Machine Learning and Mixed-Frequency Data Based Forecasting of Realized Volatility (the Case of Russian Stock Market)

Vladimir Pyrlik* Pavel Elizarov† Aleksandra Leonova†

CERGE-EI HSE University HSE University

September 2021

Abstract

Returns volatility is a crucial measure that is most commonly used to assess stock market risks. Forecasting volatility proved a challenging task that has been widely researched. Because realized volatility (RV) directly evaluates volatility from high-frequent market data and tends to provide precise forecasts of the future volatility, the RV approach has become a preferred one among practitioners and researchers. Since the baseline heterogeneous autoregressive model of realized volatility (HAR-RV) was introduced, substantial research has been devoted to improving its explanatory and predictive capabilities. Massive studies use a vast variety of additional variables to explain the changes in volatility on the

*the corresponding author; address: CERGE-EI, a joint workplace of Center for Economic Research and Graduate Education, Charles University and the Economics Institute of the Czech Academy of Sciences, Politických vězňů 7, P.O. Box 882, 111 21 Prague 1, Czech Republic; e-mail: vladimir.pyrlik@cerge-ei.cz

†address: St. Petersburg School of Economics and Management, HSE University, Kantemirovskaya 3A-1, 194100 St. Petersburg, Russia

markets, including financial indicators of firms or markets, various spillover effects, calendar effects, or macroeconomic indices. Another way to improve the predictive power of the models is to apply machine learning algorithms, yet this approach is understudied at the moment. We combine the two approaches and study the predictive performance of selected machine learning algorithms relative to the benchmark model in the task of prediction of realized volatility of stock returns on the Russian stock market based on both historical values of the volatility itself, and a broad set of potentially informative financial and macroeconomic indicators. We find that both the traditional methodology and the machine learning approach provide us with reasonable predictive power in terms of RMSE of realized volatility forecasts on rolling training samples, with the machine learning algorithms that utilize additional exogenous regressors generally outperforming the traditional baseline autoregressive model. Particularly, Lasso regression is shown to be rather efficient both in terms of relatively easy implementation, and precision of forecasts. More complicated algorithms (Random Forest, Gradient Boosting, Long Short-Term Memory) are very promising, yet we show that to benefit from them, it requires to tune them very finely and re-train rather often, which is a computationally demanding task.

KEYWORDS: Heterogeneous autoregressive model; Machine learning; Lasso; Gradient Boosting; Random Forest; Long Short-Term Memory; Realized volatility; Russian stock market; Mixed-frequency data