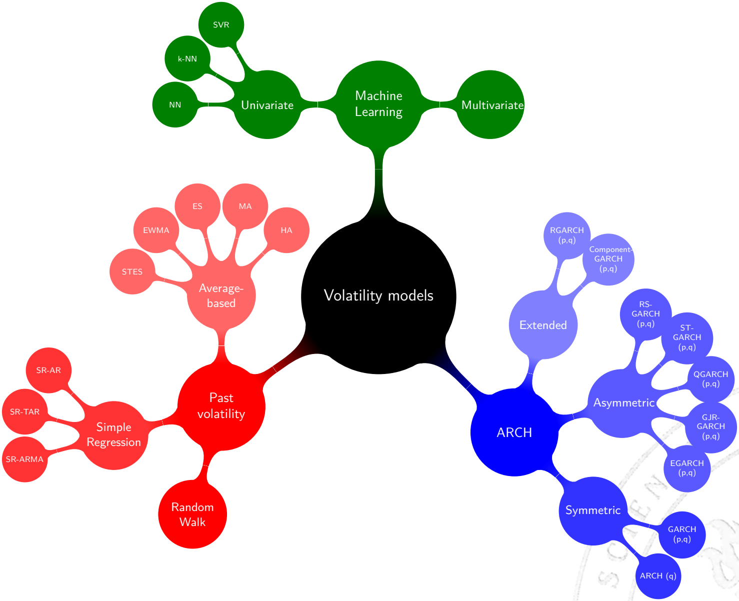


A closer look on data - Volatility proxies

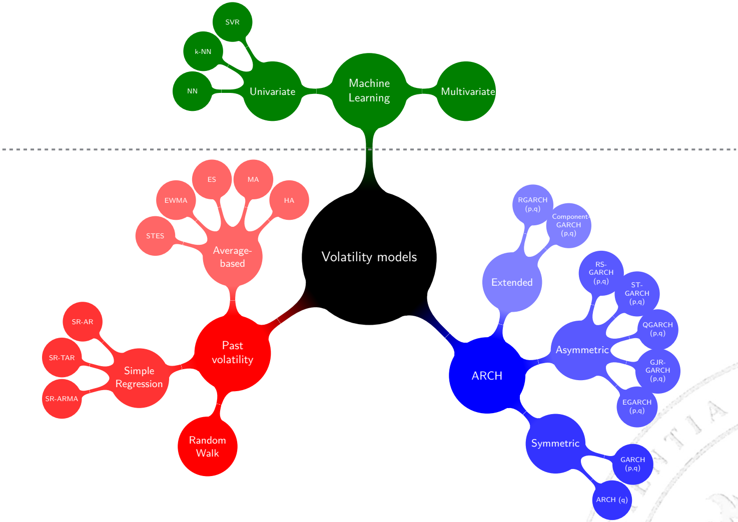




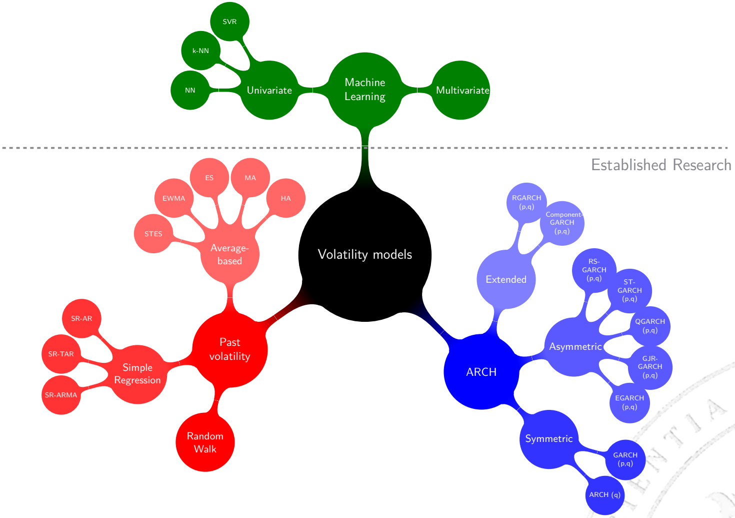
Models for volatility



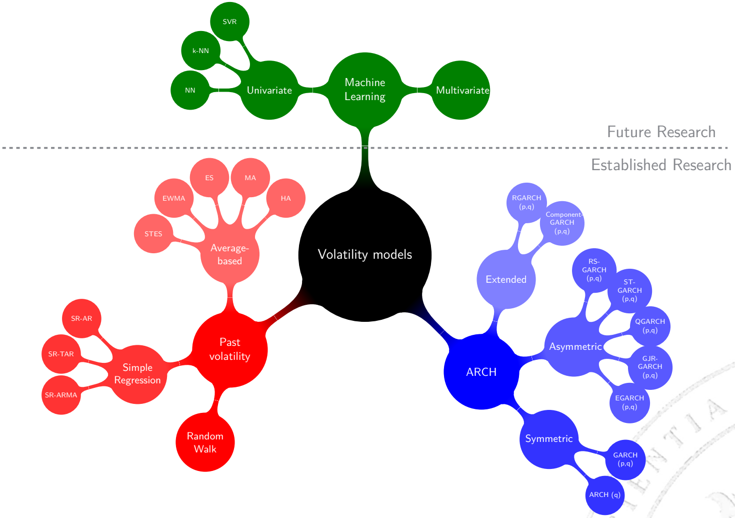
Models for volatility



Models for volatility



Models for volatility



ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Multistep ahead TS forecasting - ?

Definition

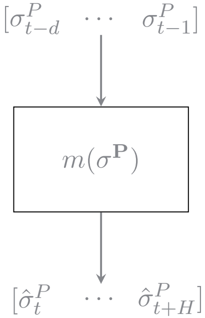
Given a univariate time series $\{y_1, \cdots, y_T\}$ comprising T observations, forecast the next H observations $\{y_{T+1}, \cdots, y_{T+H}\}$ where H is the forecast horizon.

Hypotheses:

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

Multistep ahead forecasting for volatility

State-of-the-art
NAR

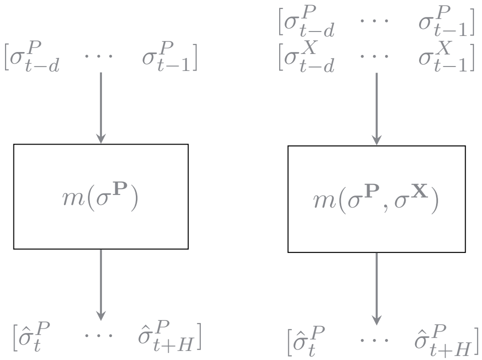


1 Input
1 Output

Multistep ahead forecasting for volatility

State-of-the-art
NAR

Proposed model
NARX



1 Input
1 Output

2 inputs
1 output

Multistep ahead forecasting for volatility

State-of-the-art: NAR

$[\sigma^P_{t-d} \quad \cdots \quad \sigma^P_{t-1}]$

↓

$m(\sigma^{\mathbf{P}})$

↓

$[\hat{\sigma}^P_{t} \quad \cdots \quad \hat{\sigma}^P_{t+H}]$

1 Input
1 Output

Proposed model: NARX

$[\sigma^P_{t-d} \quad \cdots \quad \sigma^P_{t-1}]$
$[\sigma^X_{t-d} \quad \cdots \quad \sigma^X_{t-1}]$

↓

$m(\sigma^{\mathbf{P}}, \sigma^{\mathbf{X}})$

↓

$[\hat{\sigma}^P_{t} \quad \cdots \quad \hat{\sigma}^P_{t+H}]$

2 inputs
1 output

Future work

$[\sigma^P_{t-d} \quad \cdots \quad \sigma^P_{t-1}]$
$[\cdots \quad \cdots \quad \cdots]$
$[\sigma^{X_M}_{t-d} \quad \cdots \quad \sigma^{X_M}_{t-1}]$

↓

$m(\sigma^{\mathbf{P}}, \cdots, \sigma^{\mathbf{X_M}})$

↓

$[\hat{\sigma}^P_{t} \quad \cdots \quad \hat{\sigma}^P_{t+H}]$
$[\cdots \quad \cdots \quad \cdots]$
$[\hat{\sigma}^{X_M}_{t} \quad \cdots \quad \hat{\sigma}^{X_M}_{t+H}]$

$\mathrm{M}+1$ inputs
$\mathrm{M}+1$ outputs

ULB

Multistep ahead forecasting for volatility

UNIVERSITÉ LIBRE DE BRUXELLES

Direct method

- A single model f^h for each horizon h.
- Forecast at h step is made using h^{th} model.
- Dataset examples $(d = 3,\ h = 3)$:

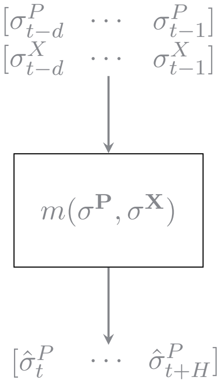
Direct NAR

x			y
σ_3^P	σ_2^P	σ_1^P	σ_5^P
σ_4^P	σ_3^P	σ_2^P	σ_6^P
...	...	...	...
σ_{T-5}^P	σ_{T-6}^P	σ_{T-7}^P	σ_{T-2}^P

Direct NARX

x						y
σ_3^P	σ_2^P	σ_1^P	σ_3^X	σ_2^X	σ_1^X	σ_5^P
σ_4^P	σ_3^P	σ_2^P	σ_4^X	σ_3^X	σ_2^X	σ_6^P
...	...	...	...	...	...	...
σ_{T-5}^P	σ_{T-6}^P	σ_{T-7}^P	σ_{T-5}^X	σ_{T-6}^X	σ_{T-7}^X	σ_{T-2}^P

Experimental setup



2 TS Input
1 TS Output

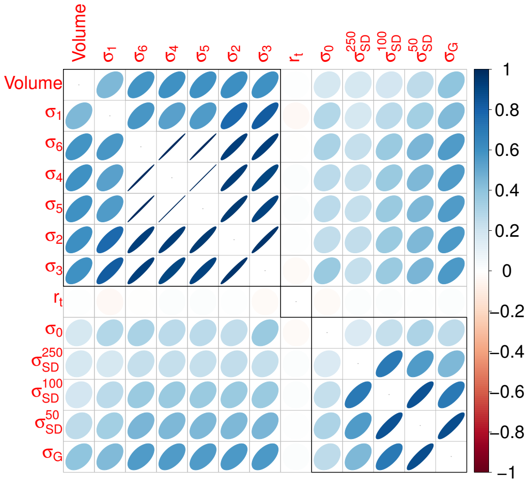
Data: Volatility proxies σ^X, σ^P from CAC40:

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

Models:

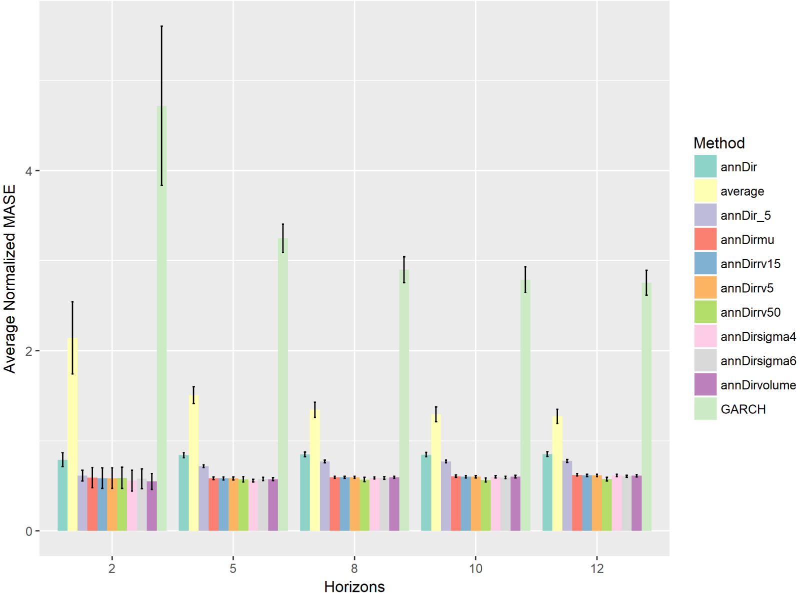
- Feedforward Neural Networks (NAR,NARX)
- k-Nearest Neighbours (NAR,NARX)
- Support Vector Regression (NAR,NARX)
- Naive (w/o σ^X)
- GARCH(1,1) (w/o σ^X)
- Average (w/o σ^X)

Correlation meta-analysis (cf. ?)



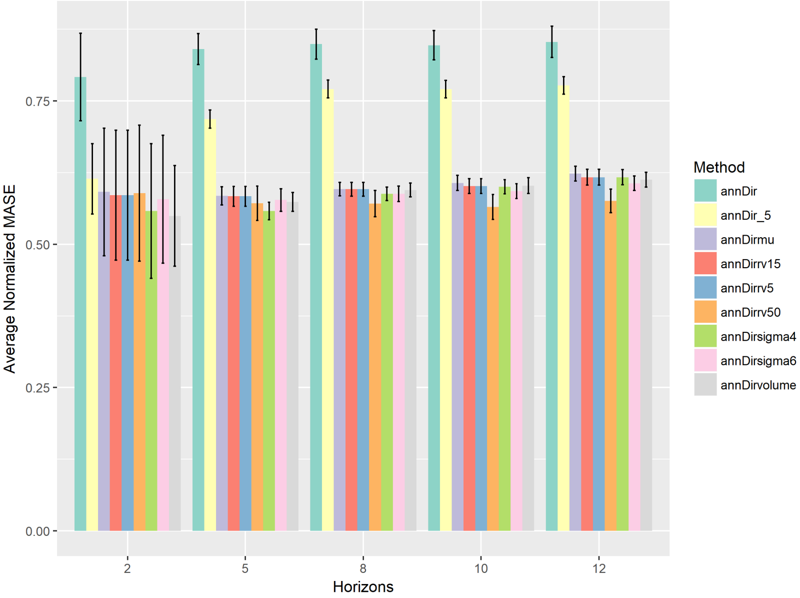
- 40 time series (CAC40)
- Time range: 05-01-2009 to 22-10-2014
- 1489 OHLC samples per TS
- Hierarchical clustering using ?
- All correlations are statistically significant

NARX forecaster - Results ANN



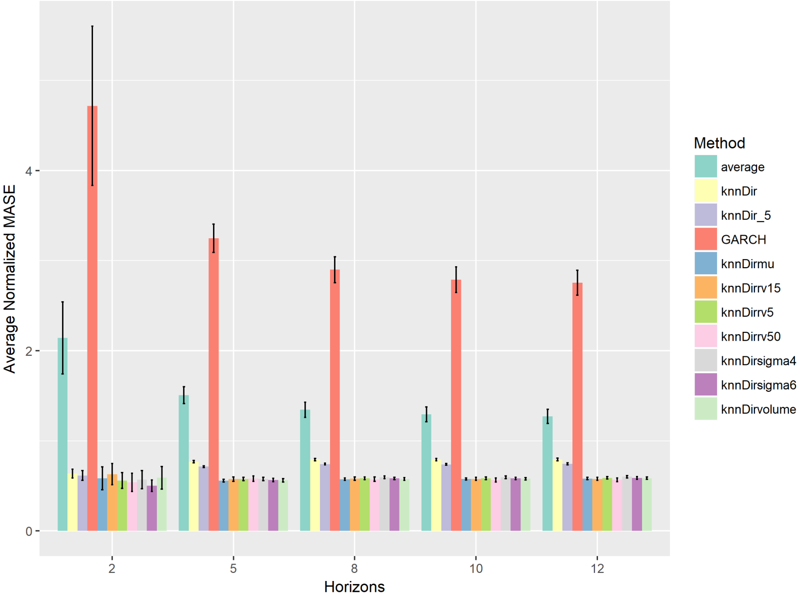
ULB

NARX forecaster - Results ANN

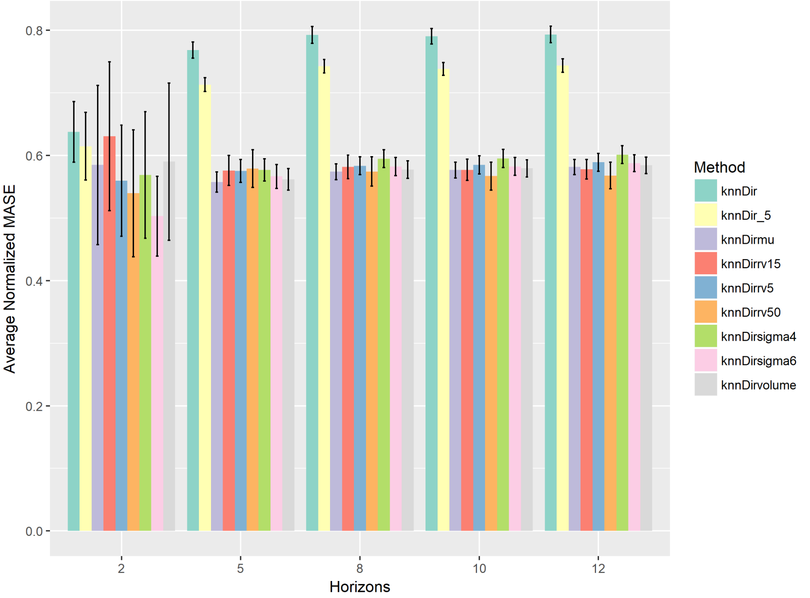


UNIVERSITÉ LIBRE DE BRUXELLES

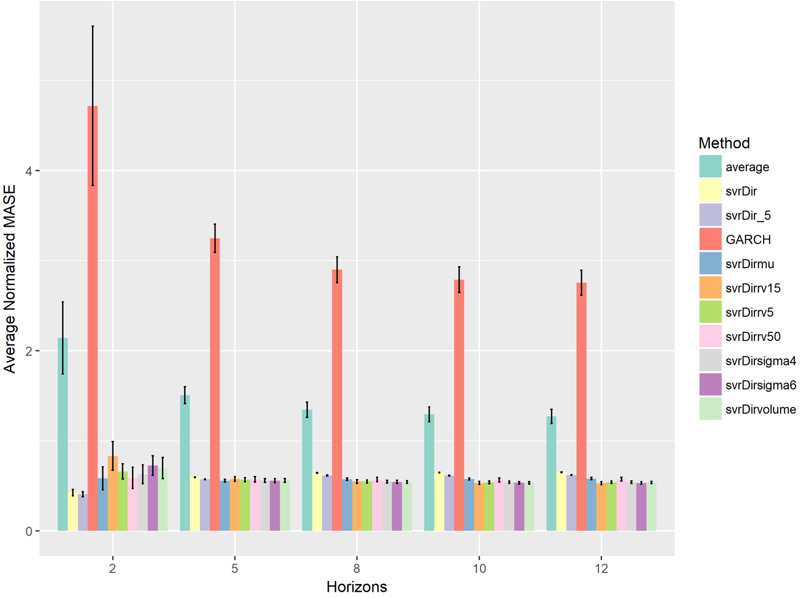
NARX forecaster - Results KNN



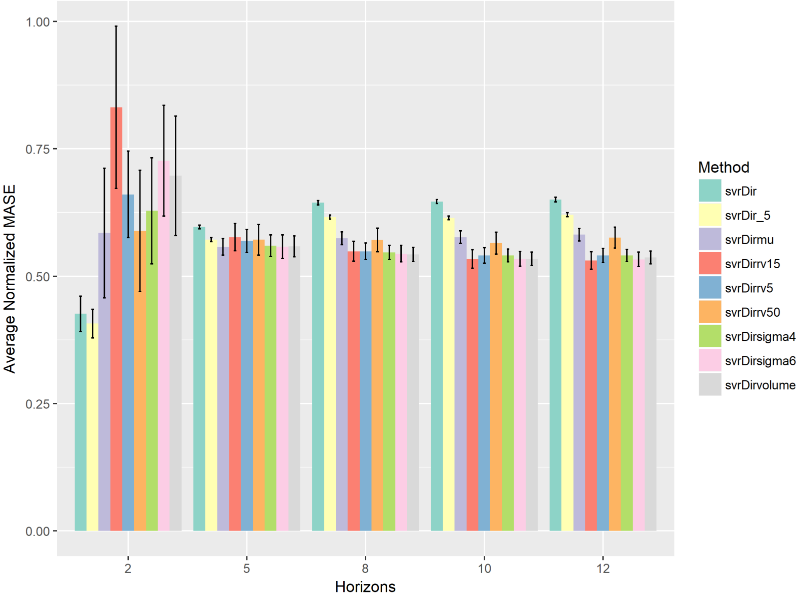
NARX forecaster - Results KNN



NARX forecaster - Results SVR



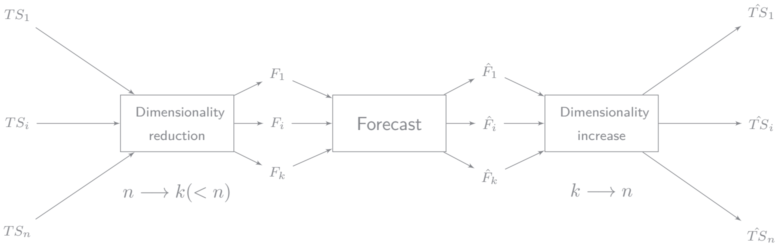
NARX forecaster - Results SVR



Conclusions

- 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 extension - DFML - ?



Dimensionality reduction/increase

- PCA
- Autoencoder

Forecast

- Lazy Learning
- Vector Auto Regressive

Outperforms Recurrent Neural Network, Partial Least Squares, Vector Auto Regressive, Singular Spectrum Analysis, Univariate and Naive.

Future work - Open problems

- Efficient multivariate predictive models?
- Large scale dimensionality reduction techniques?
  - Component-iterative PCA
  - Sample-iterative PCA
  - Autoencoders
- Streaming approaches to multivariate forecasting?

Obrigado pela vossa atenção! Alguma pergunta/observação?

jacopo.de.stefani@ulb.ac.be

Find the paper at:

Open vacancy - Post-doc:

Bibliography I

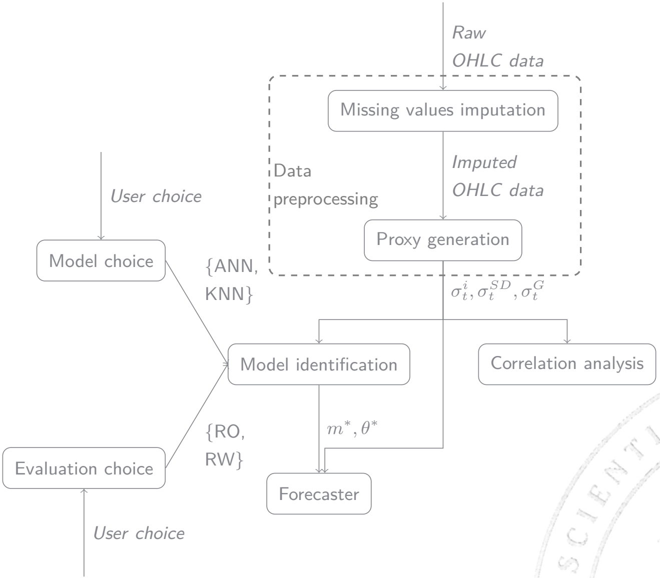
UNIVERSITÉ LIBRE DE BRUXELLES

References

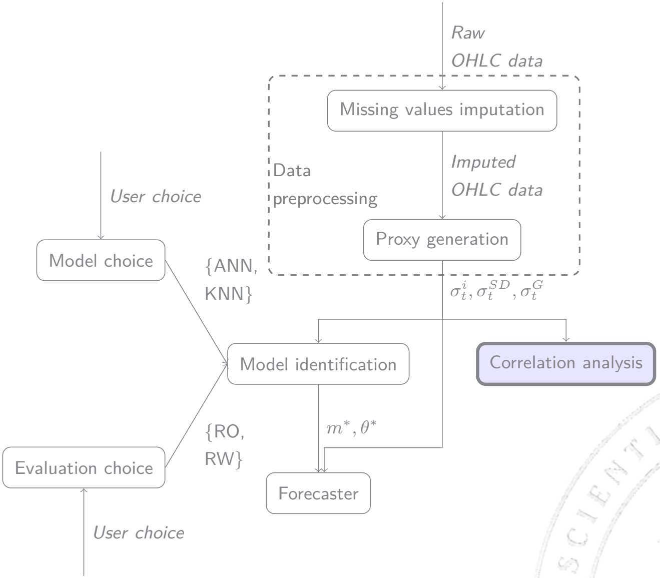
UNIVERSITÉ LIBRE DE BRUXELLES

Appendix

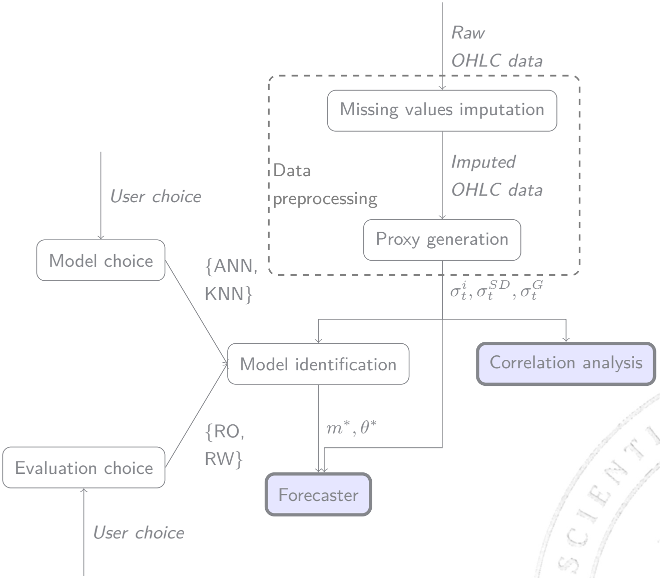
System overview



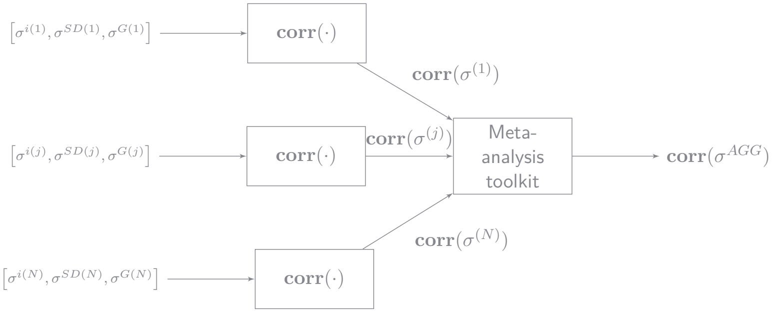
System overview



System overview

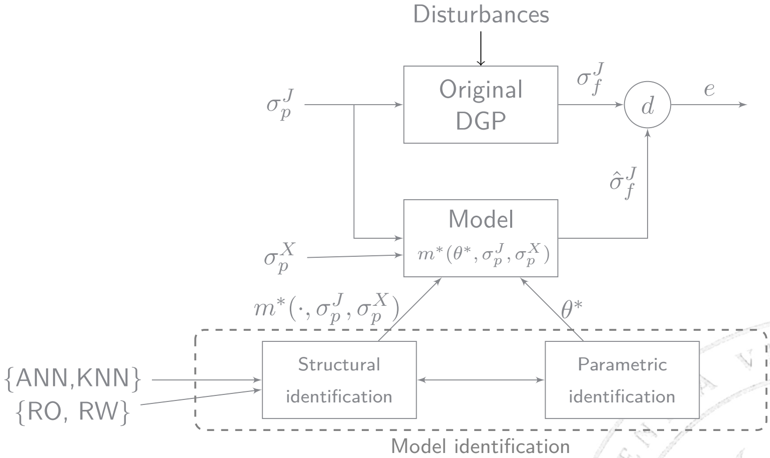


Correlation analysis - Methodology



- 40 Time series (CAC40)
- Time range: 05-01-2009 to 22-10-2014 $\Rightarrow$ 1489 OHLC samples per TS

NARX forecaster - Methodology



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) - ?

UNIVERSITÉ LIBRE DE BRUXELLES

- **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 **?**.

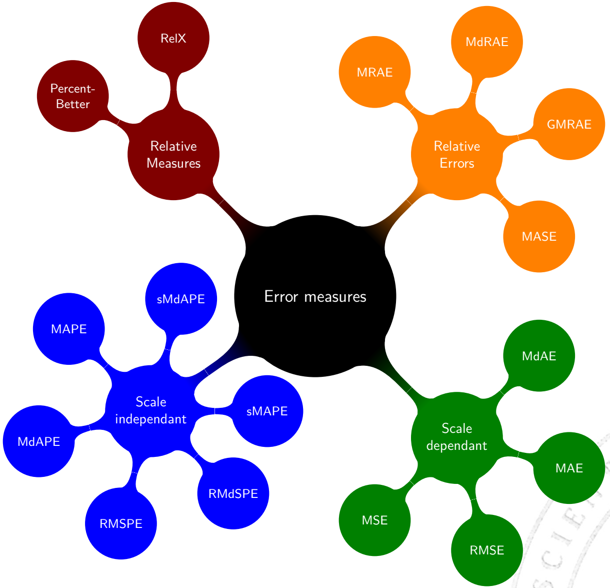
- **Sample standard deviation**

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

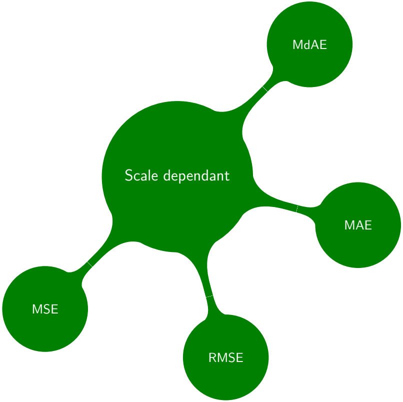
where

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

? - Error measures



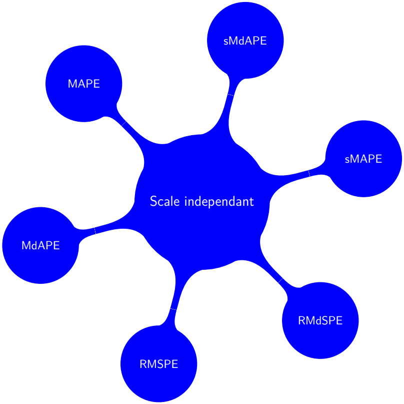
? - Scale dependant



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

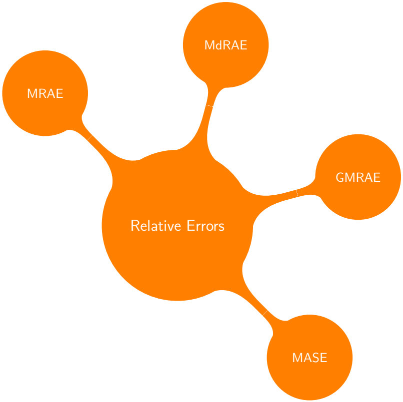
- **MSE** : $\frac{1}{n}\sum_{t=0}^{n}(y_t - \hat{y}_t)^2$
- **RMSE** : $\sqrt{\frac{1}{n}\sum_{t=0}^{n}(y_t - \hat{y}_t)^2}$
- **MAE** : $\frac{1}{n}\sum_{t=0}^{n}|y_t - \hat{y}_t|$
- **MdAE** : $Md_{t\in\{1\cdots n\}}(|y_t - \hat{y}_t|)$

? - 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** :
  $Md_{t\in\{1\cdots n\}}(200 \cdot \frac{|y_t - \hat{y}_t|}{y_t + \hat{y}_t})$

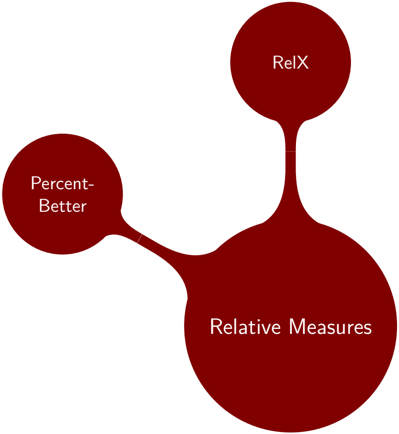
? - Relative errors



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

- **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