Expert Systems With Applications 194 (2022) 116553

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Are CDS spreads predictable during the Covid-19 pandemic? Forecasting based on SVM, GMDH, LSTM and Markov switching autoregression

Darko B. Vukovic [a,b,1,2,3,*], Kirill Romanyuk [c,4,5,6], Sergey Ivashchenko [d,e,f,7,8,9], Elena M. Grigorieva [a,g,10]

[a] International Laboratory for Finance and Financial Markets, Faculty of Economics, People's Friendship University of Russia (RUDN University), 117198, Miklukho-Maklaya str. 6, Moscow, Russia
[b] Geographical Institute "Jovan Cvijic" SASA, Djure Jaksica 9, 11000 Belgrade, Serbia
[c] HSE University, 101000, Myasnitskaya str., 20, Moscow, Russia
[d] North-Western Main Branch of the Bank of Russia, 191038, Saint Petersburg Fontanka River Embankment, 68 (107016, Moscow, 12 Neglinnaya Street), Russia
[e] Macroeconomic Analysis Division, Financial Research Institute, 127006, Moscow, Nastasyinsky Lane, 3, p. 2, Russia
[f] Theoretical Economics Laboratory, The Institute of Regional Economy Studies (Russian Academy of Sciences), 190013, Saint Petersburg, Serpukhovskaya Street 36-38, Russia
[g] Finance and Credit Department, Faculty of Economics, People's Friendship University of Russia (RUDN University), 117198, Miklukho-Maklaya str.6, Moscow, Russia

ARTICLE INFO

Keywords:
CDS spreads
SVM
GMDH
LSTM
Markov switching autoregression
Covid-19

ABSTRACT

This paper investigates the forecasting performance for credit default swap (CDS) spreads by Support Vector Machines (SVM), Group Method of Data Handling (GMDH), Long Short-Term Memory (LSTM) and Markov switching autoregression (MSA) for daily CDS spreads of the 513 leading US companies, in the period 2009–2020. The goal of this study is to test the forecasting performance of these methods before and during the Covid-19 pandemic and to check whether there are changes in the market efficiency. MSA outperforms all other methods most frequently. GMDH breaks the efficient market hypothesis more frequently (75%) than other methods. The change of the relative predictability during Covid-19 is small with some increase of the advantage of the investigated methods over a benchmark. We find that the market has been less efficient during Covid-19, however, there are no huge differences in prediction performances before and during the Covid-19 period.

1. Introduction

Forecasting financial time series aims to anticipate predictable patterns that will bring investors advantage in trading opportunities. In financial theory, this issue is explained by the efficient market theory defined by Fama (1955, 1970), Jensen (1978) and later extended by Malkiel (1992). However, the efficient market theory refers to "real time" opportunities and stable forecasting patterns will become redundant after their discovery by a large number of players (Timmermann and Granger, 2004). According to Timmermann and Granger (2004), testing events with the efficient market theory has significance (and growing popularity) for non-stationarity in financial time series. The

* Corresponding author.
E-mail addresses: vdarko@hotmail.rs (D.B. Vukovic), kromanyuk@hse.ru (K. Romanyuk), sergey.ivashchenko.ru@gmail.com (S. Ivashchenko), aroooveo@yandex.ru (E.M. Grigorieva).
[1] ORCID: 0000-0002-1165-489X.
[2] ResearcherID: G-9374-2016.
[3] Scopus Author ID: 57031563800.
[4] ORCID: 0000-0002-5420-2607.
[5] ResearcherID: F-8443-2018.
[6] Scopus Author ID: 56825751900.
[7] ORCID: 0000-0002-4248-7430.
[8] ResearcherID: N-1822-2015.
[9] Scopus Author ID: 55624584600.
[10] ORCID: 0000-0002-8345-5152.

https://doi.org/10.1016/j.eswa.2022.116553
Received 26 February 2021; Received in revised form 1 January 2022; Accepted 14 January 2022
Available online 22 January 2022
0957-4174/

forecasting of systematic patterns is mostly related to predicting the prices of various assets, however, only a few studies forecast credit default swaps (CDS). Such forecasting, which has a non-stationary nature, is very difficult (Henrique et al., 2019; Zhang et al., 2017), because it is chaotic, dynamic, and nonlinear.

Forecasting the patterns of CDS also aims to break market efficiency caused by a certain process or event. A number of researchers believe that certain events or processes in the future may impair unbiased market efficiency (Lee et al., 2020), with positive (negative) deviations affecting large positive (negative) excess returns (Abe and Nakayama, 2018) or a change in spreads caused by economic shocks (Gilchrist et al., 2016). From March 2020, the financial markets have had a new shock, caused by the global Covid-19 pandemic. This huge shock and its transmission mechanism may break market efficiency and/or existing patterns in markets. Considering that the SVM, GMDH, LSTM and MSA methods show the best performance in CDS spread forecasting (see discussion in section 2), the goal of this study is to discover outperforming methods before and during a crisis event (Covid-19 pandemic). To test this, we analyze CDS spread time series data of the leading 513 US companies from 2009 to 2020. Our study contributes to existing literature to get a better understanding of CDS spread forecasting and practically as an important tool for investment decision making. The novelty of the study lies in the fact that this is the first study of forecasting CDS spreads by testing these 4 methods, especially comparing MSA to the other three methods based on kernel or artificial intelligence and testing their predictive power before and during the Covid-19 pandemic. It is aimed to encourage future research in this area, especially given the financial crisis associated with credit derivatives and recent uncertainty in financial markets caused by the Covid-19 pandemic. Given the popularity of these methods, in this study we aim to answer the following: Do these methods have satisfactory performance in CDS spread forecasting? Do they outperform the random walk, which is impossible for efficient markets (possible brakes in efficient market theory)? Does the Covid-19 (as a process or event) affect the performance of these methods?

2. Related work

Numerous studies have tested different prediction models for time series data during Covid-19. Katris (2021) used classical (Exponential Smoothing and ARIMA) and the machine learning (Feed-Forward Artificial Neural Networks and Multivariate Adaptive Regression Splines) models for time series Covid-19 outbreak forecasting in Greece. He forecast with the Newbolt/Granger grouping scheme of models, to build a log-normal distribution of the downward trend of the outbreak. Gaglione et al. (2020) forecast time series of the Covid-19 outbreak in Italy, with the Bayesian sequential estimation and forecasting algorithm. They found that the Bayesian framework outperforms curve-fitting approaches. Pereira et al. (2020) used LSTM with modified Auto-Encoder networks to predict the Covid-19 outbreak in Brazil. According to their results, this method performs better than single LSTM. On the other hand, Chimmula and Zhang (2020) found that the LSTM network provides satisfactory performance in forecasting time series data during the Covid-19 pandemic. The authors develop a deep-learning model to forecast the Covid-19 outbreak in Canada. Peng and Nagata (2020) used SVM to forecast Covid-19 time series data in the 12 most-affected countries. They found that SVM with a kernel function shows the best performance in nonlinear out-of-sample forecasts.

In addition to studies that forecast the Covid-19 pandemic, several studies have focused on financial time series data during the Covid-19 pandemic. Given that Covid-19 is an event that affects players who fail to form rational expectations or fail to adjust their trading strategy given their expectations, the Covid-19 event has become an attractive topic in financial forecasting. For example, Ghosh & Chaudhuri (2021) applied kernel principal component analysis to refine several technical and macroeconomic indicators for separate forecasting by stacking and deep neural network models. They found that both models perform well before and during the Covid-19 pandemic. Colladon et al. (2020) used a textual data index to forecast return and volatility dynamics of financial variables in the Italian market. According to the authors' results, this model is very accurate at forecasting bond returns and volatilities during Covid-19.

The question still arises: Which models proved to be the most accurate in forecasting financial time series, including CDS spreads? Jia et al. (2020) tested a number of forecasting models based on machine learning and deep learning techniques, or metamodels to predict future values. One of the most widely used machine learning methods in forecasting is the artificial intelligence method (ANN) based on neurons and emulates the human system of learning through pattern identification (Vyklyuk et al., 2013; Laboissiere et al., 2015; Vukovic et al., 2020). However, according to Li et al. (2020), LSTM as a deep learning method, outperforms ANN in addressing the long-term dependence problem. It has become a very popular method in financial market forecasting because of the characteristics that are very efficient and predictable in occurrences, having significant deterrent qualities. Lee et al. (2021) claim that models based on deep learning have greater forecasting power than regression or time series models, that use conventional statistical methods. The same authors find that LSTM is a feasible option for predicting trends of commercial vacancy in commercial districts. Peng et al. (2020) argue that accuracy and performance power of LSTM are influenced of parameters and optimization. To increase the accuracy and power of LSTM model prediction, these authors use fruit fly optimization algorithm.

Another popular method in financial time series forecasting is the kernel-based SVM method. The main difference between artificial intelligence methods and SVM is in the fact that the first one minimizes the errors in the training stage, while SVM minimizes the upper threshold of the error of its classification (Henrique et al., 2019). SVM is a very popular method in time series prediction due to its superior learning ability (Zhou et al., 2016). Tobback et al. (2018) tested CDS spreads as part of ten different variables by using 210,000 articles for text mining and applying support vector machines (SVM) with a linear kernel, modality annotation and constructing an Economic Policy Uncertainty index for Belgium. They found that the SVM method has higher predictive power. The next most used method is Group Method of Data Handling (GMDH), developed by Ivakhnenko (1988). Even though this model is old and has been used over the last thirty years, many authors use GMDH because it is particularly suitable for the cases when little data are available (and no *a priori* information is available) on the structure of the mechanism generating the series. According to Yang et al. (2009), GMDH is a more efficient model for prediction compared to artificial intelligence methods and statistical models for a given level of data. Finally, the MSA model, as an effective predictive technique, is used in many studies. MSA has become a very popular model due to its features capturing the schematic behavior of financial return series, fat tails, continually arising periods of instability with low volatility, skewness, and time-varying correlations (Ang & Timmerman, 2012; Stutzer, 2020).

In addition to previous studies, the motivation of our research is also supported by the next several studies (concerning CDS spread forecasting)[11]:

1. Hu et al. (2019) used a series of machine learning techniques to test their prediction power for monthly CDS spreads of 69 companies, for the period from 2006 to 2016. The study is based on two parametric machine learning methods (MLM) (LASSO and Ridge) and six nonparametric MLM (Neural Network, Support Vector Regression, Bagging, Regression Tree, Gradient Boosting and Random Forest). According to the authors, nonparametric MLM outperform all other

[11] The discussed studies are presented in Table 1, with the contribution of our research concerning related work and theory.

methods.

2. Kim et al. (2020) analyzed and forecast time-series (2008–2019) of term structure of CDS spreads by using the Nelson–Siegel model, recurrent neural network, support vector regression (SVR), LSTM, and GMDH. They authors found that RNN, SVR, LSTM and GMDH outperform the Nelson–Siegel model, while GMDH was the most accurate in the prediction.

3. Fei et al. (2017) found a time-varying dependence between CDS spreads and equity prices. The authors empirically proved that the MSA bivariate copula model is an appropriate measure of forecasting (in crisis dependence and low dependence for in-sample statistical criteria).

4. Avino and Nneji (2014) forecast CDS spreads using linear and non-linear models, studying the iTraxx Europe index during the 2008 financial crisis. They noted that the Markov switching models underperform other linear models.

3. Methods

3.1. Support vector Machines

SVM can be viewed as a supervised learning approach for data analysis. The idea behind SVM is to find a hyperplane that maximizes the distance between classes. SVM for a regression can be expressed as the following optimization task (1).

$$\begin{gathered}\min_{w,b,\zeta}\frac{1}{2}w^Tw + C\sum\nolimits_{i=1}^{n}\zeta_i \\ \text{subject to } |y_i - w^T\phi(x_i) - b| \leq \epsilon + \zeta_i, \\ \zeta_i, \zeta_i^* \geq 0, i = 1, 2, \cdots n\end{gathered} \tag{1}$$

where $x_i \in R^p, i = 1,2,...,n$, and $y \in R^n$, y_i is the true outcome for x_i input and $\widehat{y_i} = w^T\phi(x_i) + b$ we can refer to as predicted outcome for x_i input. ζ_i is the penalty if the difference between the prediction $w^T\phi(x_i) + b$ and the target y_i is greater than the allowed distance ϵ. $\phi(\cdot)$ is a function that is applied when a linear separation of observations is not possible. In addition, it is common to use the radial basis function in SVM for time series analysis (Liu et al., 2011).

3.2. Group method of data handling

The group method of data handling is an inductive self-organizing method for solving data analysis tasks. A function for GMDH can be generally viewed as a combination of elementary functions (f_i) and coefficients (2). However, GMDH is usually based on polynomial reference function such as Kolmogorov-Gabor polynomial (3). Different reference functions can be also used, e.g. harmonic and logistic (Xu et al., 2021).

$$Y(x_1, x_2, \cdots, x_n) = a_0 + \sum\nolimits_{i=1}^{m} a_i f_i \tag{2}$$

$$Y(x_1, x_2, \cdots, x_n) = a_0 + \sum\nolimits_{i=1}^{m} a_i x_i + \sum\nolimits_{i=1}^{m}\sum\nolimits_{j=1}^{m} a_{ij}x_ix_j + \sum\nolimits_{i=1}^{m}\sum\nolimits_{j=1}^{m}\sum\nolimits_{k=1}^{m} a_{ijk}x_ix_jx_k + \cdots \tag{3}$$

3.3. Long Short-Term Memory

Long Short-Term Memory is a type of Recurrent Neural Network in which short-term and long-term items are stored and applied effectively. LSTM models are often applied in time series analysis. Within LSTM networks, cells form units, and units form layers. Each cell has input, output, and forget gate. Equations (4)-(8) "show the form of the forward pass of the LSTM unit (x_t : input vector to the LSTM unit, f_t : forget gate's activation vector, i_t : input gate's activation vector, o_t : output gate's activation vector, h_t : output vector of the LSTM unit, c_t : cell state vector, δ_q : sigmoid function, δ_c , δ_h: hyperbolic tangent function, *: element-wise (Hadamard) product, W, U: weight matrices to be learned, b: bias vector parameters to be learned)" (Sezer et al., 2020).

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \tag{4}$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \tag{5}$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \tag{6}$$

$$c_t = f_t * c_{t-1} + i_t * \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \tag{7}$$

$$h_t = o_t * \sigma_h(c_t) \tag{8}$$

3.4. Markov switching autoregression model

Markov switching autoregression can be viewed as a set of autoregressions connected with regime specific values, i.e. the number of autoregressions is equal to the number of regimes. For example, let s_t denote an unobservable state variable assuming the value one or zero. A simple switching model for the variable z_t involves two AR specifications (9):

$$z_t = \begin{cases} \alpha_0 + \beta z_{t-1} + \epsilon_t, & s_t = 0 \\ \alpha_0 + \alpha_1 + \beta z_{t-1} + \epsilon_t, & s_t = 1 \end{cases} \tag{9}$$

where $|\beta| < 1$ and ϵ_t are i.i.d. random variables with mean zero and variance σ_ϵ^2. This is a stationary AR(1) process with mean $\frac{\alpha_0}{1-\beta}$ when $s_t = 0$, and it switches to another stationary AR(1) process with mean $\frac{\alpha_0+\alpha_1}{1-\beta}$ when s_t changes from 0 to 1 (Kuan, 2002).

4. Results

In this section, SVM, GMDH, LSTM and MSA models are employed to analyze forecasting performance for CDS spreads before and during the Covid-19 pandemic, and to check whether efficiency of market changes due to Covid-19.

4.1. Data description

The dataset that is analyzed in this study consists of time series obtained from Bloomberg terminal. The time horizon for the training and testing is from 28 October 2009 to 16 October 2020, 2799 observations in total. We observed daily CDS spreads (1-year, 4-year, 7-year, and 10-year) for 513 US all ratings (investment and high yields, debt type - senior) companies from multiple sectors. Fig. 1 shows the times series plots of CDS spread average levels for all maturities. The dataset is created similarly like in the following studies: Avino and Nneji (2014), Kim et al. (2020), and Buse and Schienle (2019).

It can be seen on Fig. 1 that Covid-19 is not the most volatile period for CDS, and training set includes very different dynamic of CDS. The parameters of the models are estimated on training set only. The performance is computed on tests sets. It is very restrictive (tight) form of out-of-sample forecasting tests. Moreover, we use default values of the model's hyper-parameters (such as layers number for LSTM). It prevents any possibility of out-of-sample data influence on choice of the hyper-parameters (for example via criteria choice). However, it creates disadvantages for more flexible approaches due to missed possibility of their improvement (for example, LSTM can be greatly improved according to Peng et al. (2020), Lee et al. (2021).

The Fig. 2 describes how we combine data and used methods/ models. Training set starts from 28 October 2009 till 6 December 2018. Three test sets of the same size (158 observations) were formed in order to show how the above-mentioned methods will perform during Covid-19 pandemic period in comparison to periods before the pandemic. The World Health Organization declared pandemic due to coronavirus on 11th March 2020. The first test set starts from 7 December 2018 till 19

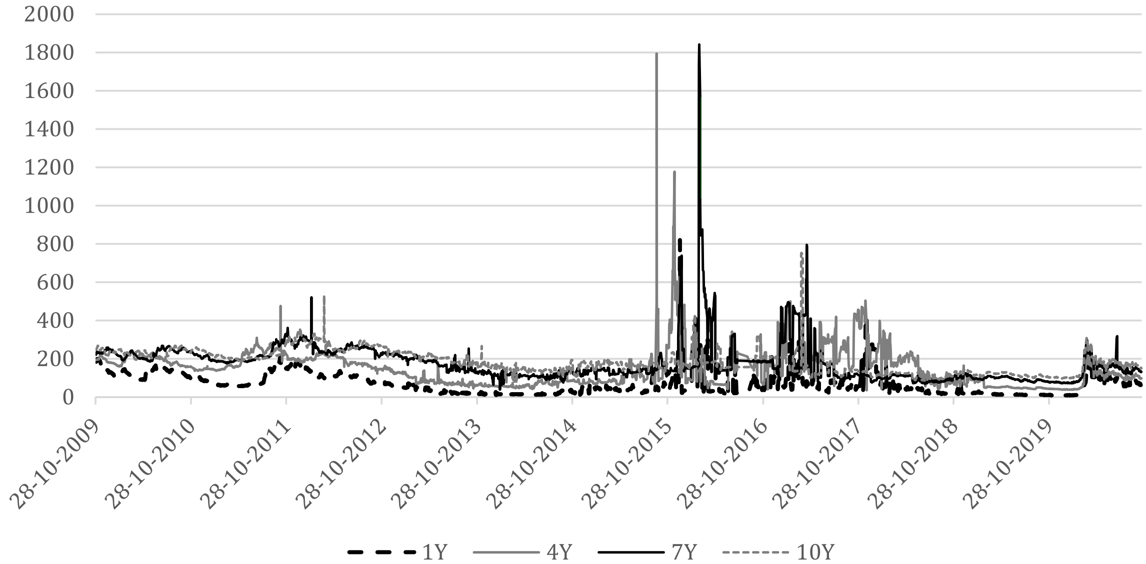

Fig. 1. Time series plots of CDS spread sample average levels for all maturities. This figure shows time series of daily CDS spreads from 1-year to 10-year maturities, over the period 10/28 /2009 to 10 /16 /2020.

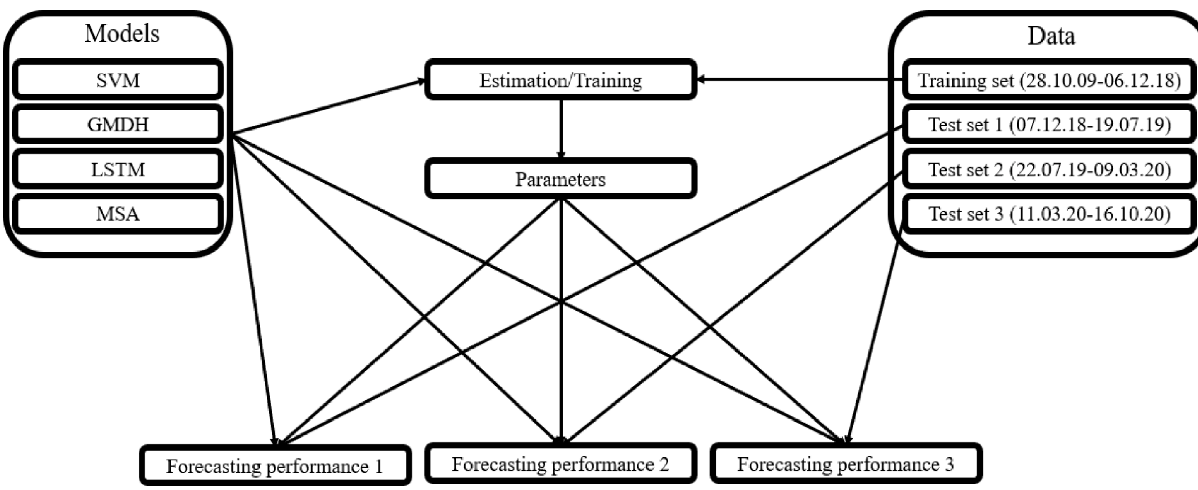

Fig. 2. A methodological framework.

July 2019, the second one from 22 July 2019 till 9 March 2020, and the third one from 11 March 2020 till 16 October 2020.

The following software were used for handling calculations:

- Sklearn python package was used for SVM. More specifically, SVR method was used from this package with radial basis function as a kernel and gamma = 0.1.
- Gmdh.py python package was used for GMDH. More specifically, multilayer GMDH method was used from this package.
- Tensorflow python package with Keras API was used for LSTM. LSTM had 100 epochs. RobustScaler method were used from sklearn package for scaling the data.
- Eviews 11 was used for building MSA. AR (10) model with 2 states (namely 'normal' and 'crisis' states or regimes (Fei et al., 2017)) was taken as the basis for CDS forecasting. $\hat{x}_t$ is predicted by the first equation if economy is in 'normal' state ($s_t = 0$) and by the second equation if economy is in 'crisis' state ($s_t = 1$) (13).

$$\hat{x}_t = \begin{cases} b_1{}^*x_{t-1} + b_2{}^*x_{t-2} + \cdots + b_{10}{}^*x_{t-10} + \epsilon_t, & \text{if } s_t = 0 \\ c_1{}^*x_{t-1} + c_2{}^*x_{t-2} + \cdots + c_{10}{}^*x_{t-10} + \epsilon_t, & \text{if } s_t = 1 \end{cases} \tag{13}$$

4.2. Performance

LSTM, SVM and GMDH produce point forecasts, and we use the main measures to estimate the efficiency of such forecasts: root mean square deviation (10), mean absolute error (11), mean absolute percentage error (12). Random walk is used as a benchmark. If an error measure is lower for a forecast than for the benchmark, then this forecast is effective, and such an instance is marked with an asterisk (Table 2).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{t+i} - x_{t+i})^2} \tag{10}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|x_{t+i} - \hat{x}_{t+i}| \tag{11}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}|(x_{t+i} - \hat{x}_{t+i})/x_{t+i}| \tag{12}$$

The results show the consistency of the methods. If a method outperforms the benchmark during the normal periods (test set 1, test set 2),

Table 1
Related work and the contribution of the study.

The efficient market theory	The break in market efficiency	The prediction models during Covid-19	Methodological framework				CDS spread forecasting
			LSTM	SVM	GMDH	MSA	
Fama (1955, 1970), Jensen (1978) Malkiel (1992), Timmermann and Granger (2004).	Gilchrist et al. (2016), Abe and Nakayama, (2018), Lee et al. (2020)	Gaglione et al. (2020), Pereira et al. (2020), Chimmula and Zhang (2020), Colladon et al. (2020), Katris (2021), Ghosh and Chaudhuri (2021).	Li et al. (2020), Peng et al. (2020), Lee et. al. (2021).	Zhou et al., (2016), Tobback et al. (2018), Henrique et al. (2019), Stuzer, (2020).	Ivakhnenko (1988), Yang et al. (2009).	Ang & Timmerman, (2012), Stuzer, (2020).	Avino and Nneji (2014), Fei et al. (2017), Hu et al. (2019), Kim et al. (2020).
The contribution of study							
Lower market efficiency (possible breaks) from March 2020. Results in line with Gaglione et al. (2020) and Katris (2021).		Results in line with Ghosh and Chaudhuri, (2021).	Satisfactory performance (better forecasting performance of longer period data).	GMDH outperforms SVM.	GMDH is the most stable in terms of outperforming random walk.	MSA outperforms all other methods more frequently than the others.	Results in line with Hu et al. (2019), Kim et al. (2020).

Note: Table 1 presents related theory and work on which study is based. The lower part of the table represents the contribution of the study in relation to each row of the table.

Table 2
Performance.

		Test Set 1			Test Set 2			Test Set 3		
		RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1Y	SVM	2.037	1.324	0.08	0.751	0.538	0.05	18.381	12.011	0.121
	GMDH	2.018	1.298	0.079	0.748	0.544	0.05	16.257	10.82	0.112
	Markov	1.852*	1.179*	0.072*	0.705*	0.496*	0.045*	14.692*	9.707*	0.1*
	LSTM	2.585	1.535	0.09	0.736	0.553	0.052	15.454	9.95	0.103
4Y	SVM	5.872*	2.217*	0.029*	1.291*	0.615*	0.013*	12.096*	7.735*	0.062*
	GMDH	5.869*	2.214*	0.029*	1.292*	0.615*	0.013*	12.094*	7.736*	0.062*
	Markov	5.994	2.474	0.033	1.345	0.649	0.014	12.306	8.023	0.064
	LSTM	6.096	2.645	0.036	1.295*	0.754	0.015	11.956*	7.638*	0.061*
7Y	SVM	1.818	1.166	0.012	2.849	1.554*	0.018*	21.948	10.154	0.06*
	GMDH	1.78*	1.141*	0.011*	2.848*	1.555*	0.018*	21.947*	10.154*	0.06*
	Markov	1.899	1.24	0.013	3.012	1.6	0.018*	21.297*	10.182	0.06*
	LSTM	3.939	2.154	0.022	3.038	1.856	0.022	23.32	12.486	0.075
10Y	SVM	2.046*	1.261*	0.01*	2.877*	1.624*	0.014*	12.058*	8.487*	0.046*
	GMDH	2.066*	1.272*	0.011*	2.89*	1.632*	0.014*	12.066*	8.495*	0.046*
	Markov	2.057*	1.297	0.01*	2.974	1.658	0.015	11.77*	8.195*	0.044*
	LSTM	2.136	1.348	0.012	2.858*	1.647	0.015	12.194	8.623	0.046*

Note: Table 2 presents measures calculated using SVM, GMDH, MSA, and LSTM forecasts for 1-year, 4-year, 7-year, and 10-year CDS spreads in 3 periods (i.e. test set 1 and test set 2 are pre Covid-19 while test set 3 takes place during Covid-19); *The error value for this forecast is lower than for the benchmark.

they most likely outperform the benchmark during the Covid-19 period (test set 3). There are only 2 notable exceptions for the opposite sentence: MSA for 7Y and LSTM for 4Y lose during normal times but win during Covid-19. The forecasting errors are larger in the Covid-19 period (test set 3) but relative performance is generally sustained. Moreover, the share of cases outperforming random walk is growing slightly during Covid-19. It means that the market became less efficient during extraordinary periods (Fig. 3). The forecasting techniques give better predictions of CDS spreads for longer maturities in the Covid-19 period (see Table 1). GMDH is the most stable in terms of outperforming random walk (75% of cases). MSA outperforms all other methods more frequently than the others (16 cases for MSA and 13 for GMDH).

Fig. 3 reports lower market efficiency (possible breaks) for all four CDS spreads from March 2020. This situation recognizes the possibility to predict systematic patterns, which gives the investor the opportunity to gain extra returns (in line with studies of: Hu et al., 2019; Kim et al., 2020). Moreover, forecasts are more accurate for the longer term CDS spreads (according to Table 2). This situation can be explained by the fact that during a crisis, the demand and volatility of short-term instruments increases. Considering this, the accuracy of forecasting financial market instruments also depends on their maturities, which many studies do not take into account (they usually analyze one or more instruments with the same maturity). Lastly, test sets 1,2 and 3 (Table 2) reports similar accuracy in forecasting, which indicates that predictions during Covid-19 do not affect higher forecasting performance (they are similar and satisfactory for pre and during Covid-19 period; this is in line with study of Ghosh & Chaudhuri, 2021).

5. Discussion

Our results are in accordance with Kim et al. (2020) where GMDH outperforms SVM and LSTM. However, these authors did not test MSA, which has previously been shown to have better performance compared to SVM and LSTM in similar time-series data (Fei et al., 2017). From a risk management perspective, our results support Ardia et al. (2018), showing that MSA is a good method for forecasting financial time series. In the case of Covid-19, we also show the possibilities of time series data forecasting, indicating breaks, which is line of studies of Gaglione et al. (2020) and Katris (2021). However, our results show better forecasting performance of longer period data (seven and ten years CDS) compared to study of Gaglione et al. (2020) (where the authors find better performance for a shorter time interval). We can also conclude that our results are in line with LSTM's satisfactory performance in the findings of Chimmula and Zhang (2020) and Peng and Nagata (2020) for SVM performance. Lastly, our findings are most similar to results of Ghosh & Chaudhuri (2021), where methods show satisfactory prediction

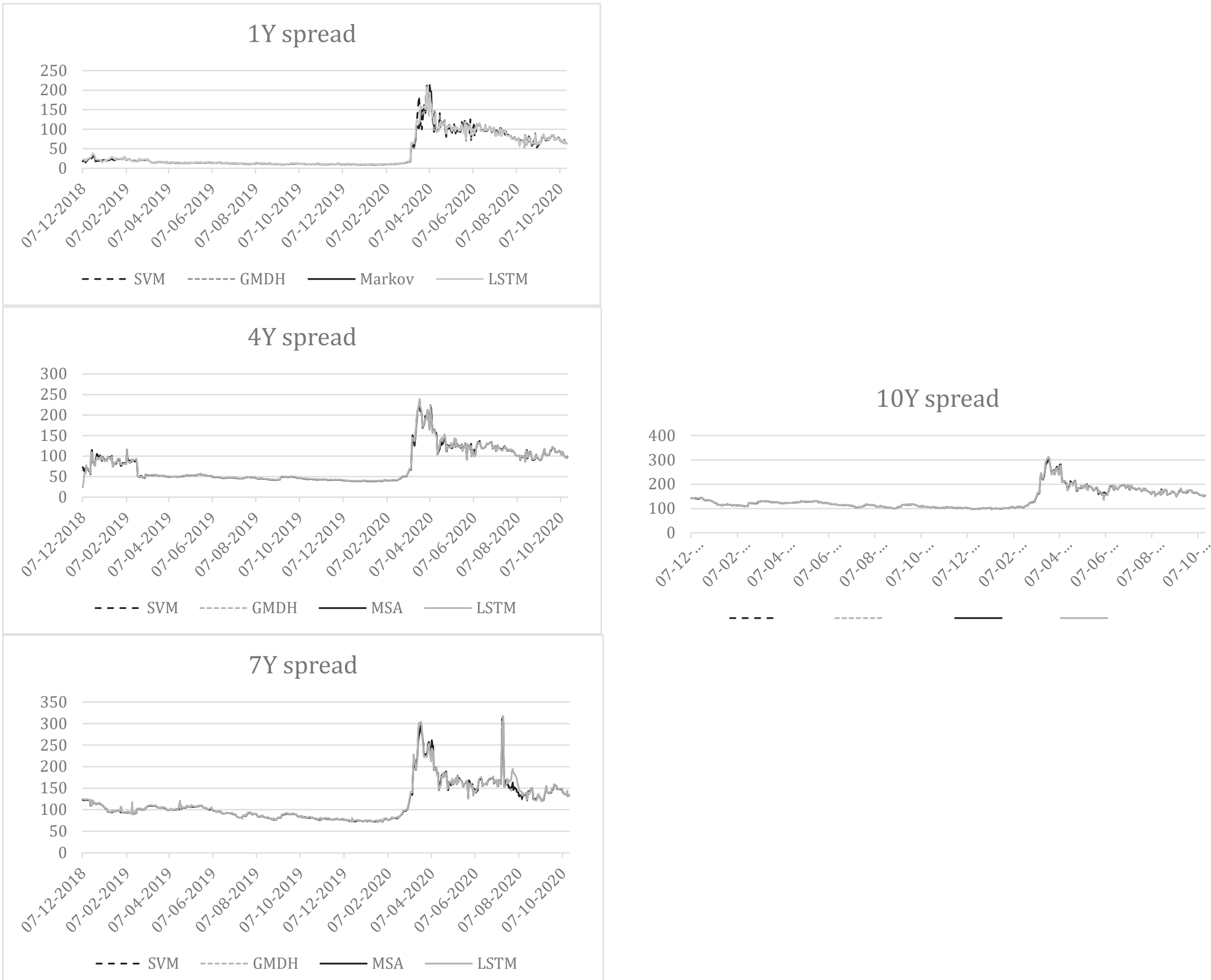

Fig. 3. Forecasts.

performances before and during the Covid-19 pandemic. Ghosh & Chaudhuri (2021) tested only a kernel-based method, while our study shows satisfactory performance of both machine learning and deep learning techniques.

In the period of increased volatility and potential structural breaks, forecasting accuracy appears to be driven by the company's nonlinear specific credit risk. We find that the market has become less efficient during the Covid-19 pandemic, which gives the investor the opportunity to forecast systematic patterns related to CDS spreads. There are no huge differences in prediction performances before and during the Covid-19 pandemic. In both periods, our tested methods show the possibility of predicting systematic patterns, where breaks in market efficiency are just slightly higher during the Covid-19 pandemic. In other words, the historic CDS data from the pre Covid-19 period can be used for training methods that are supposed to be used in the post Covid-19 period, which can be a significant practical value for investors.

Our tests are based on single variable approaches for aggregate measures with unfavorable restrictions for sophisticated forecasting methods. It opens many suggestions for future studies. It can be checked whether the predictability properties are uniformly distributed across sectors or our results are forced by properties of a few sectors. It can be checked whether the cross-dependence of different CDS has changed due to the Covid-19 pandemic that can lead to lower market efficiency. The comparison of methods producing density (instead of point) forecasts can be investigated. The analysis of forecasting error distribution, including the changing of its volatility and statistical hypothesis testing can be another direction of future studies. Finally, future studies could optimize related parameters in model to achieve higher predictive performance (like LSTM in study of Peng et al. (2020)).

CRediT authorship contribution statement

Darko B. Vukovic: Conceptualization, Formal analysis, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. **Kirill Romanyuk:** Methodology, Software, Validation, Visualization. **Sergey Ivashchenko:** Formal analysis, Validation, Visualization. **Elena M. Grigorieva:** Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

a) The manuscript has supported by Russian Science Foundation for the project No. 22-28-01553 (Making Smart Decisions in the Face of Uncertainty in Russia: Investing and Forecasting in a Crisis).

b) This paper has been supported by the RUDN University Strategic Academic Leadership Program (for authors from RUDN).

c) Support from the Basic Research Program of the National Research University Higher School of Economics is gratefully acknowledged by Kirill Romanyuk.

Funding

The manuscript has supported by Russian Science Foundation for the project No. 22-28-01553 (Making Smart Decisions in the Face of Uncertainty in Russia: Investing and Forecasting in a Crisis).

Disclaimer

This article presents the authors' personal opinion and may differ from the official position of the Bank of Russia. The Bank of Russia assumes no responsibility for the content of the article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eswa.2022.116553.

References

Abe, M., & Nakayama, H. (2018). Deep Learning for Forecasting Stock Returns in the Cross-Section, In Pacific-Asia Conference on Knowledge Discovery and Data Mining, 273-284.

Ang, A., & Timmerman, A. (2012). Regime changes in financial markets. *Annual Review of Financial Economics, 4*, 313–337.

Avino, D., & Nneji, O. (2014). Are CDS spreads predictable? An analysis of linear and non-linear forecasting models. *International Review of Financial Analysis, 34*, 262–274.

Colladon, A. F., Grassi, S., Ravazzolo, F., & Violante, F. (2020). Forecasting financial markets with semantic network analysis in the COVID-19 crisis, arXiv preprint, arXiv: 2009.04975.

Chimmula, V. K. R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals, 135*, Article 109864.

Fama, E. F. (1955). The behavior of stock market prices. *Journal of Business, 38*, 34–105.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance, 38*, 34–105.

Fei, F., Fuertes, A., & Kalotychou, E. (2017). Dependence in credit default swap and equity markets: Dynamic copula with Markov-switching. *International Journal of Forecasting, 33*(3), 662–678.

Gaglione, D., Braca, P., et al. (2020). Adaptive Bayesian learning and forecasting of epidemic evolution—Data analysis of the COVID-19 outbreak. *IEEE Access, 8*, 175244–175264.

Gilchrist, S., Zakrajsek, E., Favara, G., & Lewis, K. (2016). Recession Risk and the Excess Bond Premium, *Fed Notes*, April (8), 1-3.

Ghosh, I., & Chaudhuri, T. D. (2021). FEB-stacking and FEB-DNN models for stock trend prediction: A performance analysis for pre and post Covid-19 periods. *Decision Making: Applications in Management and Engineering, 4*(1), 51–84.

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications, 124*, 226–251.

Hu, N., Li, J., & Meyer-Cirkel, A. (2019). Completing the Market: Generating Shadow CDS Spreads by Machine Learning, IMF Working Paper, WP/19/292, International Monetary Fund.

Ivakhnenko, A. G. (1988). Sorting methods for modeling and clustering (survey of the GMDH papers for the years 1983–1988) the present stage of GMDH development. *Soviet Journal of Automation and Information Sciences (Avtomatika), 21*, 1–13.

Jensen, M. (1978). Some anomalous evidence regarding market efficiency. *Journal of Financial Economics, 6*, 95–101.

Jia, L., Alizadeh, R., Hao, J., Wang, G., Allen, J. K., & Mistree, F. (2020). A rule-based method for automated surrogate model selection. *Advanced Engineering Informatics, 45*, Article 101123.

Katris, C. (2021). A time series-based statistical approach for outbreak spread forecasting: Application of COVID-19 in Greece. *Expert Systems with Applications, 166*, Article 114077.

Kim, W.-J., Jung, G., & Choi, S.-Y. (2020). Forecasting CDS term structure based on Nelson-Siegel model and machine learning. *Complexity, Special Issue, 23*, pages. https://doi.org/10.1155/2020/2518283

Kuan, C. M. (2002). Lecture on the Markov switching model. *Institute of Economics Academia Sinica, 8*(15), 1–30.

Lee, J., Kim, H., & Kim, H. (2021). Commercial vacancy prediction using LSTM neural networks. *Sustainability, 2021*(13), 5400.

Lee, H. C. B., Stallaert, J., & Fan, M. (2020). Anomalies in probability estimates for event forecasting on prediction markets. *Production and Operations Management, 29*(9), 2077–2095.

Liu, Q., Chen, C., Zhang, Y., & Hu, Z. (2011). Feature selection for support vector machines with RBF kernel. *Artificial Intelligence Review, 36*(2), 99–115.

Malkiel, B. (1992). *Efficient market hypothesis*. London: New Palgrave Dictionary of Money and Finance.

Peng, Y., & Nagata, M. H. (2020). An empirical overview of nonlinearity and overfitting in machine learning using COVID-19 data, *Chaos. Solitons & Fractals, 139*, Article 110055.

Peng, L., Zhu, Q., Lv, S. X., et al. (2020). Effective long short-term memory with fruit fly optimization algorithm for time series forecasting. *Soft Computing, 24*, 15059–15079.

Pereira, I. G., Guerin, J. M., et al. (2020). Forecasting Covid-19 dynamics in Brazil: A data driven approach. *International Journal of Environmental Research and Public Health, 17*(14), 5115.

Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing, 90*, Article 106181.

Stutzer, M. (2020). Persistence of averages in financial Markov Switching models: A large deviations approach. *Physica A, 553*, Article 124237.

Timmermann, A., & Granger, C. W. J. (2004). Efficient market hypothesis and forecasting. *International Journal of Forecasting, 20*, 15–27.

Tobback, E., Naudts, H., Daelemans, W., Junqué de Fortuny, E., & Martensa, D. (2018). Belgian economic policy uncertainty index: Improvement through text mining. *International Journal of Forecasting, 34*(2), 355–365.

Vukovic, D., Vyklyuk, Y., Matsiuk, N., & Maiti, M. (2020). Neural network forecasting in prediction Sharpe ratio: Evidence from EU debt market. *Physica A: Statistical Mechanics and its Applications, 542*(C), Article 123331.

Vyklyuk, Y., Vukovic, D., Jovanovic, A (2013). Forex predicton with neural network: usd/eur currency pair. ***Actual problem of economics***, Issue 10 (148) /2013, 251-261.

Xu, L., Bai, L., Xiao, J., Liu, Q., Chen, E., Wang, X., et al. (2021). Multiple graph kernel learning based on GMDH-type neural network. *Information Fusion, 66*, 100–110.

Yang, C. H., Liao, M. Y., et al. (2009). Constructing financial distress prediction model using group method of data handling technique. In Proceedings of the eighth international conference on machine learning and cybernetics, Baoding, 12–15, July.

Zhang, N., Lin, A., & Shang, P. (2017). Multidimensional k-nearest neighbor model based on EEMD for financial time series forecasting. *Physica A: Statistical Mechanics and its Applications, 477*, 161–173.

Zhou, T., Gao, S., Wang, J., Chu, C., Todo, Y., & Tang, Z. (2016). Financial time series prediction using a dendritic neuron model. *Knowledge-Based Systems, 105*(C), 214–224.

Dr. Darko B. Vukovic is the Head of the International Laboratory for Finance and Financial Markets at Faculty of Economics, People's Friendship University of Russia (RUDN University), in Moscow, Russia. He is also associate professor at Finance and credit department, Faculty of Economics, People's Friendship University of Russia (RUDN University), in Moscow, Russia. He is engaged on international master program, subjects: Financial markets and institutions and Financial Risk Management. Since 2008, dr Darko Vukovic works at Geographical Institute "Jovan Cvijic" of the Serbian Academy of Sciences and Arts, at position Head of Department of regional economics and economic geography.

Kirill Romanyuk, PhD is a senior lecturer at the Department of Finance at HSE University, Moscow, Russia. His research interests contain financial time series forecasting, credit scoring, credit risk analysis. Dr. Romanyuk is an expert in mathematical modeling and forecasting.

Ivashchenko S., PhD is economic adviser of North-Western Main Branch of the Bank of Russia. He is also senior researcher at Macroeconomic Analysis Division of Financial Research Institute, and senior researcher at theoretical economics laboratory of The Institute of Regional Economy Studies (Russian Academy of Sciences).

Elena M. Grigorieva, Ph.D., is Vice Dean for science at Faculty of Economics, RUDN University and Associate Professor at Finance and credit department, People's Friendship University of Russia (RUDN University), in Moscow, Russia. Her research interests are: Banking, Financial Management, Financial instruments for infrastructure development.