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

MIDAS 2017 - ECML-PKDD
Hotel Aleksandar Palace, Skopje, FYROM
Monday 18th September, 2017

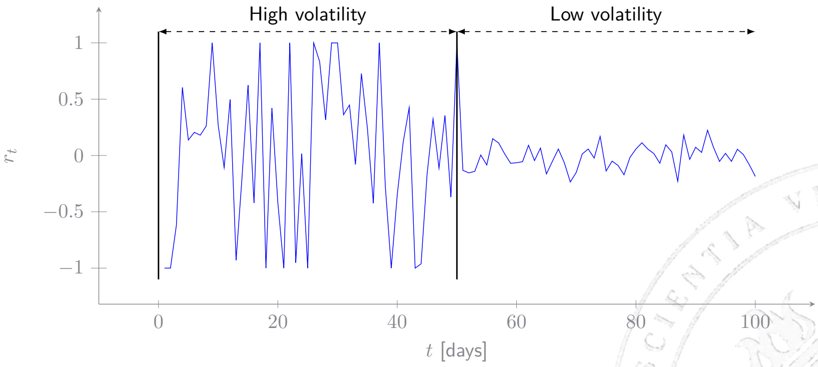
UNIVERSITÉ LIBRE DE BRUXELLES

Problem overview

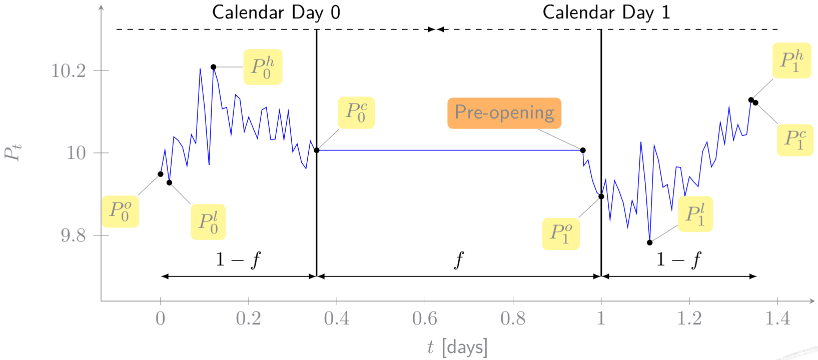
What is volatility?

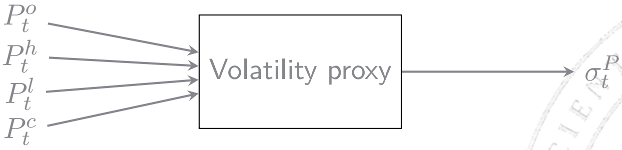
Definition

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

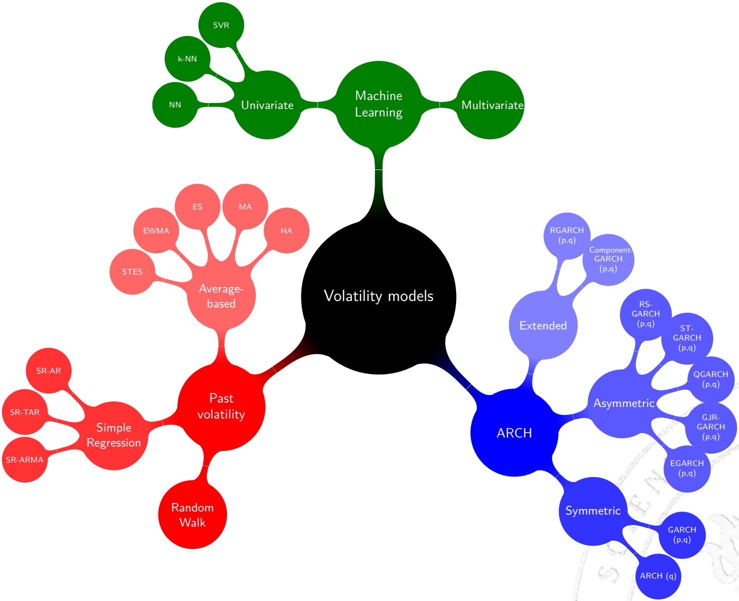


A closer look on data - Volatility proxies

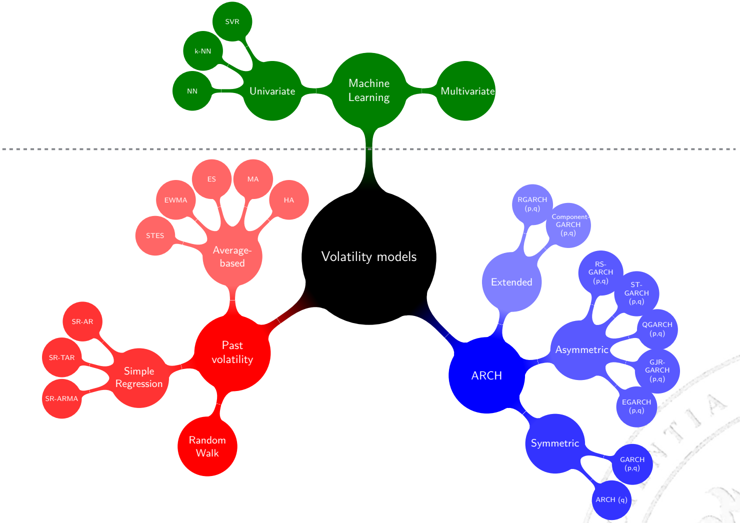




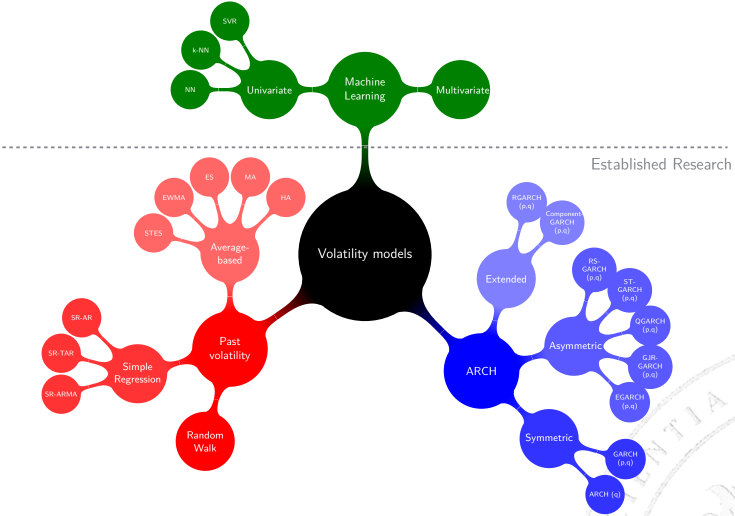
Models for volatility



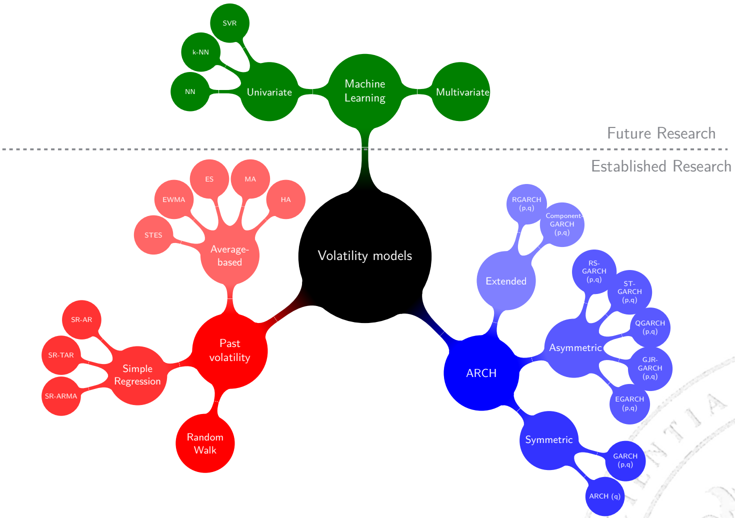
Models for volatility



Models for volatility



Models for volatility



Multistep ahead TS forecasting - Taieb [2014]

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

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

$$\downarrow$$

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

$$\downarrow$$

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

1 Input
1 Output

Multistep ahead forecasting for volatility

State-of-the-art
NAR

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

$$\downarrow$$

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

$$\downarrow$$

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

1 Input
1 Output

Proposed model
NARX

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

$$\downarrow$$

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

$$\downarrow$$

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

2 inputs
1 output

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Multistep ahead forecasting for volatility

State-of-the-art NAR

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

$$\downarrow$$

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

$$\downarrow$$

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

1 Input
1 Output

Proposed model NARX

$$\begin{matrix} [\sigma^P_{t-d} & \cdots & \sigma^P_{t-1}] \\ [\sigma^X_{t-d} & \cdots & \sigma^X_{t-1}] \end{matrix}$$

$$\downarrow$$

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

$$\downarrow$$

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

2 inputs
1 output

Future work

$$\begin{matrix} [\sigma^P_{t-d} & \cdots & \sigma^P_{t-1}] \\ [\cdots & \cdots & \cdots] \\ [\sigma^{X_M}_{t-d} & \cdots & \sigma^{X_M}_{t-1}] \end{matrix}$$

$$\downarrow$$

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

$$\downarrow$$

$$\begin{matrix} [\hat{\sigma}^P_{t} & \cdots & \hat{\sigma}^P_{t+H}] \\ [\cdots & \cdots & \cdots] \\ [\hat{\sigma}^{X_M}_{t} & \cdots & \hat{\sigma}^{X_M}_{t+H}] \end{matrix}$$

M + 1 inputs
M + 1 outputs

Multistep ahead forecasting for volatility

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

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Experimental setup

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

$$\downarrow$$

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

$$\downarrow$$

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

2 TS Input
1 TS Output

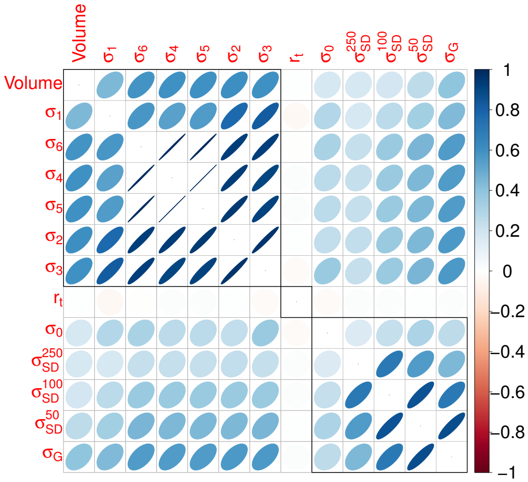
Data: Volatility proxies σ^X, σ^P from CAC40:

- **Price based**
  - $\sigma_{\mathbf{i}}$ **family** - Garman and Klass [1980]
- **Return based**
  - **GARCH (1,1) model** - Hansen and Lunde [2005]
  - **Sample standard deviation**

Models:

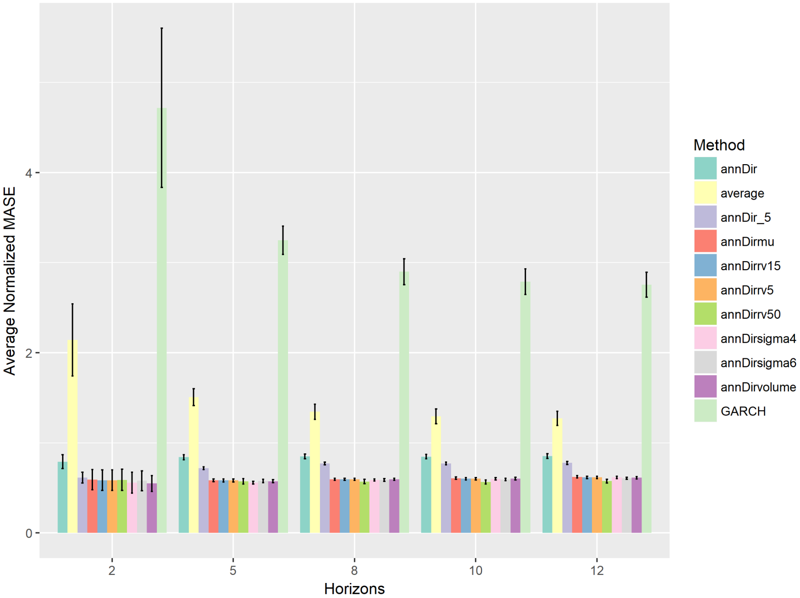
- Feedforward Neural Networks (NAR,NARX)
- k-Nearest Neighbours (NAR,NARX)
- Support Vector Regression (NAR,NARX)
- Naive (w/o σ^X)
- GARCH(1,1) (w/o σ^X)
- Average (w/o σ^X)

Correlation meta-analysis (cf. Field [2001])

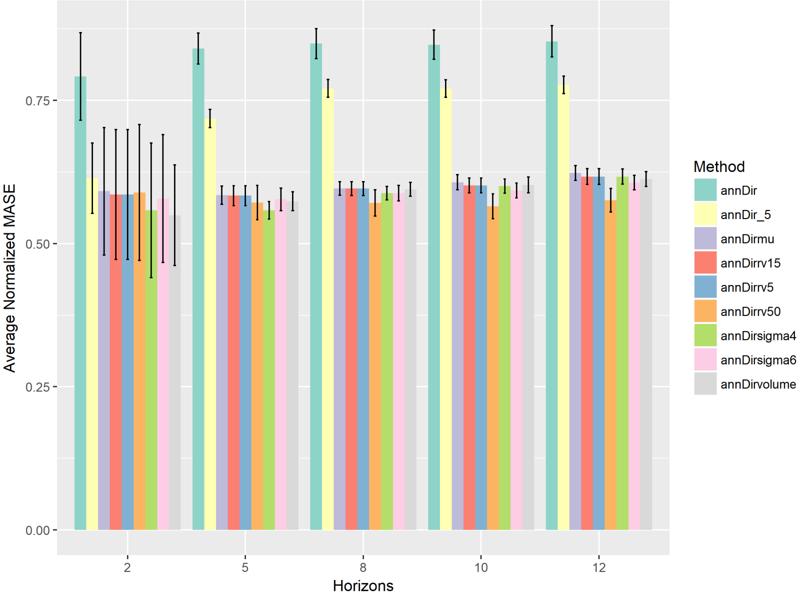


- 40 time series (CAC40)
- Time range: 05-01-2009 to 22-10-2014
- 1489 OHLC samples per TS
- Hierarchical clustering using Ward Jr [1963]
- All correlations are statistically significant

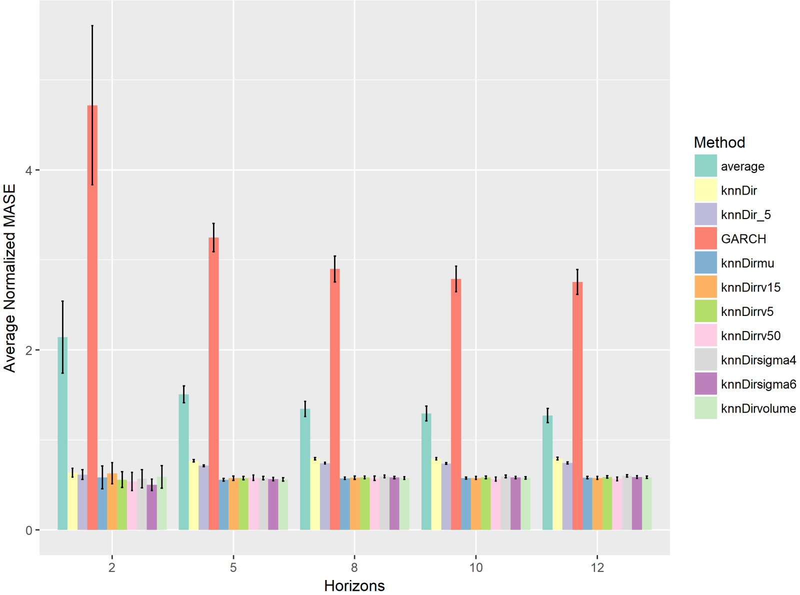
NARX forecaster - Results ANN



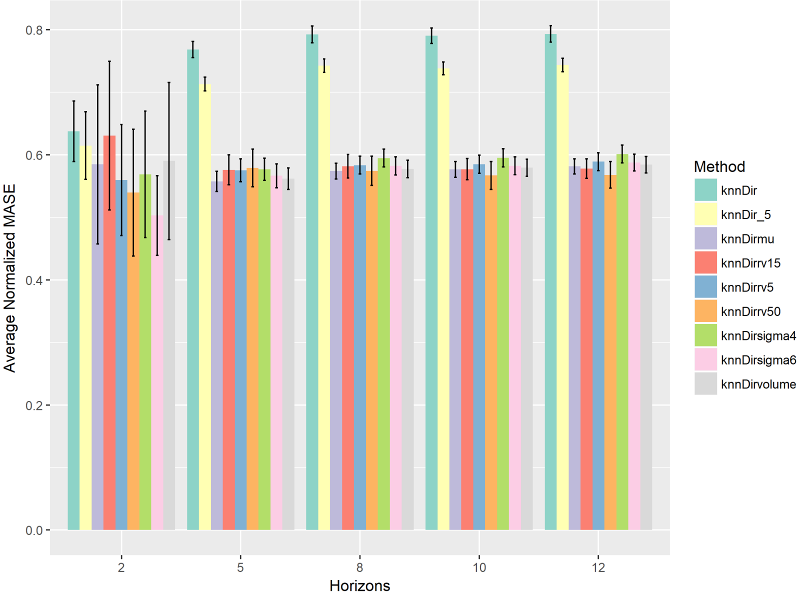
NARX forecaster - Results ANN



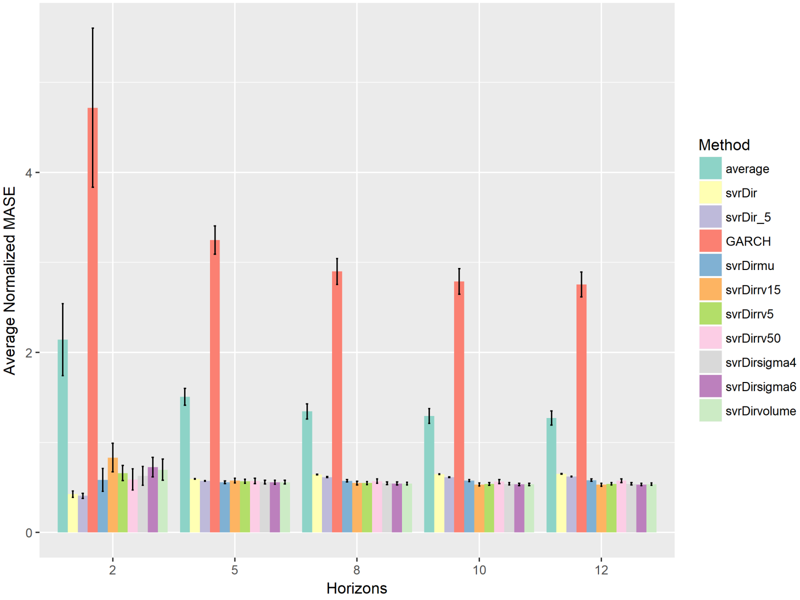
NARX forecaster - Results KNN



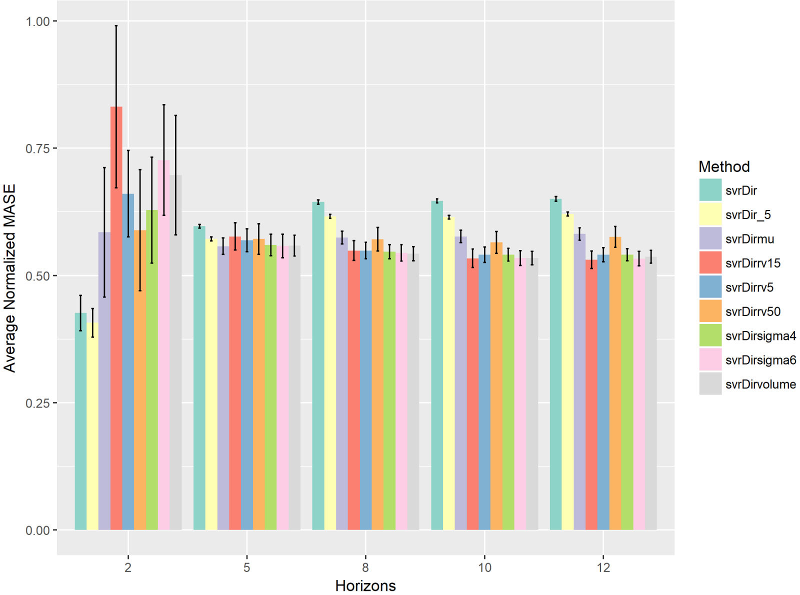
NARX forecaster - Results KNN



NARX forecaster - Results SVR



NARX forecaster - Results SVR



UNIVERSITÉ LIBRE DE BRUXELLES

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.

Thank you for your attention! Any questions/comments?

jacopo.de.stefani@ulb.ac.be

Find the paper at:

Bibliography I

References

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

Andy P Field. Meta-analysis of correlation coefficients: a monte carlo comparison of fixed-and random-effects methods. *Psychological methods*, 6(2):161, 2001.

Mark B Garman and Michael J Klass. On the estimation of security price volatilities from historical data. *Journal of business*, pages 67–78, 1980.

ULB

Bibliography II

UNIVERSITÉ LIBRE DE BRUXELLES

Peter R Hansen and Asger Lunde. A forecast comparison of volatility models: does anything beat a garch (1, 1)? *Journal of applied econometrics*, 20(7):873–889, 2005.

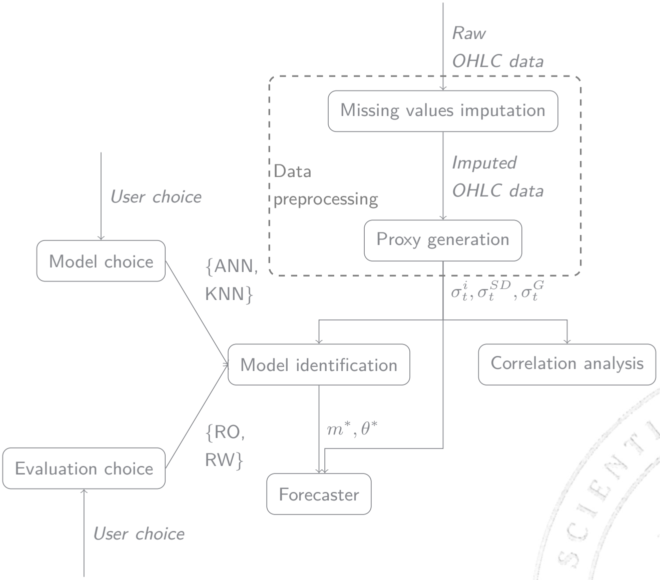
Rob J Hyndman and Anne B Koehler. Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4): 679–688, 2006.

Souhaib Ben Taieb. *Machine learning strategies for multi-step-ahead time series forecasting*. PhD thesis, Ph. D. Thesis, 2014.

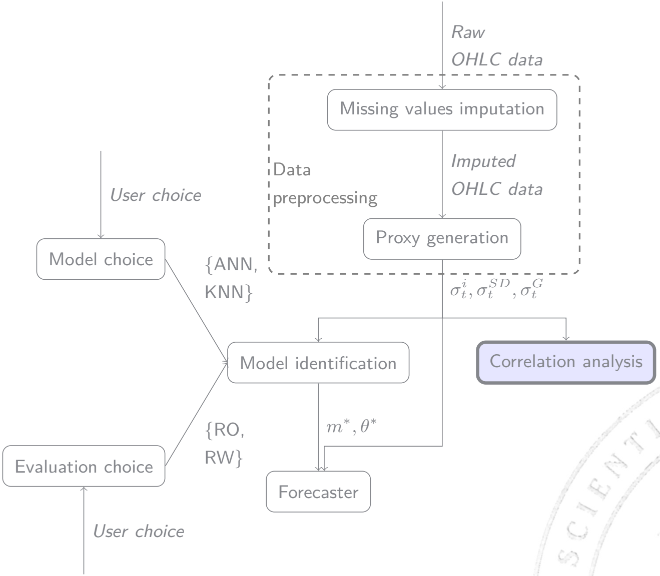
Joe H Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58 (301):236–244, 1963.

Appendix

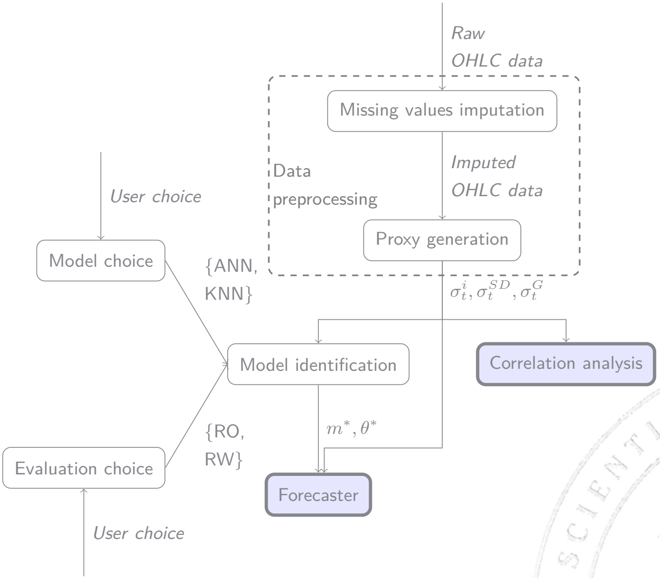
System overview



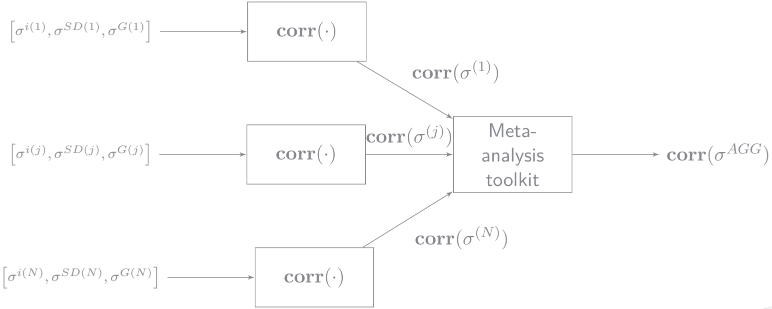
System overview



System overview

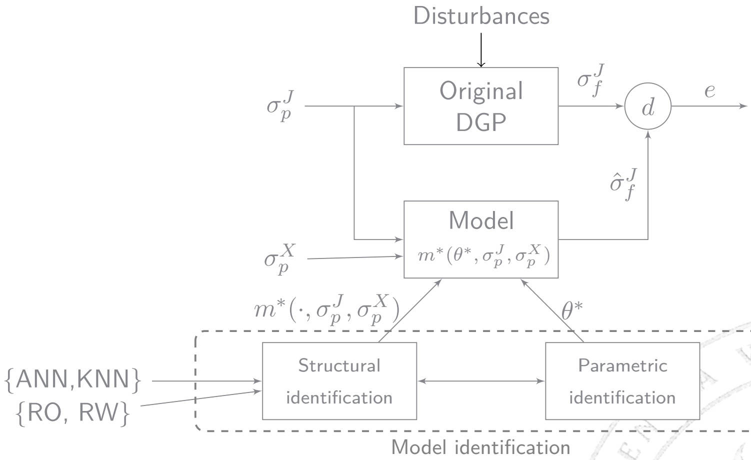


Correlation analysis - Methodology



- 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) - Garman and Klass [1980]

- **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) - Garman and Klass [1980]

- **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** - Hansen and Lunde [2005]

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

  where $\varepsilon_{t-i} \sim \mathcal{N}(0,1)$, with the coefficients $\omega, \alpha_i, \beta_j$ fitted according to Bollerslev [1986].

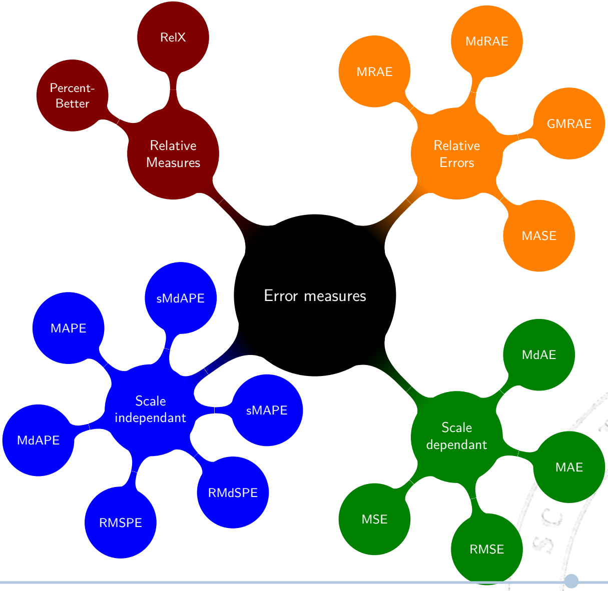
- **Sample standard deviation**

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

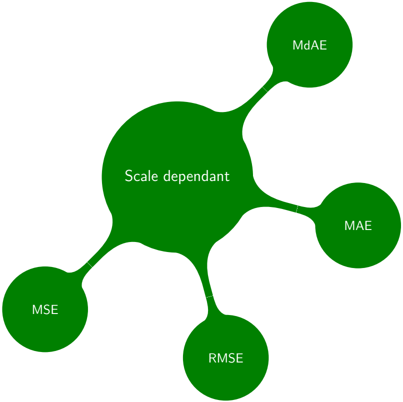
  where

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

Hyndman and Koehler [2006] - Error measures



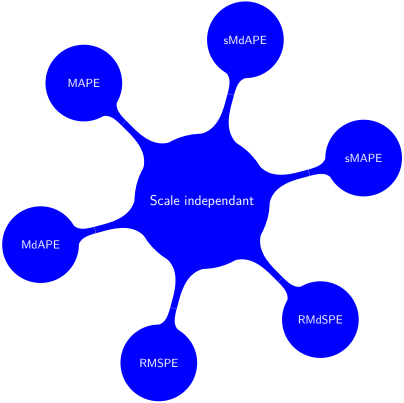
Hyndman and Koehler [2006] - Scale dependant



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

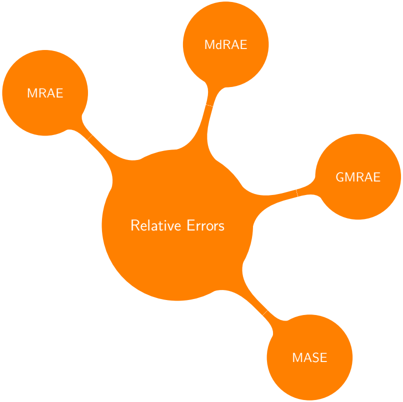
- **MSE** : $\frac{1}{n}\sum_{t=0}^{n}(y_t - \hat{y}_t)^2$
- **RMSE** : $\sqrt{\frac{1}{n}\sum_{t=0}^{n}(y_t - \hat{y}_t)^2}$
- **MAE** : $\frac{1}{n}\sum_{t=0}^{n}|y_t - \hat{y}_t|$
- **MdAE** : $Md_{t\in\{1\cdots n\}}(|y_t - \hat{y}_t|)$

Hyndman and Koehler [2006] - 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** :
  $Md_{t\in\{1\cdots n\}}(200 \cdot \frac{|y_t - \hat{y}_t|}{y_t + \hat{y}_t})$

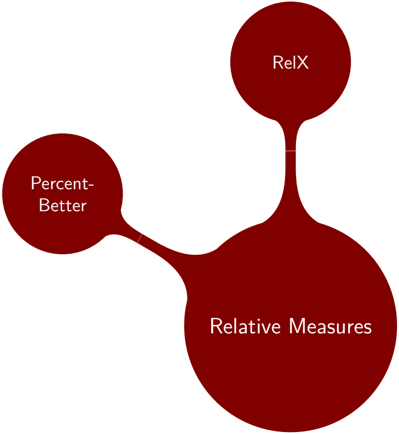
Hyndman and Koehler [2006] - Relative errors



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

- **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** : $\frac{1}{T}\sum_{t=1}^{T}\left(\frac{|e_t|}{\frac{1}{T-1}\sum_{i=2}^{T}|Y_i - Y_{i-1}|}\right)$

Hyndman and Koehler [2006] - 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