

Contents

1	Abstract	3
2	Introduction	3
3	Literature Review	4
4	Methodology	6
5	Anticipated Results	7
6	Conclusion	8

1. Abstract

This purpose of this study is to investigate the relationship between a list of macroeconomic variables and quality with which Deep Learning can predict Russian stock market. The connection between commodity prices, exchange rates, inflation, and foreign markets is a well-established area of economic and econometric research. On the other hand, in the prediction field, while modern Deep Learning methods are becoming more and more popular for financial time series forecasting, those fundamental economic factors are rarely utilized. Using daily data on several Russian publicly traded companies in the period of 2011-2019, I plan to determine if the mentioned factors are suitable as inputs for LSTM model, improving its performance compared to baselines and ARIMA in terms of RMSE.

2. Introduction

In recent years, the surge in the power of Machine Learning, and especially Deep Learning, has lead to dramatic advancements in computer vision, natural language processing, and various other tasks, including time series forecasting, and, in particular, financial time series forecasting [?]. However, this stream of literature is inherently technical. Machine Learning specialists view financial data merely as one of time series domains, alongside fields like energy consumption, weather forecasting or turnout rates. Their works, while contributing to minimizing prediction error, use little knowledge of economics or finance to provide any meaningful insights of the way financial markets work of behave. The series considered are either univariate, use standard High-Low-Open-Close data on stocks, or other ML-suitable data like tweets or news articles [?]. Macroeconomic factors such as exchange rates or oil prices can hardly be found.

On the other hand, economists searching for causality between stock markets and fundamental factors based on theoretical insights is a different stream of literature. Connections of stock markets with oil prices, exchange rates, inflation and foreign markets have been documented for countries around the world [?]. However, these works use different, statistical toolkit, consisting of mostly linear and ARCH methods. In terms of pure prediction power, Machine Learning methods like Random Forest or SVM are proved to produce better results than linear Box-Jenkins approaches [?]. Deep learning approaches like LSTM outperform them as well [?].

My aim therefore is to test applicability of macroeconomic data for the task of forecasting Russian stock prices time series. Compared to univariate setting, do factors like oil prices, exchange rates or foreign market indexes improve predictions produces by Deep Learning algorithms?

It is not a classic econometric study, as improving prediction does not imply causality. However, improving prediction would mean there is certainly some relevant information, which powerful ML methods are able to retrieve and utilize.

The rest of this proposal is organized as follows. Relevant literature is reviewed in the next section. Methodology section explains data acquisition procedure, training pipeline, optimization algorithms, proposes the list of models, and sketches variable selection approach. Then I speculate on possible results, and, finally, conclude the proposal with limitations of the study.

3. Literature Review

Among the most studied relationships of external variables with stock prices are oil prices, exchange rates and inflation. They make up a significant branch of financial econometric literature.

The link between oil prices and economic activity, and, hence, stock prices can be considered common knowledge. There are various theoretical explanations for this relationship [?]. Firstly, oil prices directly affect future cash flows of oil-producer and oil-consumer firms, which get evaluated in their stock price. Therefore, oil-producer stocks move in the same direction as oil price as it affects their profits, while oil-consumers experience a rise or reduction in their costs which should drive their stock's price to the opposite direction [? ?]. Secondly, oil prices affect basic macroeconomic variables such as output and inflation, which in turn have their effect on the whole financial market. Finally, fluctuations in oil price increase uncertainty [?].

As for empirical proofs, the situation is unclear. For example, [?] and [?] find negative relationship, [?] document positive effect, and studies like [?] or [?] report no effect at all. Moreover, it is apparent that different countries exhibit different effects from oil prices. The studies of [?] and [?] show that for oil-exporting countries the effect of oil price change is positive, and the opposite is true for oil-importing countries. The same can be said about different sectors. Energy-supplying sectors have been found to be affected positively by oil prices, while energy-consumers such as Buildings, Manufacturing, Transport, Food, Chemical, Medical, Computer and other sectors are more likely to suffer from them [? ?].

Another factor considered to be affecting stock prices is exchange rate. Economic reasoning is the following. Similar to oil prices, exchange rates affect future cash flows of exporting and importing firms [?]. Depreciation of domestic currency benefits exporters, increasing their profits, and hurts importers, increasing their costs, and, through this, affects their stock prices. However, the reverse link is known as well: as the rise (or fall) in stock prices indicates rise (or fall) in their holders' wealth, it also affects demand for imports, driving exchange rate up (or down).

Empirical evidence here is even more ambiguous than for oil prices. A number of studies document existence of relationship only for a subsample of studies countries [? ? , among many others]. The direction and existence of causality also depends on market condition, flowing from stock prices to exchange rates in crisis and from exchange rates to stock prices otherwise [? ?].

There is also evidence of inflation [? ?], interest rates [?], real output [? ?], commodity prices other than oil [? ?] and many other factors influencing stock market causally.

As for the case of Russia, there are studies examining causal determinants of Russian stock market. [?] consider the period between 1997 and 2012 and oil prices and eastern European stock markets performance as variables to affect stock market in Russia. According to their findings, oil prices stopped being significant after 2006, while correlation with other markets increased in that period. [?] studies several emerging economies' markets, including Brazil, India, China and Russia in the period of 1999-2006. They report a not significant result for both oil prices and exchange rates and all four countries. [?] study the impact of oil prices, exchange rates, foreign stock indexes and interest rates on Russian stock market from 2003 to 2018. The authors document a varying relationship, with variations caused by structural breaks. This list of studies is by no means inclusive.

Econometric models employed by authors in most of mentioned articles are either some variations of GARCH (generalized autoregressive conditional heteroskedasticity) model or VAR, a multivariate generalization of ARIMA (autoregressive integrated moving average) model. While being suitable for causality determination, in terms of predictive power, those are by far not the state of the art at the moment [?]. It is apparent that for as complex entity as stock market prices or returns simple linear or ARCH models do not suffice.

Deep Learning has gained popularity with remarkable increase of its power, complexity and number of possible applications. Financial time series is no exception. An increasing number of articles are being published by ML specialists, introducing new models and approaches. I will cover the basic and most popular architectures.

The basic deep learning architecture is Artificial Neural Network [?]. It is, broadly speaking, a combination of parametric matrix multiplications with non-linear activation function applied to each of them, stacked together. For each vector of variables (in time series context, lags of the variable of interest) a single value y_{t+1} is produced. Then, it is compared with the true future value using a loss function. Because loss functions and all layer operations are differentiable with respect to parameters, gradients can be calculated, and the parameters get updated in the direction of minimizing the loss, with a procedure called backpropagation. After sufficient number of iterations, the model learns the best parameters.

The simple ANN model, however, is somewhat equivalent to running a linear regression on lagged time series data: it does not account for time dependence and autocorrelation [?]. For this purpose, Recurrent Neural Networks were introduced, first in the field of Natural Language Processing [?]. An RNN is made of a single cell (still a combination of parametric matrix multiplication, concatenation and nonlinear functions) rolling over instances of a time-dependent data. In this architecture an input is a sequence of some pre-defined length, and the model gets applied to every sequence in the data in a rolling window manner.

There are several types of RNN cells, each carefully designed by ML researches to better accomplish prediction tasks. The most widely used is called LSTM, Long Short-Term Memory [?]. Initially designed to deal with the vanishing gradient problem, LSTM can also utilize both short- and long-term temporal dependence by receiving previous step state as input, which makes it very good for time series tasks with autocorrelation of higher orders.

Most of papers in this field employ DL architectures in single-stock forecasting task. A number of basic variables like Open, Close prices, Volume and Volatility of a stock or a stock index are fed into a neural network of some new design, then the results are compared to previously developed models and baselines. Some of the authors use technical indicators [e.g. ? ?]. There is a stream of literature incorporating complex information like tweets, news or Google trends into forecasting framework, using different specific NN architectures for processing that data [? ? ?]. Some just use a very large number of stocks at the same time for training [?].

However, meaningful external macroeconomic factors are rarely used. The closest work in some sense is by ?]. The authors test, whether information on gold and oil prices is suitable for predicting future stock movement (a classification task, contrary to regression task of predicting the future value), using a Convolution Neural Network (CNN). They conclude that this external information enhance forecasts for firms from semiconductor, petroleum, and automotive industries, while not affecting apparel, fast food, and copy processing industries. I plan a somewhat similar approach.

4. Methodology

The task of this proposed study is multivariate time series forecasting, i.e. predicting next day's stock price using a sequence of previous price values as well as different explanatory variables.

Daily data on Russian stocks will be collected from open sources, for example, finam.ru. Period of observation is preliminarily decided to be from 2011 to 2019 years: data on earlier dates is inaccessible, and later dates exhibit COVID-19 structural shock, likely changing the pattern of connections in data, so it would be unfair to test the model on these new conditions. Several Russian publicly traded companies from different sectors will be selected to preform the forecasts.

The following fundamental macroeconomic factors are going to be considered. These variables are freely available in open sources as well.

- Oil, gas, gold prices
- Dollar and Euro exchange rates
- Interest rates on government bonds
- NASDAQ, NYSE, London exchange, etc. indexes

Following the standard machine learning procedure, the data will be sequentially split into training, validation and testing samples. The training

sample will be used to optimize model’s parameters using gradient methods; validation period will be used to tune the model’s hyperparameters, i.e. chose the external, non-optimizable parameters which give the best performance on this validation sample; testing data is for final evaluation and comparison between different models.

The criterion is Root Mean Squared Error (RMSE) metric, squared deviation of predicted value from the actual one.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}} \quad (1)$$

Performance of various models is going to be compared to determine if the most advanced Deep learning models do indeed provide better results. The list of models and baselines currently considered is the following:

- Linear models
 - ARIMA
- Deep Learning models
 - ANN
 - LSTM
- Baselines
 - rolling mean
 - previous value

To analyze variable importance, each group of variables is going to be iteratively excluded from training sample, to compare resulting prediction qualities. If model performance with the factor included is statistically better across multiple trials, the factor will be considered important in a sense of bearing useful information which the model is able to extract.

The work will be conducted in Python, using PyTorch, sklearn packages, and likely many others [? ? ?].

5. Anticipated Results

The result of the study, firstly, is a choice of a best-performing model, in terms of RMSE. It is likely going to be LSTM or some variation of it, though it is not certain, hence the application of other models for comparison is planned. Researches have established that Neural Network approaches generally achieve higher quality predictions compared to linear models like ARIMA [?]. However, sometimes it does not hold for various reasons, like the choice of hyperparameters, suboptimal training procedure or the nature of the data. As training and testing the models on Russian stock has not been done previously, this study can provide an additional example of successful or unsuccessful application of DL in finance.

The other planned result is an interpretation of importance of suggested variables. In principle, the hypothesis for all of them is that they do bear significant information about stock market, and therefore must increase prediction quality. As reviewed literature suggest, most of the factors even have direct causal connection with the variable of interest. However, this does not imply that the variables will be important in this paper’s setting. Determining it can motivate researches to use them in their works.

6. Conclusion

In the proposed study I plan to assess whether a list of different macroeconomic variables can increase quality of predictions on Russian stock market using Deep Learning models. As literature suggests, variables like oil prices, exchange rates, interest rates, foreign market indexes, and others can causally interact with stock prices of Russian companies. However, most of the variables have never been used outside of standard econometric pipelines, and modern Deep Learning models have not been exposed to that kind of information.

Therefore, training DL models on daily time series of Russian stock prices together with supplementary fundamental economic factors, this study aims to determine if those variables are applicable to the task of stock forecasting. It can further strengthen or weaken the hypotheses about their links with the stock market, though not prove or disprove them. The proposed study is not a causal study; improving prediction power does not directly mean that a variable is causally significant, the opposite is not true as well.

Moreover, the results on Russia most likely will not be applicable to other countries or the future Russian stock market. Huge structural shocks can very drastically change the nature of stock markets and disrupt links between factors, rendering analysis inapplicable. However, this provides opportunity for future research to expand proposed methodology on other countries and different periods.

References

- Anari, A., & Kolari, J. (2001). Stock Prices and Inflation. *Journal of Financial Research*, 24(4), 587–602. <https://doi.org/10.1111/j.1475-6803.2001.tb00832.x>
- Bahmani-Oskooee, M., & Saha, S. (2015). On the relation between stock prices and exchange rates: A review article. *Journal of Economic Studies*, 42(4), 707–732. <https://doi.org/10.1108/JES-03-2015-0043>
- Basher, S. A., & Sadorsky, P. (2006). Oil price risk and emerging stock markets. *Global Finance Journal*, 17(2), 224–251. <https://doi.org/10.1016/j.gfj.2006.04.001>
- Basher, S. A., Haug, A. A., & Sadorsky, P. (2012). Oil prices, exchange rates and emerging stock markets. *Energy Economics*, 34(1), 227–240. <https://doi.org/10.1016/j.eneco.2011.10.005>

- Brown, S. P. A., & Yücel, M. K. (2002). Energy prices and aggregate economic activity: An interpretative survey. *The Quarterly Review of Economics and Finance*, 42(2), 193–208. [https://doi.org/10.1016/S1062-9769\(02\)00138-2](https://doi.org/10.1016/S1062-9769(02)00138-2)
- Chen, K., Zhou, Y., & Dai, F. (2015). A LSTM-based method for stock returns prediction: A case study of China stock market. *2015 IEEE International Conference on Big Data (Big Data)*, 2823–2824. <https://doi.org/10.1109/BigData.2015.7364089>
- Chen, S.-W., & Chen, T.-C. (2012). Untangling the non-linear causal nexus between exchange rates and stock prices: New evidence from the OECD countries. *Journal of Economic Studies*, 39(2), 231–259. <https://doi.org/10.1108/01443581211222671>
- Chen, Y.-C., & Huang, W.-C. (2021). Constructing a stock-price forecast CNN model with gold and crude oil indicators. *Applied Soft Computing*, 112, 107760. <https://doi.org/10.1016/j.asoc.2021.107760>
- Degiannakis, S., Filis, G., & Arora, V. (2018). Oil Prices and Stock Markets: A Review of the Theory and Empirical Evidence. *The Energy Journal*, 39(01). <https://doi.org/10.5547/01956574.39.5.sdeg>
- Durai, S. R. S., & Bhaduri, S. N. (2009). Stock prices, inflation and output: Evidence from wavelet analysis. *Economic Modelling*, 26(5), 1089–1092. <https://doi.org/10.1016/j.econmod.2009.04.005>
- Eldomiaty, T., Saeed, Y., Hammam, R., & AboulSoud, S. (2019). The associations between stock prices, inflation rates, interest rates are still persistent: Empirical evidence from stock duration model. *Journal of Economics, Finance and Administrative Science*, 25(49), 149–161. <https://doi.org/10.1108/JEFAS-10-2018-0105>
- Elyasiani, E., Mansur, I., & Odusami, B. (2011). Oil price shocks and industry stock returns. *Energy Economics*, 33(5), 966–974. <https://doi.org/10.1016/j.eneco.2011.03.013>
- Filis, G. (2010). Macro economy, stock market and oil prices: Do meaningful relationships exist among their cyclical fluctuations? *Energy Economics*, 32(4), 877–886. <https://doi.org/10.1016/j.eneco.2010.03.010>
- Filis, G., Degiannakis, S., & Floros, C. (2011). Dynamic correlation between stock market and oil prices: The case of oil-importing and oil-exporting countries. *International Review of Financial Analysis*, 20(3), 152–164. <https://doi.org/10.1016/j.irfa.2011.02.014>
- Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10), 2222–2232. <https://doi.org/10.1109/TNNLS.2016.2582924>
- Hashemzadeh, N., & Taylor, P. (1988). Stock prices, money supply, and interest rates: The question of causality. *Applied Economics*, 20(12), 1603–1611. <https://doi.org/10.1080/000368488000000091>
- Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent Neural Networks for Time Series Forecasting: Current status and future directions.

- International Journal of Forecasting*, 37(1), 388–427. <https://doi.org/10.1016/j.ijforecast.2020.06.008>
- Hu, H., Tang, L., Zhang, S., & Wang, H. (2018). Predicting the direction of stock markets using optimized neural networks with Google Trends. *Neurocomputing*, 285, 188–195. <https://doi.org/10.1016/j.neucom.2018.01.038>
- Huang, J.-Y., & Liu, J.-H. (2020). Using social media mining technology to improve stock price forecast accuracy. *Journal of Forecasting*, 39(1), 104–116. <https://doi.org/10.1002/for.2616>
- Inci, A. C., & Lee, B. S. (2014). Dynamic Relations between Stock Returns and Exchange Rate Changes. *European Financial Management*, 20(1), 71–106. <https://doi.org/10.1111/j.1468-036X.2011.00621.x>
- Kollias, C., Mylonidis, N., & Paleologou, S.-M. (2012). The nexus between exchange rates and stock markets: Evidence from the euro-dollar rate and composite European stock indices using rolling analysis. *Journal of Economics and Finance*, 36(1), 136–147. <https://doi.org/10.1007/s12197-010-9129-8>
- Kordonis, J., Symeonidis, S., & Arampatzis, A. (2016). Stock Price Forecasting via Sentiment Analysis on Twitter. *Proceedings of the 20th Pan-Hellenic Conference on Informatics*, 1–6. <https://doi.org/10.1145/3003733.3003787>
- Kumar, M., & Thenmozhi, M. (2006). *Forecasting Stock Index Movement: A Comparison of Support Vector Machines and Random Forest* (SSRN Scholarly Paper No. ID 876544). Social Science Research Network. Rochester, NY. <https://doi.org/10.2139/ssrn.876544>
- Lara-Benítez, P., Carranza-García, M., & Riquelme, J. C. (2021). An Experimental Review on Deep Learning Architectures for Time Series Forecasting. *International Journal of Neural Systems*, 31(03), 2130001. <https://doi.org/10.1142/S0129065721300011>
- Li, H., Shen, Y., & Zhu, Y. (2018). Stock Price Prediction Using Attention-based Multi-Input LSTM. *Proceedings of The 10th Asian Conference on Machine Learning*, 454–469.
- Lozinskaia, A., & Saltykova, A. (2019). *Fundamental Factors Affecting The Moex Russia Index: Structural Break Detection In A Long-Term Time Series* (tech. rep. WP BRP 77/FE/2019). National Research University Higher School of Economics.
- Mohanty, S. K., Nandha, M., Turkistani, A. Q., & Alaitani, M. Y. (2011). Oil price movements and stock market returns: Evidence from Gulf Cooperation Council (GCC) countries. *Global Finance Journal*, 22(1), 42–55. <https://doi.org/10.1016/j.gfj.2011.05.004>
- Narayan, P. K., & Narayan, S. (2010). Modelling the impact of oil prices on Vietnam’s stock prices. *Applied Energy*, 87(1), 356–361. <https://doi.org/10.1016/j.apenergy.2009.05.037>
- Narayan, P. K., & Sharma, S. S. (2011). New evidence on oil price and firm returns. *Journal of Banking & Finance*, 35(12), 3253–3262. <https://doi.org/10.1016/j.jbankfin.2011.05.010>

- Nelson, D. M. Q., Pereira, A. C. M., & de Oliveira, R. A. (2017). Stock market's price movement prediction with LSTM neural networks. *2017 International Joint Conference on Neural Networks (IJCNN)*, 1419–1426. <https://doi.org/10.1109/IJCNN.2017.7966019>
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., . . . Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d. Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 32* (pp. 8024–8035). Curran Associates, Inc.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Raza, N., Jawad Hussain Shahzad, S., Tiwari, A. K., & Shahbaz, M. (2016). Asymmetric impact of gold, oil prices and their volatilities on stock prices of emerging markets. *Resources Policy*, 49, 290–301. <https://doi.org/10.1016/j.resourpol.2016.06.011>
- Robert D. Gay, J. (2008). Effect Of Macroeconomic Variables On Stock Market Returns For Four Emerging Economies: Brazil, Russia, India, And China. *International Business & Economics Research Journal (IBER)*, 7(3). <https://doi.org/10.19030/iber.v7i3.3229>
- Sadorsky, P. (2014). Modeling volatility and correlations between emerging market stock prices and the prices of copper, oil and wheat. *Energy Economics*, 43, 72–81. <https://doi.org/10.1016/j.eneco.2014.02.014>
- Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2019). Financial Time Series Forecasting with Deep Learning : A Systematic Literature Review: 2005-2019. *arXiv:1911.13288 [cs, q-fin, stat]*.
- Siami-Namini, S., & Namin, A. S. (2018). Forecasting Economics and Financial Time Series: ARIMA vs. LSTM. *arXiv:1803.06386 [cs, q-fin, stat]*.
- Siami-Namini, S., Tavakoli, N., & Siami Namin, A. (2018). A Comparison of ARIMA and LSTM in Forecasting Time Series. *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 1394–1401. <https://doi.org/10.1109/ICMLA.2018.00227>
- Silvapulle, P., Smyth, R., Zhang, X., & Fenech, J.-P. (2017). Nonparametric panel data model for crude oil and stock market prices in net oil importing countries. *Energy Economics*, 67, 255–267. <https://doi.org/10.1016/j.eneco.2017.08.017>
- Torres, J. F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F., & Troncoso, A. (2021). Deep Learning for Time Series Forecasting: A Survey. *Big Data*, 9(1), 3–21. <https://doi.org/10.1089/big.2020.0159>
- Tsagkanos, A., & Siriopoulos, C. (2013). A long-run relationship between stock price index and exchange rate: A structural nonparametric cointegrat-

- ing regression approach. *Journal of International Financial Markets, Institutions and Money*, 25, 106–118. <https://doi.org/10.1016/j.intfin.2013.01.008>
- Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.
- Verma, R. K., & Bansal, R. (2021). Impact of macroeconomic variables on the performance of stock exchange: A systematic review. *International Journal of Emerging Markets*, 16(7), 1291–1329. <https://doi.org/10.1108/IJOEM-11-2019-0993>
- Wang, Y., Wu, C., & Yang, L. (2013). Oil price shocks and stock market activities: Evidence from oil-importing and oil-exporting countries. *Journal of Comparative Economics*, 41(4), 1220–1239. <https://doi.org/10.1016/j.jce.2012.12.004>
- Zhao, X.-Q. (1999). Stock prices, inflation and output: Evidence from China. *Applied Economics Letters*, 6(8), 509–511. <https://doi.org/10.1080/135048599352835>
- Zhu, H.-M., Li, R., & Li, S. (2014). Modelling dynamic dependence between crude oil prices and Asia-Pacific stock market returns. *International Review of Economics & Finance*, 29, 208–223. <https://doi.org/10.1016/j.iref.2013.05.015>