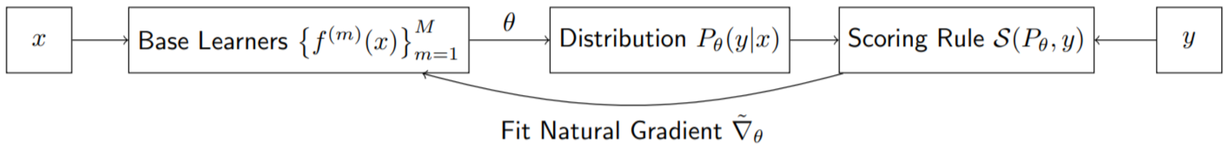
Value at Risk estimation using **N**atural **G**radient **Boosting** for Probabilistic Prediction **model**

Author:
Michał Woźniak (id 385190)

Uniwersytet Warszawski
Wydział Nauk Ekonomicznych

Short introduction to the analyzed problem

- **Motivations:** The area of market risk forecasting using machine learning is not very popular among scientists. Only a few scientific publications deal with this problem. Most often them use very complex neural networks model (Bayesian Long Short-Term Memory, Variational Autoencoders etc.), which in the end are not able to rationally outperform the benchmark (classical econometric models) in case of VaR and capital requirements exceedances. The relatively new NGBoost model seems to be an interesting alternative to well established approaches.
- **Novel approach introduction:** NGBoost allows for probabilistic predictions which is the approach where the model outputs a full probability distribution over the entire outcome space. This functionality fits perfectly, among others to the problem of quantile distribution estimation in highly nonlinear environments.



- **Purpose of the work and main hypothesis:** The aim of this project is to estimate the 1-day 1% and 2.5% Value at Risk measure (based on daily logarithmized rates of return) for S&P 500 index using the NGBoost model with multiple explanatory variables. The model will be tested under special conditions of sudden increased volatility, i.e. during the first wave of the COVID-19 pandemic. I put forward a hypothesis: Will the NGBoost model perform better in the formal backtesting procedure than the GARCH econometric models in the case of 1% and 2.5% VaR estimation during COVID-19 period?

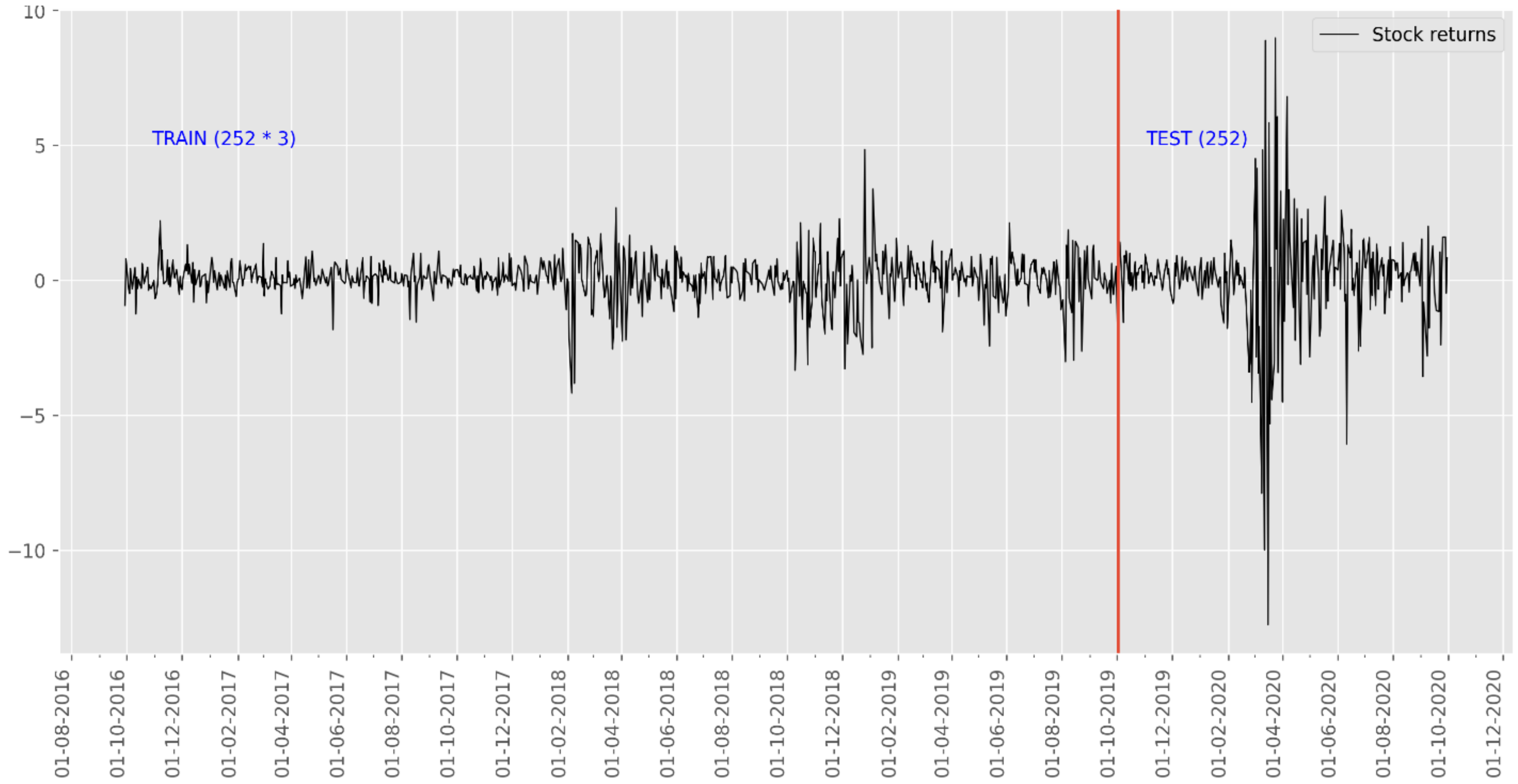
Uniwersytet Warszawski
Wydział Nauk Ekonomicznych

Dataset description

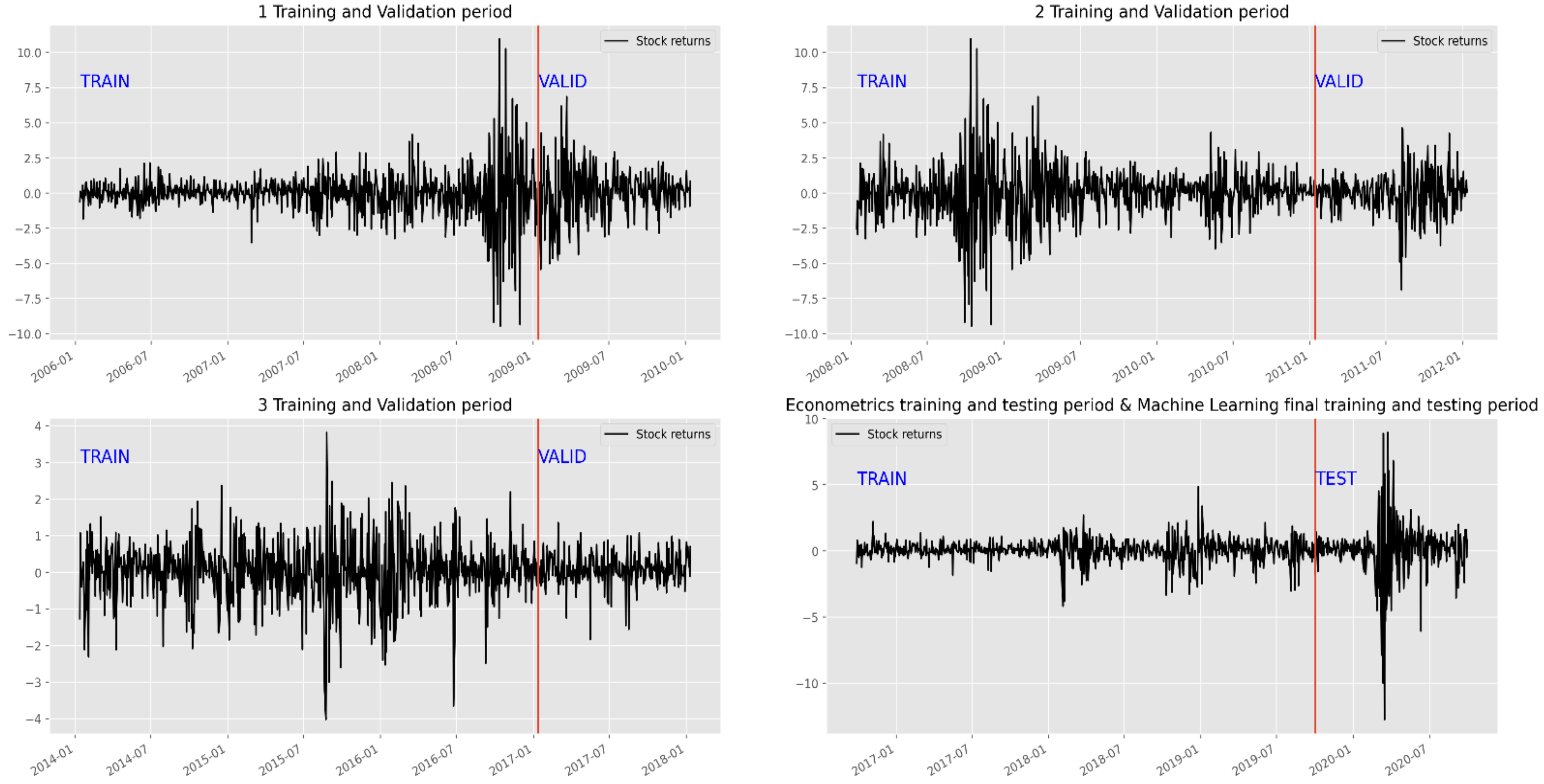
- Dataset provides daily Open price, Close price, Highest price, Lowest price, Volume information for the S&P 500 index.

  For the purposes of the study, the target variable is the logarithmized rate of return $r_t = \ln(\frac{p_t}{p_{t-1}})$. The data covers the period from January 1, 2006 to September 30, 2020. The data comes from the website stooq.pl.
- The models will be trained, validated and tested in a one-step-ahead approach in a moving/sliding time window (always the same length of the training set - 3 * 252 days, and the validation/test set – 252 days).
- The dataset is divided into time windows: in-sample dataset (train), out-of-sample dataset (validation) and out-of-sample out-of-time dataset (test, which will undergo final backtesting). These windows vary depending on the modeling approach:
  - In case of econometric models I defined only two periods: training and testing, because here we do not deal with hyperparameters tuning [visualization on the following pages]
  - In case of machine learning models I definded 4 periods: 1 training and validation , 2 training and validation, 3 training and validation, final training and testing (analogy to econometric models periods!) [visualization on the following pages]. First three periods will be used in a cross validation with „number of VaR exceeds" metric!

Dataset description – cont'd



Dataset description – cont’d



Uniwersytet Warszawski
Wydział Nauk Ekonomicznych

Dataset description – cont’d

Descriptive statistics for final training and testing period (same for econometric and machine learning models)

	count	mean	std	min	25%	50%	75%	max	skewness	kurtosiss
Final training period (2016-09-29 - 2019-10-01)	756	0.040097	0.805931	-4.18426	-0.2421	0.058973	0.443732	4.840324	-0.64889	5.512918
Final testing period (2019-10-02 - 2020-09-30)	252	0.053309	2.146715	-12.7652	-0.46783	0.232585	0.823434	8.968316	-0.90692	9.552796

Description of the selected models

Benchmark models:

1. Simple statistical models
   - a. Historical simulation
   - b. Parametric quantile from Normal distribution
   - c. Parametric quantile from Skewed Normal distribution
   - d. Parametric quantile from T distribution
   - e. Parametric quantile from Laplace distribution
   - f. Parametric quantile from Asymmetric Laplace distribution
   - g. Parametric quantile from Generalized extreme value distribution

2. Econometric models
   - a. AR(1) – GARCH(1,1) Normal
   - b. AR(1) – GARCH(1,1) T
   - c. AR(1) – GARCH(1,1) Skewed T
   - d. AR(1) – GARCH(1,1) GED
   - e. AR(1) – QML-GARCH(1,1)

Description of the selected models – cont'd

My challengers models:

1. NGBoost – VaR estimation as the quantile of the full probability distribution of stock returns returned by the model

2. NGBoost – VaR estimation using quasi GARCH approach: VaR = model.forecasted.mean + model.forecasted.variance * (model.standardized_residuals)

3. Ensembling by simple average for QML-GARCH model (best GARCH model) and NGBoost models from 1.

4. Switching model NGBoost from 1. and QML-GARCH model (best GARCH model) – NGBoost works in normal periods and QML-GARCH works in increasing volatility periods

For each of the above approaches, I estimated the following NGBoost model pool (choice based on greedy grid search):

1. base learner: ExtraTreeRegressor (max_depth = 3, min_samples_split = 2), distribution: Laplace, ETA: 0.01, boosting iterations: 100

2. base learner: ExtraTreeRegressor (max_depth = 3, min_samples_split = 2), distribution: Laplace, ETA: 0.01, boosting iterations: 250

3. base learner: ExtraTreeRegressor (max_depth = 3, min_samples_split = 2), distribution: Laplace, ETA: 0.01, boosting iterations: 500

4. base learner: ExtraTreeRegressor (max_depth = 3, min_samples_split = 2), distribution: T-Student, ETA: 0.01, boosting iterations: 250

5. base learner: ExtraTreeRegressor (max_depth = 3, min_samples_split = 2), distribution: T-Student, ETA: 0.01, boosting iterations: 100

6. base learner: ExtraTreeRegressor (max_depth = 3, min_samples_split = 2), distribution: T-Student, ETA: 0.1, boosting iterations: 250

Uniwersytet Warszawski
Wydział Nauk Ekonomicznych

Description of the selected metrics / backtesting

- **Loss function for NGBoost models**: negative log-likelihood
- **Main evaluation metrics:** number of exceeds VaR and amount of financial provisions (starting capital 1e6 USD)
- **Additional formal VaR tests:** Kupiec test (unconditional coverage), Christoffersen test (conditional coverage) and Engle test (dynamic quantile).

List of the models parameters

In case of **econometric** models I optimized following parameters:

1. Variance model (ARCH, GARCH, EGARCH, HARCH) and its parameters (p, o, q)
2. Mean model (Constant, Zero, AR) and its parameters (p)
3. Distribution (Normal, T, Skewed T, GED)

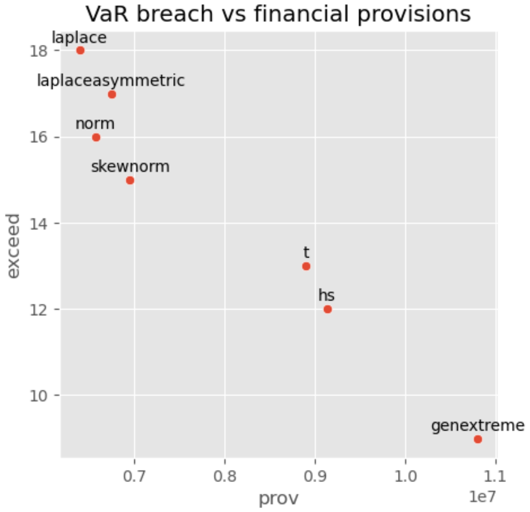
In case of **NGBoost** models I optimized following hyperparameters:

1. Base learner (Ridge Regression, Extra Decision Tree, Decision Tree and their hyperparamters)
2. Distribution (Laplace, Normal, T)
3. Learning rate/ETA (0.01, 0.1, 0.25)
4. Number of boosting rounds (100, 250, 500)

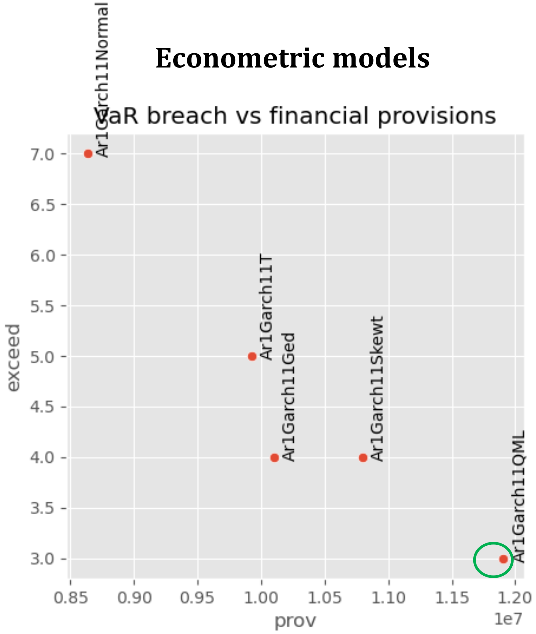
Results – 1% VaR

VaR breaches vs Financial provision

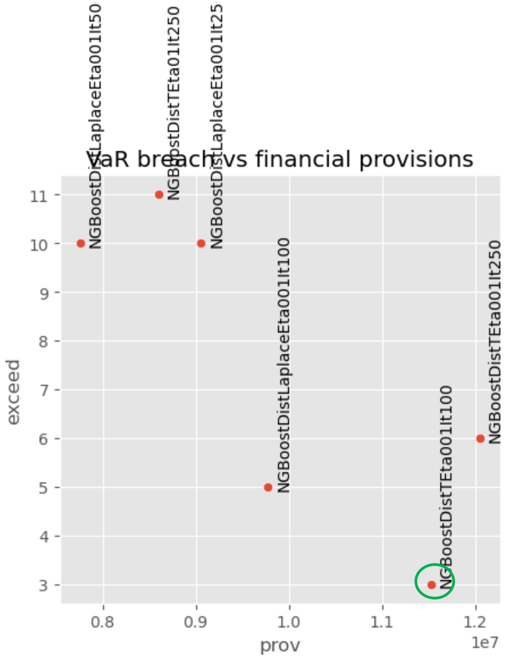
Simple statistical models



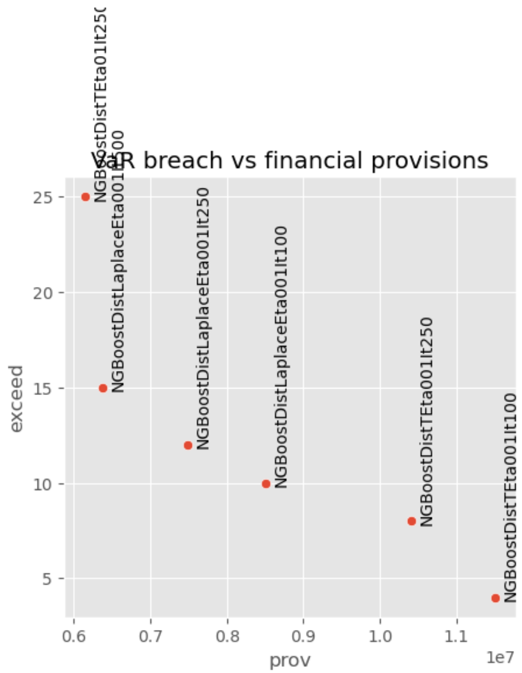
Econometric models



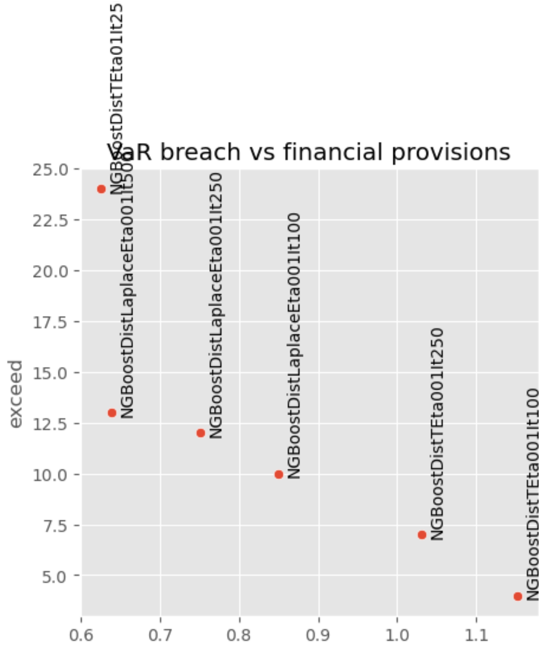
NGBoost 1st approach models



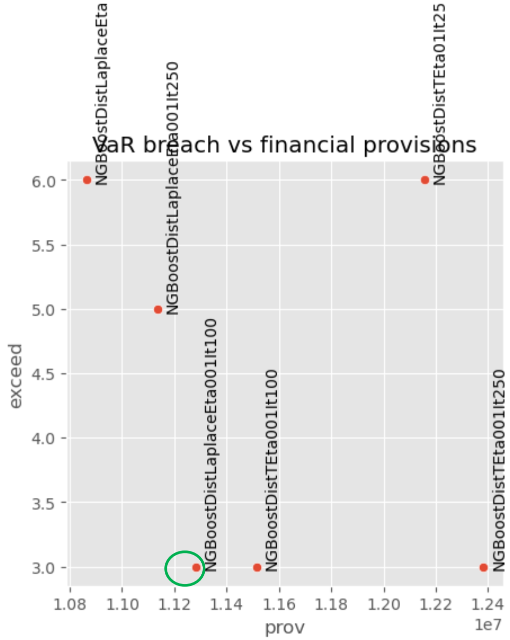
NGBoost 2nd approach models



NGBoost 3rd approach models



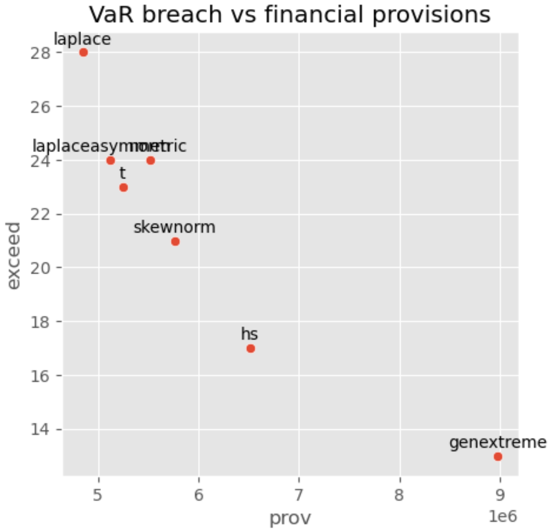
NGBoost 4th approach models



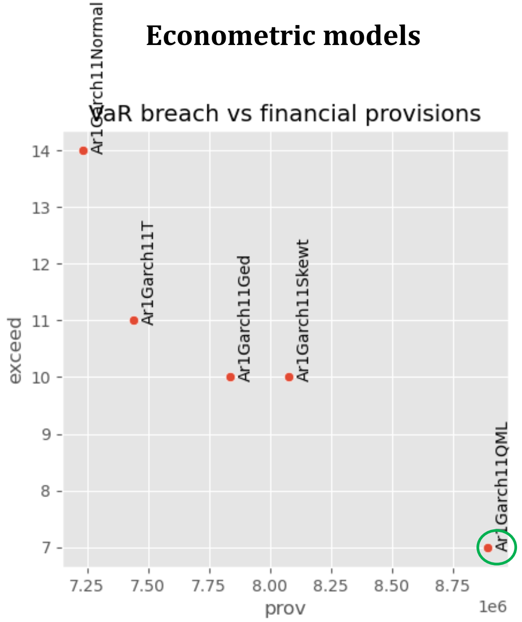
Results – 2.5% VaR

VaR breaches vs Financial provision

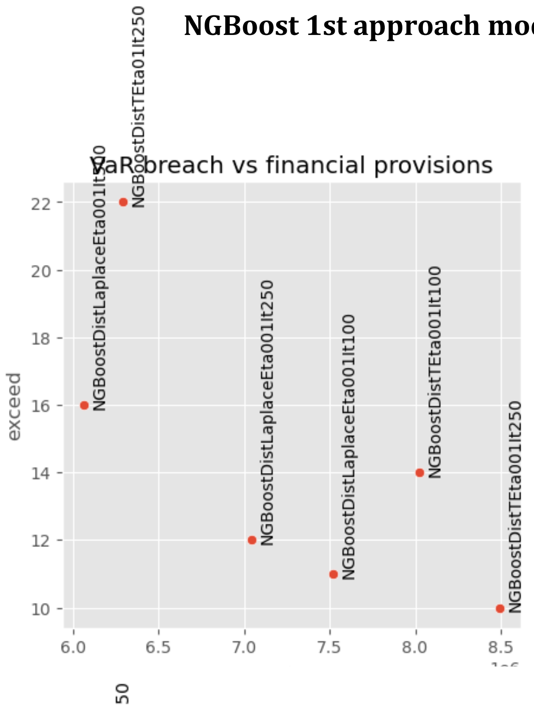
Simple statistical models



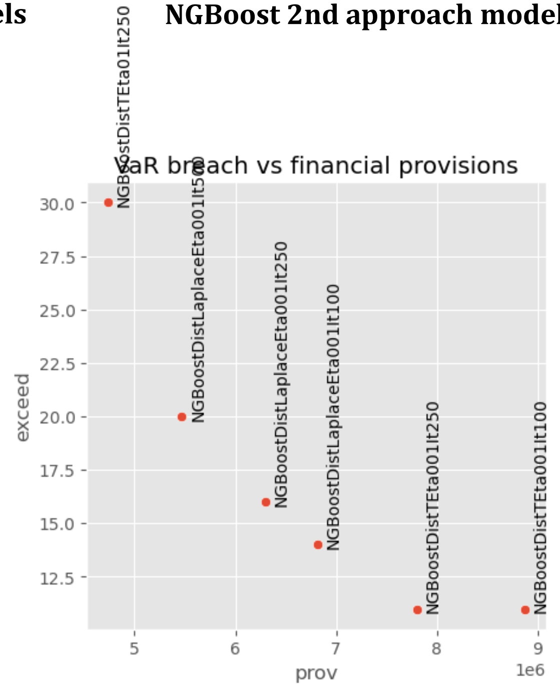
Econometric models



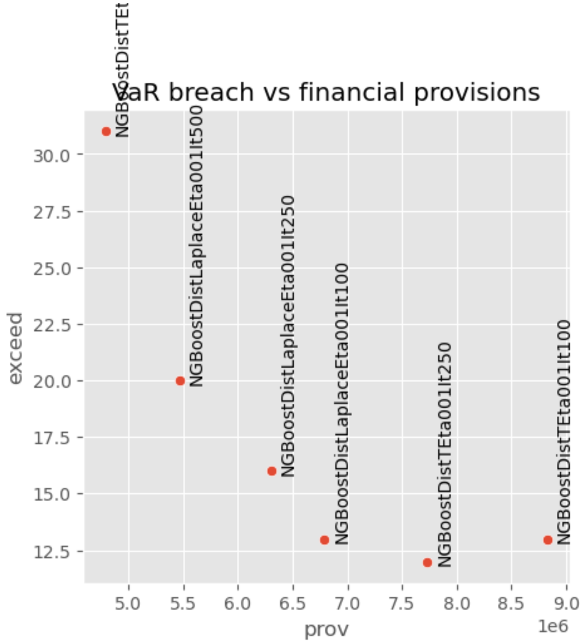
NGBoost 1st approach models



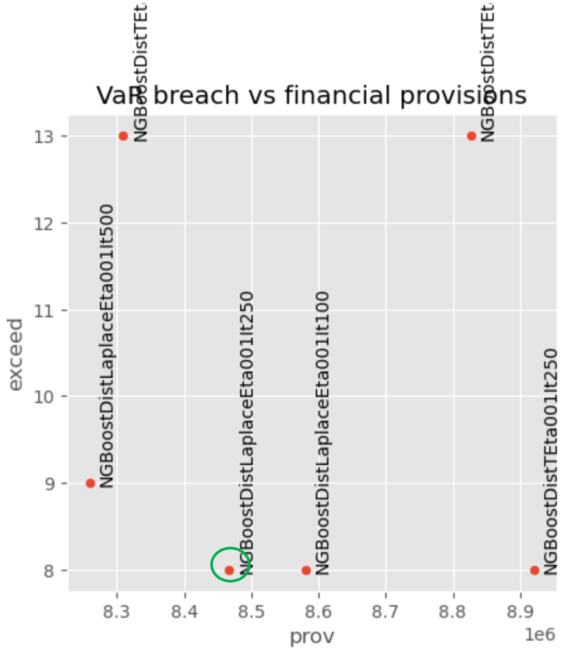
NGBoost 2nd approach models



NGBoost 3rd approach models



NGBoost 4th approach models



Conclusions

The aim of this project was to estimate the 1-day 1% and 2.5% Value at Risk measure (based on daily logarithmized rates of return) for S&P 500 index using the NGBoost model with multiple explanatory variables. The model was tested under special conditions of sudden increased volatility, i.e., during the first wave of the COVID-19 pandemic. This goal has been achieved! At the beginning I put forward a hypothesis: Can the NGBoost model perform better in the formal backtesting procedure than the GARCH econometric models in the case of 1% and 2.5% VaR estimation during COVID-19 period? Now, I can state that **NGBoost model perform better in the formal backtesting procedure than the GARCH econometric models in the case of 1% VaR estimation during COVID-19 period**. **In case of 2.5 % VaR NGBoost results were slightly worse due to number of VaR breaches but generated lower financial provisions.** It leads to the trade-off between number of VaR breaches and financial provisions. But generally NGBoost can successfully compete with GARCH class models. **At the same time, it is worth noting that for 2.5% VaR, the GARCH models performed much better in times of sudden increased volatility, and the NGBoost model in times of relative calm in the markets.**

References

- Abad, P., Benito, S. and López, C., 2014. A comprehensive review of Value at Risk methodologies. The Spanish Review of Financial Economics, 12(1), pp.15-32.
- Buczyński, M. and Chlebus, M., 2018. Comparison of Semi-Parametric and Benchmark Value-At-Risk Models in Several Time Periods with Different Volatility Levels. e-Finanse, 14(2), pp.67-82.
- Buczyński, M. and Chlebus, M., 2020. Old-fashioned parametric models are still the best: a comparison of value-at-risk approaches in several volatility states. The Journal of Risk Model Validation.
- Chlebus, M., Dyczko, M. and Woźniak, M. 2019. Nvidia's stock returns prediction using machine learning techniques for time series forecasting problem. https://www.wne.uw.edu.pl/files/6415/9481/5844/WNE_WP328.pdf
- Duan, T., Avati, A., Ding, D., Thai, K., Basu, S., Ng, A. and Schuler, A. 2019. NGBoost: Natural Gradient Boosting for Probabilistic Prediction. https://arxiv.org/abs/1910.03225v4
- Szubzda, F. and Chlebus, M., 2020. Comparison of Block Maxima and Peaks Over Threshold Value-at-Risk models for market risk in various economic conditions. Central European Economic Journal, 6(53), pp.70-85.