Applying Machine Learning to predict the realized volatility of various financial assets

Ivan Mitkov
IMitkov@feba.uni-sofia.bg

Faculty of Economics and Business Administration
Sofia University "Sv. Kliment Ohridski"
www.uni-sofia.bg/feba

Outline

1. Motivation
2. Methodology
3. Data
4. Forecasting accuracy
5. VaR backtesting
6. Conclusion
7. Bibliography

Motivation

- To evaluate volatility forecasts, one needs a good approximation for the true volatility
- High frequency data and realized volatility (RV) possibly a good solution
- Under specific assumptions RV a consistent measure of the true volatility

Realized volatility

Realized volatility (RV) relies on the sum of squared intraday returns

$$RV_t^{(d)} = \sqrt{\sum_{j=0}^{M-1} r_{t-j\cdot\Delta}^2}$$

with

$$\Delta = 1d/M$$

$$r_{t-j\cdot\Delta} = p(t - j\cdot\Delta) - p(t-(j+1)\cdot\Delta)$$

Researchers showed that if $M \to \infty$, then

$$RV_t \xrightarrow{p} IV_t$$

Realized volatility - stylized facts

- ACF is dying out at a hyperbolic rate rather than exponentially, i.e. existence of long memory
- The logarithm of RV is nearly Gaussian
- Its distribution in levels is right-skewed and leptokurtic

Questions

- How to predict the realized volatility correctly?
- Linear or nonlinear models?
- How do the models perform with less training data?
- Could we use them in high and low volatility times?
- Do we find any statistically significant differences in the predicted time series?
- Financial applications of the forecasts?

A linear model

Heterogeneous autoregressive model of realized volatility (HAR-RV)

$$\tilde{\sigma}^{(d)}_{t+1d} = c + \beta^{(d)} RV^{(d)}_t + \beta^{(w)} RV^{(w)}_t + \beta^{(m)} RV^{(m)}_t + \tilde{\omega}^{(d)}_{t+1d}$$

with $RV^{(d)}_t$, $RV^{(w)}_t$ and $RV^{(m)}_t$ being respectively the daily, weekly and monthly realized volatilities for period t

A popular hybrid nonlinear model

- FNNHAR - A hybrid between a feedforward neural network (FNN) and HAR-RV, i.e. $RV_t^{(d)}$, $RV_t^{(w)}$ and $RV_t^{(m)}$ as inputs in FNN
  Arneric, Josip, Tea Poklepovic, and Juin Wen Teai. "Neural network approach in forecasting realized variance using high-frequency data." Business Systems Research Journal 9.2 (2018): 18-34.

Methodology proposal

- A step further in the research of *Arneric, Josip, Tea Poklepovic, and Juin Wen Teai*
- Machine Learning approach with Recurrent Neuronal Netowks
- Usage of SRN, LSTM and GRU with the same three input variables - $RV_t^{(d)}$, $RV_t^{(w)}$ and $RV_t^{(m)}$

DAX30

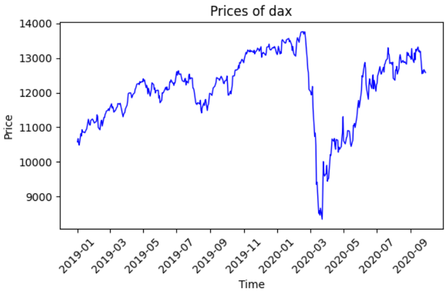

Figure 1: Price evolution the German DAX30

SP500

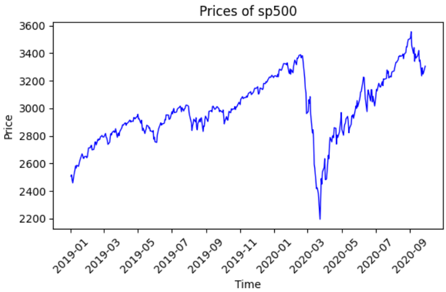

Figure 2: Price evolution of SP500

High volatility times
Jan - Mar 2020

Accuracy - long case

Asset	Loss function	HAR	FNNHAR	SRN	LSTM	GRU
dax30	RMSE	2.51	3.91	6.29	4.21	6.23
sp500	RMSE	1.74	3.01	2.06	2.72	4.45

Table 1: Prediction errors for the high volatility times (long)

Accuracy - short case

Asset	Loss function	HAR	FNNHAR	SRN	LSTM	GRU
dax30	RMSE	2.51	4.21	6.32	4.99	4.66
sp500	RMSE	3.61	2.72	6.28	4.96	4.41

Table 2: Prediction errors for the high volatility times (short)

Low volatility times
May - Aug 2020

Accuracy - long case

Asset	Loss function	HAR	FNNHAR	SRN	LSTM	GRU
dax30	RMSE	3.61	3.51	3.55	3.54	3.51
sp500	RMSE	3.41	5.51	3.28	3.26	3.27

Table 3: Prediction errors for the low volatility times (long)

Accuracy - short case

Asset	Loss function	HAR	FNNHAR	SRN	LSTM	GRU
dax30	RMSE	12.16	8.07	5.34	5.45	5.95
sp500	RMSE	15.23	2.51	2.66	2.31	2.32

Table 4: Prediction errors for the low volatility times (short)

Backtesting strategies for calculating VaR at Alpha = 0.05

The listed below strategies were applied.

- Kupiec's Conditional Coverage Test
- Kupiec's Unconditional Coverage Test
- Christoffersen's Exceedence Independence Test

Accuracy of predictions

High volatility times

- Long case: RNNs have smaller errors than the competing FNN-HAR, valid only for SP500
- Short case: the competing model give better predictions than the RNNs, HAR is the best model

Low volatility times

- Long case: The prosed RNNs have smaller errors
- Short case: The prosed RNNs outperform

Further research

Financial application

- Are the predictions good enough for forecasting VaR and ES

Enhance the predictions with deep neuronal networks

- Adding more neurons and layers

Use another linear models as benchmarks, so jumps in the data would be incorporated as well

- ARFIMA
- HAR-RV-J

Think of another input variables

Bibliography

- Andersen, Torben G and Bollerslev, Tim and Diebold, Francis X and Labys, Paul
  The distribution of realized exchange rate volatility
  available on https://www.tandfonline.com

- Arnerić, Josip and Poklepović, Tea and Teai, Juin Wen
  Neural Network Approach in Forecasting Realized Variance Using High-Frequency Data
  available on https://content.sciendo.com

- Corsi, Fulvio
  A simple approximate long-memory model of realized volatility
  available on http://statmath.wu.ac.at

Bibliography

Kruse, Robinson
Can realized volatility improve the accuracy of Value-at-Risk forecasts
available on https://www.researchgate.net

Louzis, Dimitrios P and Xanthopoulos-Sisinis, Spyros and Refenes, Apostolos P
Realized volatility models and alternative Value-at-Risk prediction strategies
available on https://www.sciencedirect.com

Bibliography

Tsay, Ruey S
Analysis of financial time series

Zhang, Guoqiang and Patuwo, B Eddy and Hu, Michael Y
Forecasting with artificial neural networks:: The state of the art

available on www.citeseerx.ist.psu.edu

Zheng, Fengxia and Zhong, Shouming
Time series forecasting using a hybrid RBF neural network and AR model based on binomial smoothing
available on www.citeseerx.ist.psu.edu