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

Vladimir Pyrlik* CERGE-EI

Pavel Elizarov† HSE University

Aleksandra Leonova† HSE University

November 2021

Abstract

We assess the performance of selected machine learning algorithms (lasso, random forest, gradient boosting, and long short-term memory) in forecasting the daily realized volatility of returns of selected top stocks in the Russian stock market in comparison with a heterogeneous autoregressive realized volatility benchmark in 2018-2020. We seek to improve the predictive power of the models by including various economic indicators that carry information about future volatility. We find that lasso delivers a good combination of easy implementation and forecast precision. The other algorithms require fine-tuning and frequent re-training, otherwise they are likely to fail to outperform the benchmark often enough. Only the basic lagged log-RV values are significant explanatory variables in terms of the benchmark in-sample quality. Many economic indicators of mixed frequencies improve the predictive power of lasso though, including calendar and overnight effects, financial spillovers from local and global markets, and various macroeconomics indicators.

KEYWORDS: heterogeneous autoregressive model; machine learning; lasso; gradient boosting; random forest; long short-term memory; realized volatility; Russian stock market; mixed-frequency data

*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