Paisii Hilendarski University of Plovdiv
Faculty of Mathematics and Informatics
Department of "Mathematical Analysis"

Ivaylo Vladimirov Boyoukliev

MODELING AND RESEARCH OF FOREIGN EXCHANGE FINANCIAL MARKETS

ABSTRACT

of a dissertation for awarding the educational and scientific degree "Doctor"
in the field of higher education 4. Natural sciences, mathematics and informatics
Professional direction 4.5 Mathematics
Doctoral Program: Mathematical Modeling and Application of Mathematics

Scientific supervisors:

Prof. D.Sc. Snezhana Georgieva Gocheva-Ilieva

Associate Prof. Dr. Hristina Nikolova Kulina

Plovdiv 2024

The dissertation work has been discussed and scheduled for the defense of an extended departmental council of the Department of "Mathematical Analysis" at the Faculty of Mathematics and Informatics of the "Paisii Hilendarski" University of Plovdiv, the city of Plovdiv, held on 15.01.2024.

The dissertation has a total volume of 160 pages, in which 46 tables and 68 figures, with an introduction, four chapters, a conclusion and a bibliography consisting of 135 sources. The list of author publications includes 4 titles.

The defense of the dissertation work will take place on 12.04.2024 at .... time in the Meeting Hall of the New Building of the "Paisii Hilendarski" University of Plovdiv at an open meeting of a scientific jury composed of:

1. Prof. Dr. Hristo Stefanov Kiskinov - chairman
2. Prof. Dr. Mikhail Dimov Todorov
3. Prof. Ph.D. Leda Dimitrova Minkova
4. Prof. Ph.D. Miglena Nikolaeva Koleva
5. Assoc. Ph.D. Ivanka Mitkova Zheleva

The numbering of formulas, citations, examples, tables and figures coincides with their numbering in the dissertation work.

The defense materials are available to those interested in the secretariat of FMI, new building of PU, cab. 330, every working day from 8:30 AM to 5:00 PM.

Author of the dissertation: Ivaylo Vladimirov Boyoukliev
Title: Modeling and Research of Foreign Exchange Financial Markets

Content

GENERAL CHARACTERISTICS OF THE DISSERTATION

Timeliness of the issues

This dissertation develops and applies both classical and state-of-the-art intelligent machine learning (ML) methods for statistical modeling and data analysis in the applied field of foreign exchange and financial markets.

The results of the research, modeling, analysis and forecasting of the financial and foreign exchange markets are widely used in the daily activities of a number of participants both in these markets and in all other economic areas. In order to increase the quality of the decisions made, their scientific validity, systematicity and credibility, in recent years, an increasingly intensive application of statistical modeling, primarily with intelligent methods, has been observed.

Along with this, in a scientific-applied aspect, there are a number of current tasks for the further development of modeling and the study of time series. They arise from the fact that in order to obtain adequate forecasts of the investigated dependent variable forward in time, it is necessary to have values of the predictors in the same future time period. In addition, predictors are often completely missing or very difficult to find. Another difficult but important issue for practice is the identification of the factors that have the strongest influence in the forecasting process..

A main goal and tasks of the dissertation work

The main goal of the present dissertation is the development and application of new approaches to modeling time series based on stochastic and powerful intelligent methods for the study, analysis and forecasting of currency financial markets.

The object of the research is actually existing historical data from the financial markets and systems, used as a basis for forecasting their future movements, values and dynamics, as well as for determining effective investment strategies in current financial systems and software.

To achieve the set goal of the dissertation work, **the following tasks** were formulated:

1. Development of a general framework for modeling univariate and multivariate time series with intelligent methods based on machine learning.
2. Study of time series from the sphere of currency financial markets and development of methods for selecting predictors for the purpose of short-term future forecasts.
3. Application of ensemble methods for modeling and forecasting one-dimensional time series for deposits in foreign currency of Bulgarian citizens.
4. Application of ensemble methods for modeling multidimensional time series for exchange rates and macroeconomic indicators.
5. Development of a hybrid approach with ARIMA correction for modeling fluctuations in citizens' bank deposits.

Scientific hypotheses

1. Application of the developed framework for analysis and modeling of one-dimensional time series with bank deposit data.
2. Obtaining a preliminary estimate and predicting the dynamics of the EUR/USD exchange rate forward in time by modeling, analyzing and forecasting multidimensional time series for financial data of a stochastic nature.
3. Establishing the dependence of exchange rates on various main macroeconomic indicators such as the main interest rate (PRI), gross domestic product (GDP), inflation, yield on 10 years. public debt and others.

Structure of the dissertation

The dissertation contains an introduction, 4 chapters, a conclusion, a declaration and a bibliography with a total volume of 160 printed pages.

In Chapter 1, the main elements and stages of statistical modeling are presented, the application area is defined, a detailed literature review is made on the subject of the dissertation, the methods used are explained, the auxiliary software is indicated, and the goals and tasks of the research are formulated.

In Chapter 2, the developed general framework of the research is described, a new approach is developed for modeling and forecasting one-dimensional time series from the financial sector with official bank data from the Bulgarian National Bank for the short-term deposits of Bulgarian citizens in USD currency. The approach consists in a detailed study of the autoregressive and partial autoregressive functions (ACF and PACF) of the initial time series, considered as the dependent variable, in order to identify possible predictor series. The presence of multicollinearity between the found predictors and the type of distribution is investigated. Using the candidate predictors, models are built with the CART Ensembles and Bagging (EBag) machine learning ensemble method. EBag models are built and evaluated in two cases – with pre-transformation of the data to improve the distribution towards normality and without transformation of the data. Examination of the order, ACF, and PACF indicated the possibility of using 6 lagged variables and a first-order trend as candidate predictors. To satisfy the ARIMA conditions, the initial time series was transformed with the Yeo-Johnson formula. A benchmark ARIMA(6,1,1) model was built and tested. For modeling with CART Ensembles and Bagging, multiple models were built and analyzed by calibrating the model training hyperparameters. The best model with transformed data was an ensemble of 60 CART trees W60a, which achieved coefficient of determination R2=94.3%, RMSE=$28.8 million, and MAPE=13.8%. Of the EBag models without data transformation, the models have better statistics. The selected best EBag model EYY30, with R2=97.7%, RMSE=$18.2 million and MAPE=10.2%.

In Chapter 3, a new approach is developed for modeling and short-term forecasting of the dynamics of the Euro/US Dollar (EUR/USD) exchange rate as a function of 8 main macroeconomic indicators. Modeling is conducted according to the general research framework. In the first stage, the components of the multivariate time series were analyzed with the univariate ARIMA method, statistical tests and other analyses. In the second stage, the results of three different ensemble methods with ML were applied and compared: CART-Ensembles and bagging (EBag), Arcing and Random Forests (RF). The obtained results are such that predictors are determined from the values of the dependent variable Y=EUR/USD_FX_RATE and the lagged variables of the eight macroeconomic factors X1, X2, ..., X8. The three ensemble methods EBag, Arcing and RF are applied to model and forecast the exchange rate one month ahead. Multiple models were built and analyzed with each of the ensemble methods. Of these, four selected models are presented, using four groups of predictors. For EBag, the models with the best statistical performance were obtained with model G4, which matched the exchange rate data with R2=98.8% and had accuracy statistics RMSE=0.0174, MAPE=1.03%. Of the Arcing models with the best performance is model AG4 with R2=99.1%, RMSE=0.0152, MAPE=1.00%. Of the RF models, the best is model RG4 with: R2=98.6%, RMSE=0.0194, MAPE=1.20%. The adequacy of the models was established by diagnosing their residues with the Ljung-Box test. The MAPE statistics of the described models range from 1% to 1.2%, i.e. are below 10%, with which we can classify the obtained models as very accurate.

In Chapter 4, hybrid methods are developed and explored. A one-dimensional dynamic series with monthly data on deposits of Bulgarian individuals in US dollars for a period of 227 months was studied. The modeled data are divided into two parts, of which the larger observed parts are used to build the models, and the last three values are left to test and verify the accuracy of the models to predict "new data". First, hybrid Arcing-ARIMA models were built with transformed data separately for two data samples – a sample with all available data (A) and a reduced sample (B) including 70% of the initial data. After transforming the data, parametric ARIMA models were built for comparison, which showed fit to the data with up to R2=58% and R2=49%, respectively, for the two samples A and B. Subsequently, four lagged variables were identified as predictors (respectively for lags 1, 2, 3, 6), first order trend and one time variable (Month). Several Arcing models were built, of which A50 was the chosen model for sample A and B60 for sample B. To eliminate serial correlation, the errors of these models were further modeled with a one-dimensional ARIMA model and hybrid Arcing-ARIMA models were constructed. Their indicators after retransformation for the full sample A were estimated at R2=98.5% and MAPE=12.5%, and for the reduced sample B – at R2=98.1% and MAPE=16.22%, respectively. Arcing-ARIMA hybrid models were developed with untransformed data. The selected best HA_AL120 hybrid model for the 224 data sample has statistics R2=99.6% and MAPE=12.9% and shows the most accurate predictions for the three test "future" months. Finally, Random Forests models were built with the untransformed data for comparison. By varying the hyperparameters, three models were identified, of which RF Model 2 was selected as the best model. Its residuals do not contain autocorrelation and have a near-normal distribution that does not require hybridization. RF Model 2 statistics are: R2=98.2% and MAPE=17.7%. Of the three modeling types in this chapter, the one with the best statistics and most accurate predictions for the test data is the hybrid Arcing-ARIMA model HA_AL120.

The conclusion presents a brief summary, a list of publications on the dissertation work, approvals, and claims for scientific contributions of the doctoral student are systematized.

The bibliography includes 135 sources.

The dissertation work was developed at the "Mathematical Analysis" department of the Faculty of Mathematics and Informatics of "Paisii Hilendarski" University of Plovdiv in the period 2020-2023.

BRIEF CONTENTS OF THE DISSERTATION

CHAPTER 1. INTRODUCTION

1.1. Basic concepts of time series data

The study of foreign exchange financial markets boils down to the processing, analysis and modeling of time series. We will consider the modeling task with regression type methods. The purpose of the regression type of model is both to describe and predict the observed data, and to make it possible to obtain predictions for the next points in time. The general form of a time series $Y(t)$ is a sequence of numbers of the type

$$Y(t) = \{y_1, y_2, ...y_t, ,..., y_N\}, \ y_t \in \mathbb{R} \tag{1.1}$$

where t is time index, N is number of observations (sample size), $\mathbb{R}$ is the set of real numbers. To build a regression model, in the one-dimensional case, we look for a relationship between each observation y_t and several of its previous ones $y_{t-1}, y_{t-2}, ...$. This process is called autoregressive and is only characteristic of dynamic data. In the presence of other lines related to (1.1) in the same scale and time period, the task becomes multidimensional. In a real situation, for a regression type of model, it will be possible to account for both autoregressive components of $Y(t)$, and the influences of the other rows with their autoregressive terms. We will consider that r independent rows $X_{1,t}, X_{2,t}, .., X_{r,t}$ are set in a vector for each t $\mathbf{X}_t = (X_{1,t}, X_{2,t}, .., X_{r,t})$. We are looking for an explicit form of the dependence (model) of Y on $\mathbf{X}_t$ on the form:

$$\hat{Y}(t) = f(\mathbf{X}_t) \tag{1.2}$$

The error (residue) of the model $\hat{Y}(t)$ for each *t* denote with $\varepsilon(t) = Y(t) - f(\mathbf{X}_t)$. This error is usually assumed to be uncorrelated with zero mean and constant variance σ^2 (white noise), i.e. $\varepsilon(t) \in \mathrm{N}(0, \sigma^2)$.

Another basic concept of time series is the concept of stationarity. We say that a series is stationary if its mean and standard deviation do not change over time. Otherwise, it is non-stationary and has a trend. A first-order trend is calculated by numerical differentiation according to the formula $Trend(t) = Y(t) - Y(t-1)$ for each *t*.

1.2. Methods used for modeling and research of foreign exchange financial markets

As a general rule, in all studies in applied topics, it is noticed that the database of the central banks or statistical institutes in the studied countries is most often used as a source [22, 42], and for their statistical processing, the help of certain specialized and automated software products and statistical software [57, 103].

Autoregressive integrated moving average method (Box Jenkins ARIMA)

The Box-Jenkins forecasting methodology uses an iterative approach to determine a valid linear type parametric model. The general form of the model sets with a general formula the value of the row at each time point *t*. The model is written as ARIMA(p,d,q)s, where the parameters are: p - number of previous values for current t (lagged terms), d - number of trends, q - number of stochastic moving averages, *s* - seasonality. The selected model is compared with historical data to check how accurately it describes the temporal order. A model is considered acceptable if its residuals are small, randomly distributed, and contain no useful information. If a model is not satisfactory, the process is repeated with updated parameters. This procedure is repeated until a satisfactory model is found. The resulting model can be used for predictive purposes.

Ensemble method of classification and regression trees with bagging (CART Ensembles and Bagging)

EBag is an intelligent machine learning (ML) method with CART trees [19]. It is obtained by averaging the predictions of multiple models generated with the same algorithm for different redistributed (bootstrap) samples from the initial data. It is trained by default with cross-validation. EBag provides additional functions for mean and median relative weighting of predictors in a regression predictive model, differential weights in multiclass classification, missing value imputation, the results of all individual tree statistics in the ensemble, comparisons of all tree statistics, and more. (Fig. 1.3). EBag is poorly studied for financial data.

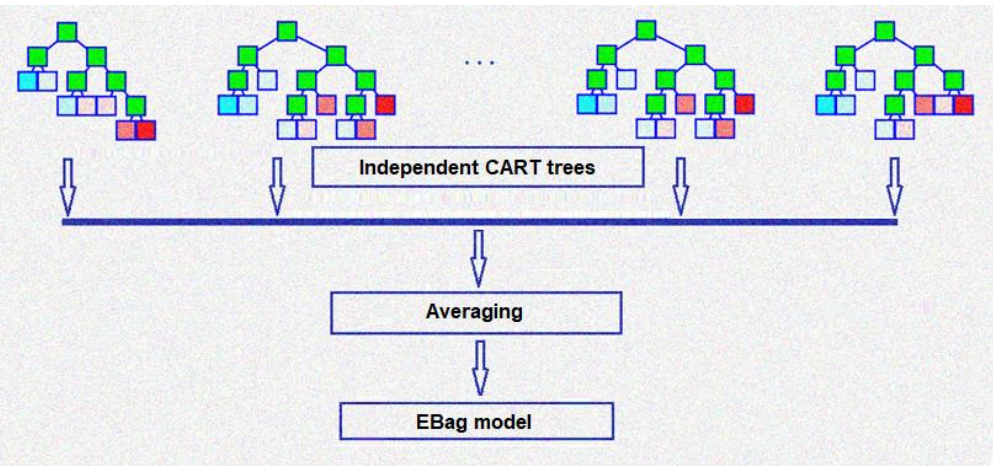

Figure 1.3. Scheme of CART ensemble model.

Ensemble method with adaptive boosting (Adaptively resampling and combine) – Arcing

Arcing is an algorithm based on CART resolution trees of the gradient boosting class [18]. It sequentially combines the models, with each subsequent one taking into account and improving the qualities of the previous models. The idea of the method is illustrated in Fig. 1.4.

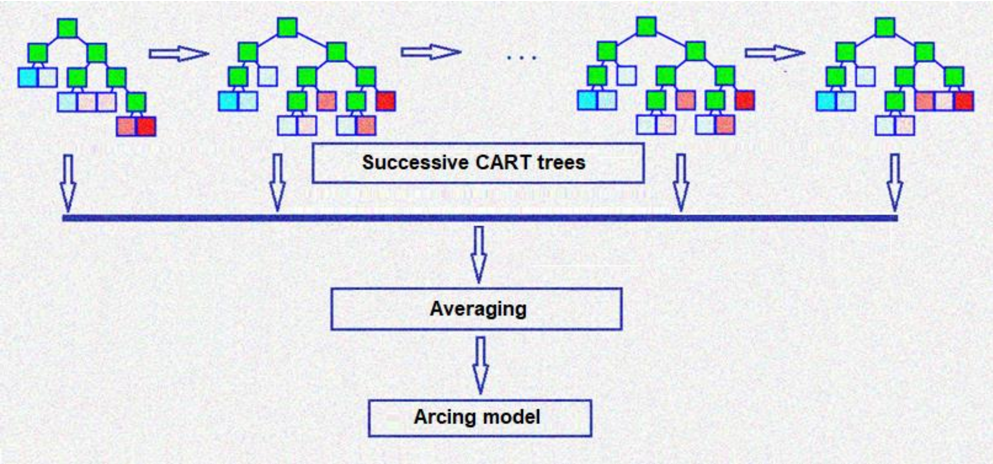

Figure 1.4. Scheme of Arcing model.

Arcing enables more efficient training of single models, better prediction of test samples and hence obtaining an alternative ensemble model. Provides information as EBag. So far, this method has not been well researched for empirical data and has not been used in the field of foreign exchange financial markets.

Random Forests method (Random Forests, RF)

The use of RF is as a major 'benchmark' for machine learning applications [20]. It is one of the most preferred methods for prediction and multi-class classification. The way of learning and forecasting new data with OOB provides a great opportunity to obtain highly efficient models and forecasts in the complex realm of currency financial markets. The results of the RF models will be compared with the other approaches.

Hybridization of models

Sometimes it is necessary to further refine the results for the resulting model. In the case of time series, the statistical adequacy and applicability of the models require checks included in model error diagnostics or other requirements. For this purpose, error modeling and residual dependency extraction can be applied. In the current dissertation, the errors (residuals) are further modeled with ARIMA models. The hybrid model is obtained after summing the output model with the resulting ARIMA model.

The theoretical foundations of ensemble methods are developed in [45, 66, 110, 119].

CHAPTER 2. MODELING OF FOREIGN CURRENCY DEPOSITS WITH ENSEMBLE METHODS

This chapter presents a general framework of research in the dissertation. The EBag method is applied to predict monthly foreign currency deposit data.

2.1. General framework of research in the dissertation work

For the construction and analysis of predictive regression models, we propose and follow in the dissertation the following general framework, presented in Fig. 2.1.

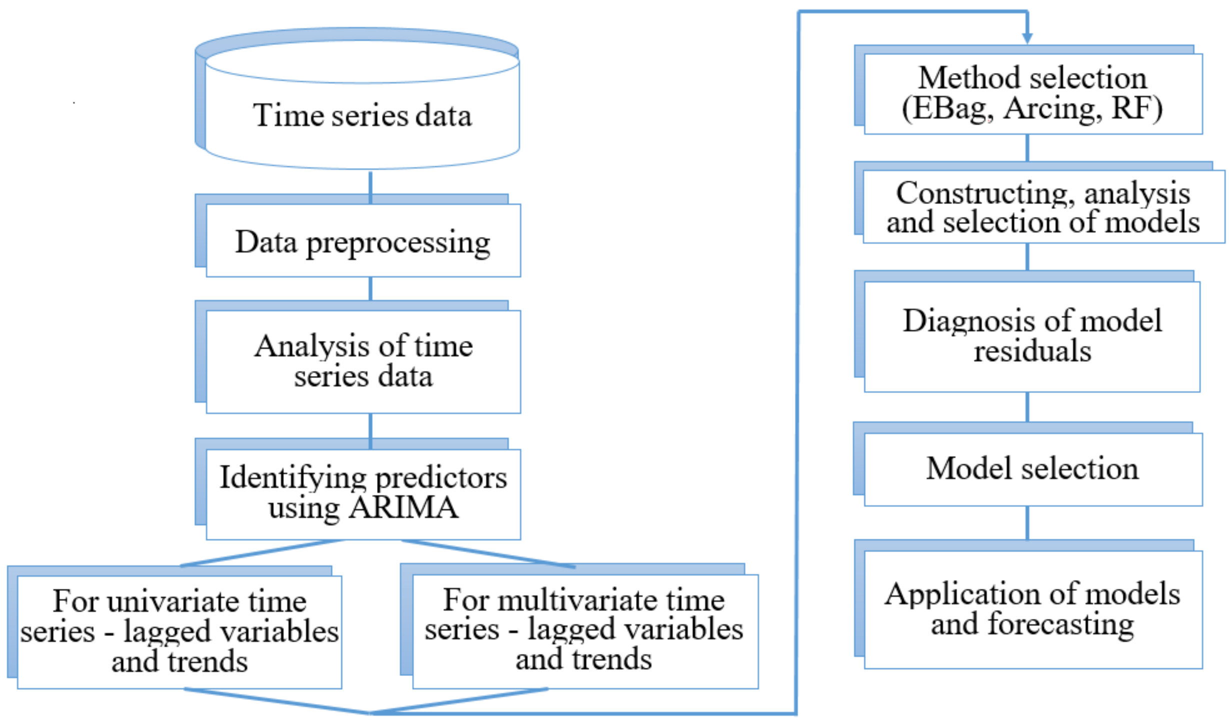

Figure 2.1. General framework scheme.

2.2. Data description, initial processing, and building a benchmark ARIMA model to identify predictors

The research is based on official data on the deposits of Bulgarian citizens in the period February 2004 - April 2021, a total of N=207 records. Data are in millions of US dollars (USD), averaged over months. Their source is the Bulgarian National Bank [22].

The variable of the corresponding one-dimensional dynamic series is denoted by Depos. Descriptive statistics of initial deposit values are given in Table 2.2.

Table 2.2. Descriptive statistics of the dependent variable Depos

Variable	Average value	Median	Std. Deviation	Dispersion	Asymmetry	Std. asymmetry error	Excess	Std. kurtosis error
Depos, USD	158.28	139.44	111.55	12444.39	1.009	0.169	0.983	0.337

These indicators suggest that the distribution of Depos is not close to normal. Kolmogorov-Smirnov and Shapiro-Wilk tests are significant with values Sig.=0.001 and Sig.=0.000. We conclude that the initial variable Depos is not normally distributed. To stabilize the variance and improve the distribution of Depos, we apply the following Yeo-Johnson transformation:

$$Tr\_Y = YJ(Y,\lambda) = \begin{cases} \dfrac{(Y+1)^{\lambda}-1}{\lambda} & Y \geq 0,\ \lambda \neq 0 \\ \log(Y+1) & Y \geq 0,\ \lambda = 0 \\ -\dfrac{(-Y+1)^{2-\lambda}-1}{2-\lambda} & Y < 0,\ \lambda \neq 2 \\ -\log(-Y+1) & Y < 0,\ \lambda = 2 \end{cases},\ \lambda \in [-2,2] \tag{2.1}$$

where Tr_Y is the transformed variable, and the parameter λ is found by the researcher.

After the transformation, we obtain the Tr_Depos variable with a normal distribution and use the ARIMA method to build a reference model. The model is:

$$\text{ARIMA}(6,1,1)_{12}. \tag{2.2}$$

The statistical indicators of model (2.2) are: $R^2 = 51.7\%$ and RMSE=3.837 million dollars.

2.3. Application of ensemble CART method to data on currency deposits of Bulgarian citizens

The constructed ARIMA model (2.2) suggests choosing predictors from the lagged variables with up to p=6 lags (lags back from each current). For our case, we will also consider the differentiated time series of the first order, describing the presence of a trend. We apply the EBag method with different numbers of trees in the ensemble and with different numbers of lagged variables. To avoid possible overfitting of the EBag models, they are trained by 10-fold cross-validation. Of all the obtained models, we choose the W40 and W60a models (with 40 and 60 trees in the ensemble, respectively) as the best. After retransforming the predicted values and returning to the initial variable Depos, the predicted values from the W40 and W60a models are calculated. The resulting variables are denoted by Predicted_W40 and Predicted_W60a, respectively. Their statistics calculated for all data (at N=207) show a slight decline. They are given in Table 2.9.

Table 2.9. Statistics of the models W40 и W60a after retransformation.

Model	R^2	RMSE	MAPE
Predicted_W40	0.941	29.951	0.152
Predicted_W60a	0.943	28.820	0.138

To analyze the adequacy of the selected W40 and W60a models, their residuals were examined. The residuals are within the theoretical error $\pm 2/\sqrt{N}$ as can be seen from their ACF diagram in Figure 2.13.

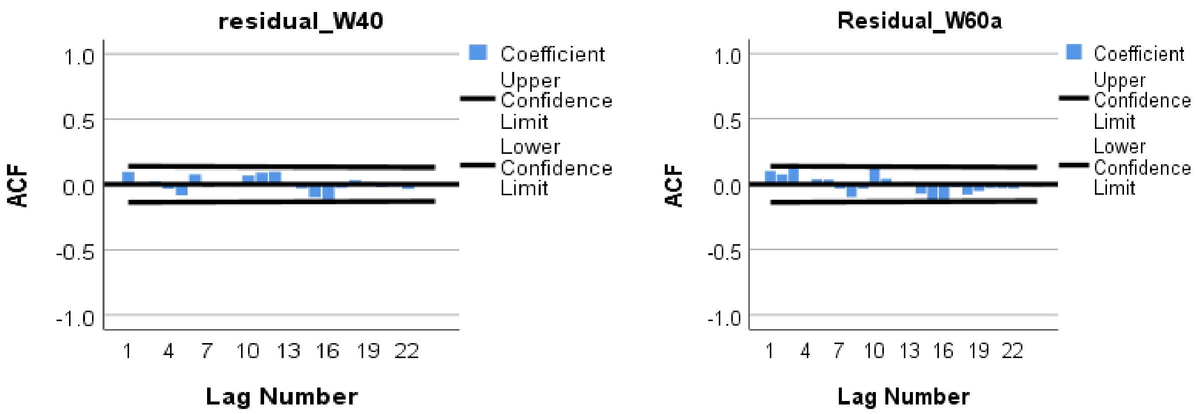

Figure 2.13. Autocorrelation function (ACF) of the residuals of models W40 and W60a.

The constructed EBag models W40 and W60a can be considered statistically valid. The numerical results for the last three months predicted by these models are given in Table 2.12. A better prediction of W60a is reported, especially for the first two months.

Table 2.12. Comparison of predicted values from Tr_Depos204's EBag models for the three months of testing.

Variable	February, 2021	March, 2021	April, 2021
Depos	30.562	48.952	16.083
ARIMA(6,1,1)	18.642	6.878	8.779
*W*40	18.184	19.587	19.392
*W*60a	27.711	36.409	35.323

2.4. Application of the EBag method for Depos204 without transforming the data

We repeat the study using the same data described in paragraph 2.2. Here again we have separated the last three data (months February, March and April 2021) for testing. As already noted in 2.3, since the data are not normally distributed, the use of classical parametric methods is not recommended. In our case, we will use EBag, which has no restrictions on the allocation of variables. We conduct the research without transformation of the source data.

Following paragraph 2.3 possible predictors of Depos204 are its lagged variables Depos <1, 2, 3, 6> and Trend204. A large number of models were built and evaluated. To calibrate the models, the following hyperparameters were varied: number of CART trees T in the ensemble of 40, 50, 60 and 70; type of training – cross-validation CV 5 and CV 10; ratio between minimum observations in parent to descendant node m1:m2 – 5:5, 10:5. Table 2.14 shows the statistics of selected EBag models. From the first group of models with m1:m2=10:5, the best performance is model EY30. But it is inferior to model EYY30 from the remaining group with m1:m2=5:5, which we will choose as the best.

Table 2.14. Statistics of selected Depos204 EBag models without data transformation.

EBag model	Number of trees	m1:m2	CV	Predictors	R^2	RMSE	MAPE
***EY*30**	**30**	**10:5**	**CV10**	***Depos*204<1>, <6>, *Depos*204_d1**	**0.9516**	**25.623**	**0.163**
*EY*50	50	10:5	CV10	*Depos*204<1>, <6>, *Depos*204_d1	0.9512	25.948	0.166
*EY*60	60	10:5	CV10	*Depos*204<1>, <6>, *Depos*204_d1	0.9513	25.971	0.164

_EY_60a	**60**	**10:5**	**CV10**	**_Depos_204<1>, <3>, <6>, _Depos_204_d1**	**0.9560**	**25.074**	**0.155**
_EY_60b	60	10:5	CV10	_Depos_204<1>, <2>, <3>, <6>, _Depos_204_d1	0.9559	25.178	0.164
_EYY_30	**30**	**5:5**	**CV10**	**_Depos_204<1>, <6>, _Depos_204_d1**	**0.9765**	**18.218**	**0.102**
_EYY_30a	30	5:5	CV5	_Depos_204<1>, <6>, _Depos_204_d1	0.9765	18.195	0.102
_EYY_40	40	5:5	CV10	_Depos_204<1>, <6>, _Depos_204_d1	0.9767	17.966	0.103
_EYY_50	50	5:5	CV10	_Depos_204<1>, <6>, _Depos_204_d1	0.9777	18.032	0.104
_EYY_60	60	5:5	CV10	_Depos_204<1>, <6>, _Depos_204_d1	0.9777	18.162	0.103

Figure 2.17 shows a plot comparing the original data with the predicted values of EYY30. A very good approximation is observed. After recalculating the statistics for N=207 data, we get: R2=0.9766 or 97.7% data fit and error RMSE=$18.174 million, MAPE=0.109 or 10.9%.

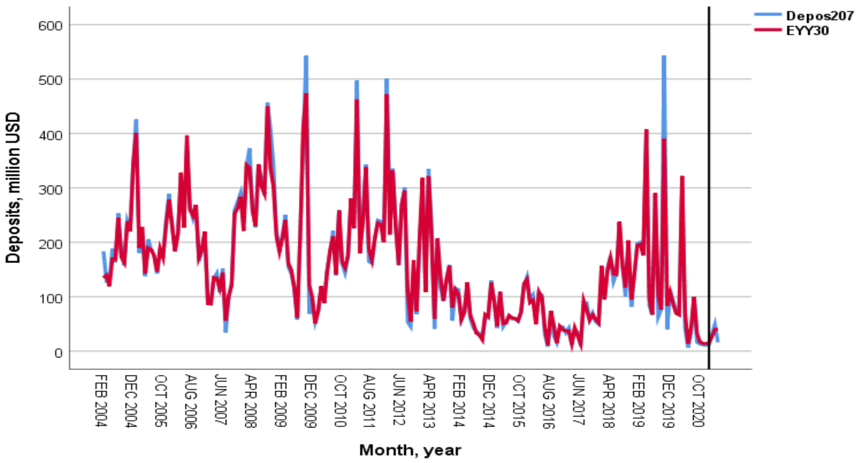

Figure 2.17. Observed values of the deposit time series Depos (blue line) and those predicted by the EYY30 model (red line).

In order to demonstrate the absence of autocorrelation in the residuals of the selected EYY30 model, Figure 2.20 presents the residual autocorrelation function (ACF). It can be seen that all the ACF coefficients of the residuals up to lag 24 are within the confidence intervals.

We can conclude that an EBag model without data transformations is adequate and can be applied to predict the order Depos207.

Estimated values for the last three months not involved in modeling for models EY30 and EYY30 are given in the following Table 2.15. One can note the good forecasting of most models for the first and second months, while for the third month there are greater inaccuracies. This is due to the use of the small number of predictors.

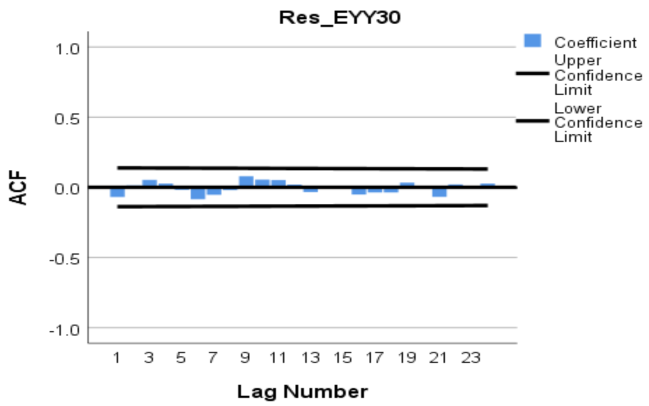

Figure 2.20. ACF of the residuals on model EYY30

Table 2.15. Estimated values from Depos204's EBag models without data transformation for the three test months

Variable	February, 2021	March, 2021	April, 2021
Depos	30.562	48.952	16.083
EY30	31.284	39.643	39.643
EYY30	26.811	40.000	40.000

CHAPTER 3. MODELING OF THE EUR/USD EXCHANGE RATE

In this chapter, in order to achieve the most accurate forecast possible of the development of the EUR/USD exchange rate, we will focus on the dependence of foreign exchange trading, taking into account the influence of key macroeconomic indicators.

Modeling is carried out in two main stages, namely predictor selection and model building according to the general framework. In the first stage, we use time series analysis and the Box-Jenkins univariate ARIMA approach [15, 87, 48], as well as statistical tests. In the second stage, after we have determined the relevant predictor time series from the macroeconomic factors for modeling and forecasting the EUR/USD exchange rate, we apply and compare the results of three different ensemble methods with MO: CART-Ensembles and bagging (EBag), Arcing and Random Forests (RF).

3.1. Data used for analysis and forecasting of the EUR/USD exchange rate and initial processing

This chapter uses official data provided by Bloomberg Professional Terminal. The data are arranged by month, with the dynamic row covering the period December 1998 - December 2021, or a total of N=277 cases. The variables and their descriptive statistics are given in Table 3.1.

Table 3.1. Descriptive statistics of the data used to model the EUR/USD exchange rate.

Variable	Description	Unit	Average vallue	Median	Std deviation
Y	EUR/USD_FX_RATE	unit	1.1997	1.1996	0.1603
X_1	EUROZONE_UNEMPLOYMENT	%	9.3401	9.1000	1.3455
X_2	USA_UNEMPLOYMENT	%	5.8769	5.4000	1.9333
X_3	EUROZONE_INFLATION	index	92.619	93.310	9.6647
X_4	USA_INFLATION	%	2.2415	2.1000	1.3511
X_5	ECB_INTEREST_RATE	%	0.8529	0.2500	1.2893
X_6	FED_INTEREST_RATE	%	1.8673	1.2500	1.9222
X_7	EUR_10Y_BUND_YIELD	%	2.4485	3.0250	1.9060

X_8	USD_10Y_BOND_YIELD	%	3.2976	3.1435	1.4196

The magnitude of the EUR/USD exchange rate, denoted by Y=EUR/USD_FX_RATE, is considered the dependent variable. The specificity of the dynamic series data in Table 3.1 is that they are measured simultaneously, which makes it difficult to make real-time predictions about their future values. Figure 3.2 shows the boxplots of the standardized variables of Y. From these it can be concluded that the distributions of all the variables are close to normal. This is most clearly expressed for the dependent variable Y= EUR/USD_FX_RATE.

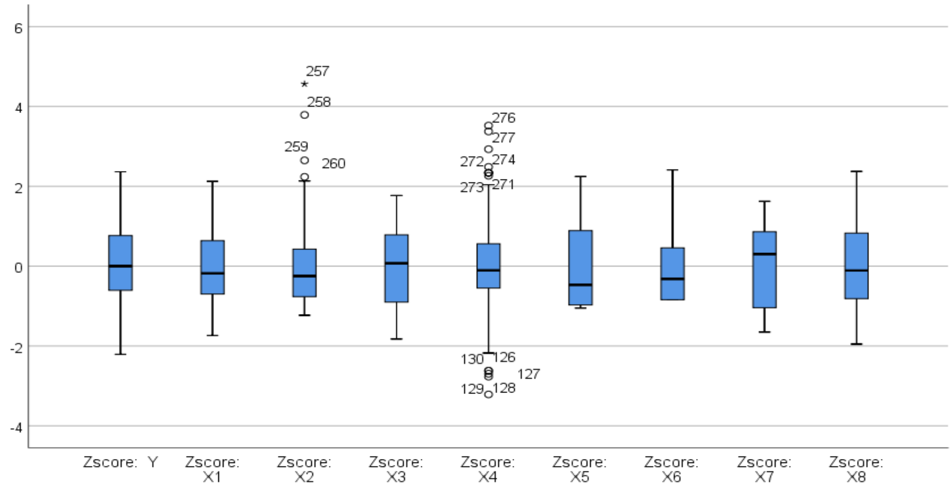

Figure 3.2. Box plot of the time series distributions used to model the exchange rate.

After checking the necessary normality assumption of the original independent time series, their autocorrelation functions were examined and univariate ARIMA analysis modeling was performed for each of them separately. Some of the constructed models are shown in Table 3.3.

Table 3.3. Statistics of the resulting univariate ARIMA models for selecting predictors for modeling the EUR/USD exchange rate

Variable	Description	ARIMA $(p,d,q)_{12}$	Significance of Ljung-Box test	R^2
Y	EUR/USD_FX_RATE	$(0,1,0)_{12}$	0.103	0.956
X_1	EUROZONE_UNEMPLOYMENT	$(0,1,9)_{12}$	0.198	0.996
X_2	USA_UNEMPLOYMENT	$(0,1,1)_{12}$	0.987	0.867
X_3	EUROZONE_INFLATION	$(0,1,6)_{12}$	0.071	1.000
X_4	USA_INFLATION	$(0,1,1)_{12}$	0.202	0.946
X_5	ECB_INTEREST_RATE	$(0,1,3)_{12}$	0.056	0.989
X_6	FED_INTEREST_RATE	$(0,1,9)_{12}$	0.684	0.991
X_7	EUR_10Y_BUND_YIELD	$(0,1,3)_{12}$	0.740	0.991
X_8	USD_10Y_BOND_YIELD	$(0,1,0)_{12}$	0.383	0.969

The models have excellent predictive qualities with a large coefficient of determination showing a match with the modeled data from 87% for USA_UNEMPLOYMENT to 100% for EUROZONE_INFLATION. The most important information for us from Table 3.3 is that a first-order trend is identified in all models, i.e. d=1 for all variables and a zero autoregressive (AR) term (ie, p=0).

3.2. Application of CART Ensembles and Bagging to Modeling the Exchange Rate

Aiming to build EBag models, we use the lagged variables of X1,X2…,X8, divided into four predictor groups according to the type of their information, namely:

- Data with the unemployment indices in the EU and the USA;
- Variables with EU and US inflation data;
- The main interest rates of the ECB and the FED;
- The yield of 10-year US and European bonds.

In the modeling process, a large number of models were built by varying their hyperparameters, and the best-performing models turned out to be the EBag models with 40 trees, standard 10-fold cross-validation, and a minimum of 5:5 endpoint observations. Table 3.6 shows the results achieved, where the models are labeled G1, G2, G3 and G4.

Table 3.6. Statistics of the selected EBag models for the exchange rate

Model	Predictors	RMSE	MAPE	R^2
*G*1	LX_1, LX_2	0.0408	0.0222	0.9377
*G*2	LX_1, LX_2, LX_3, LX_4	0.0182	0.0108	0.9874
*G*3	$LX_1, LX_2, LX_3, LX_4, LX_5, LX_6$	0.0185	0.0108	0.9868
*G*4	$LX_1, LX_2, LX_3, LX_4, LX_5, LX_6, LX_7, LX_8$	0.0174	0.0103	0.9884

Figure 3.3 illustrates a row plot of the original EUR/USD_FX_RATE data compared to that predicted by the best G4 model.

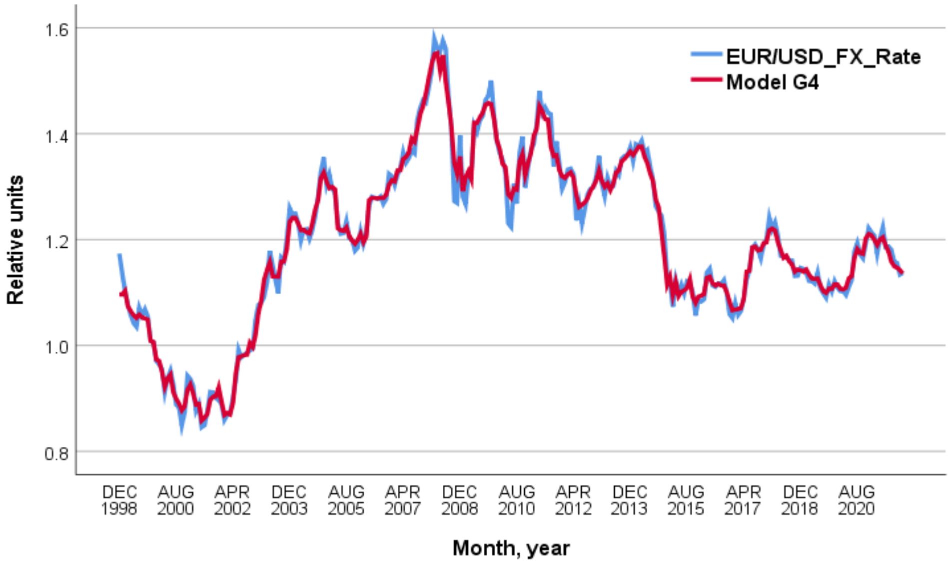

Figure 3.3. Comparison of the original data for Y = EUR/USD_FX_RATE with the predicted values from the EBag model G4.

3.3. An Arcing application for modeling exchange rate data

We repeat the study as predictors using the same variables as for the EBag models, divided into the four groups according to the type of their information (see paragraph 3.2). The dependent variable is Y=EUR/USD_FX_RATE, from which the last value (for December 2021) has been removed and used for verification.

In the modeling process, a large number of models were built by selecting and calibrating the main hyperparameters. Table 3.7 shows the results achieved, where the models are labeled AG1, AG2, AG3 and AG4.

Table 3.7. Statistics of the received Arcing models of the exchange rate.

Model	Predictors	RMSE	MAPE	R^2
$AG1$	LX_1, LX_2	0.0349	0.0227	0.9580
$AG2$	LX_1, LX_2, LX_3, LX_4	0.0164	0.0107	0.9897
$AG3$	LX_1, LX_2, LX_3, LX_4, LX_5, LX_6	0.0160	0.0104	0.9901
$AG4$	LX_1, LX_2, LX_3, LX_4, LX_5, LX_6, LX_7, LX_8	0.0152	0.0100	0.9912

Figure 3.9 shows a plot of all the original EUR/USD_FX_RATE data compared to the predicted values from the best AG4 model. A very good match is observed, except for some extreme values, which is characteristic of ensemble methods whose models are calculated by averaging.

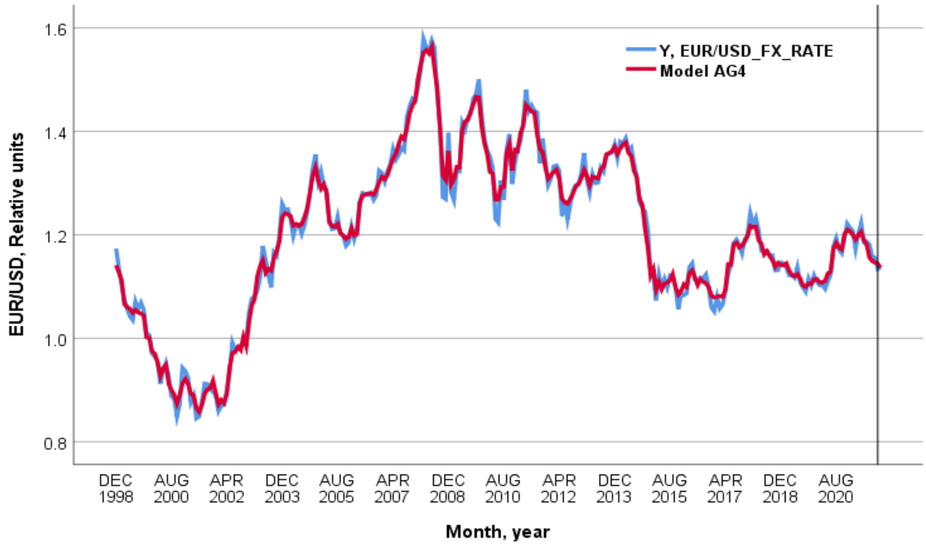

Figure 3.9. Comparison of the original data for Y = EUR/USD_FX_RATE with the predicted values from Arcing model AG4.

The quality of approximation of the original EUR/USD_FX_RATE data with the selected best Arcing model AГ4 is illustrated in Figure 3.11 with 95% confidence intervals.

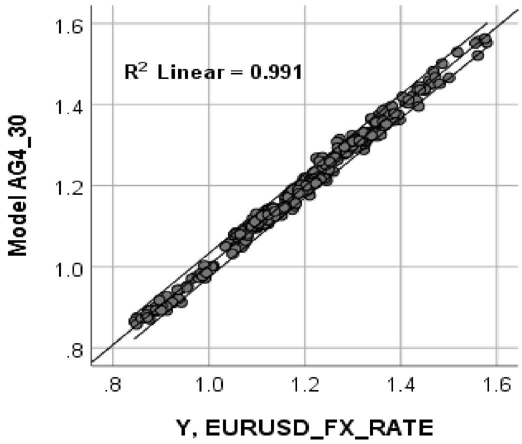

Figure 3.11. Scatterdot of EUR/USD_FX_RATE against predicted values from Arcing model AG4.

An important advantage of the studied ensemble models is that information is extracted about the weights of the used predictors. For model AG4, Figure 3.12 shows the mean and median values of the relative contribution of the predictors. The biggest relative influence is EU inflation (100), followed by EUR_10Y Bund yield (88), USD_10Y Bund yield (63) and ECB interest rate (57).

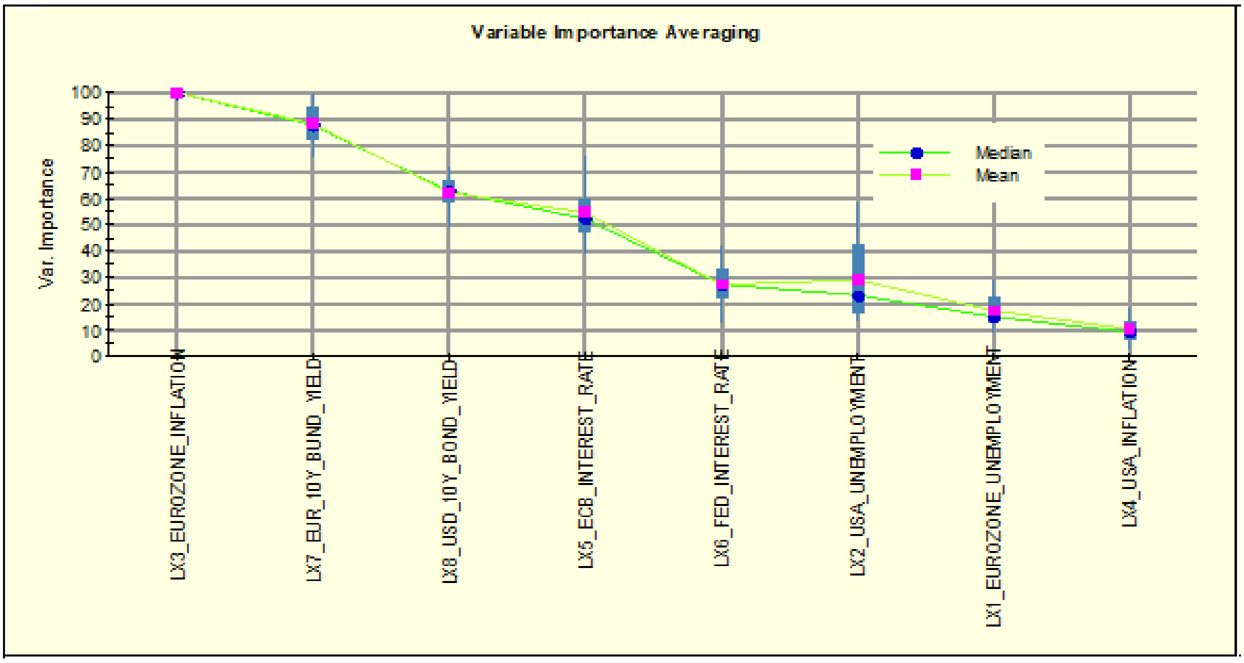

Figure 3.12. Relative weight of predictors in model AG4.

3.4. Modeling the exchange rate with Random Forests

We repeat the study using the same predictor variables as for the EBag and Arcing models, following the general research framework. For RF modeling, the inclusion of the first lagged dependent variable Y=EUR/USD_FX_RATE to the eight predictors proved appropriate. I.e. here the predictor variables are nine. Multiple models were built varying the RF hyperparameters, with the best statistical results with RF models obtained with T=350 trees, minimum observations at terminal node Atm=5, and mtry=3. The selected models are labeled RG1, RG2, RG3 and RG4, and their corresponding statistics are given in Table 3.9. From these tables it can be seen that the last three models have very close measurements, with the RG4 model being the best.

Table 3.9. Statistics of selected RF exchange rate models

Model	Predictors		RMSE	MAPE	R^2
$RG1$	LY, LX_1, LX_2		0.0208	0.0131	0.9837
$RG2$	$LY, LX_1, LX_2, LX_3, LX_4$		0.0196	0.0120	0.9858
$RG3$	$LY, LX_1, LX_2, LX_3, LX_4, LX_5, LX_6$		0.0194	0.0121	0.9861
$RG4$	$LY, LX_1, LX_2, LX_3, LX_4, LX_5, LX_6, LX_7, LX_8$		0.0194	0.0120	0.9862

The relative influence of the individual predictors in the RG4 model is visualized in Figure 3.17.

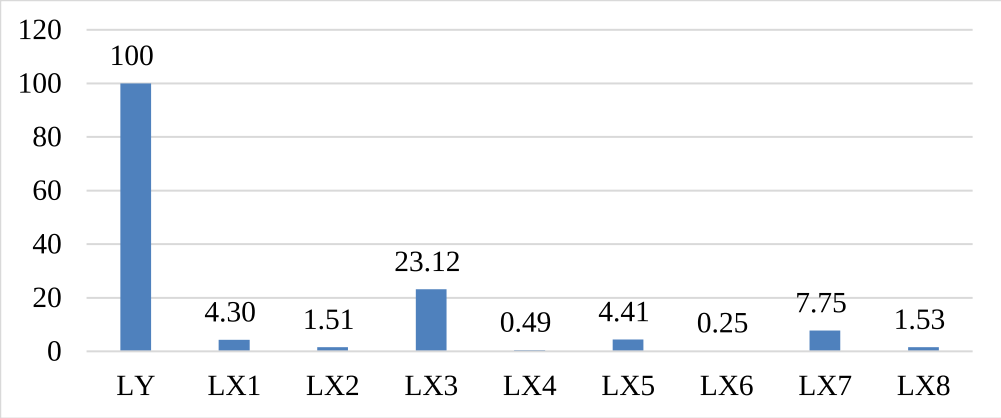

Figure 3.17. Influence of the predictors of RF model RG4.

Autocorrelation examination of the residuals of model RG4 with Figure 3.18 shows that they contain no autocorrelation.

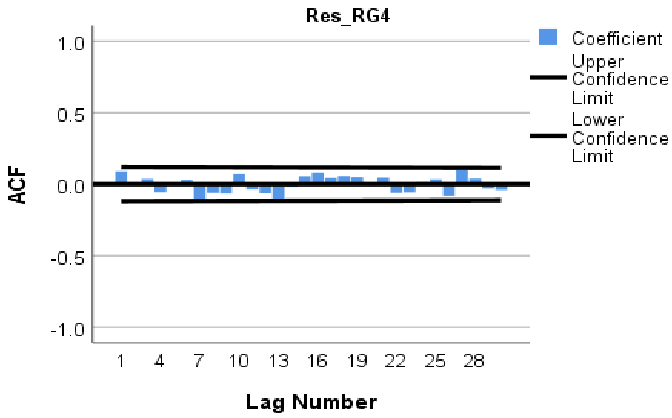

Figure 3.18. Autocorrelation function of RF model RG4 residuals

For the last test value for the last month for the EUR/USD exchange rate equal to 1.1370 the RF model predictions are as follows:
1.12445 (RG1), 1.12640 (RG2), 1.13115 (RG3), 1.13674 (RG4). The predicted value of the RG4 model is very accurate, with an absolute error of less than 0.0003.

CHAPTER 4. MODELING OF BANK DEPOSITS WITH HYBRID METHODS

In this chapter, results are obtained for univariate time series modeling with a hybrid approach based on intelligent methods with ML and univariate seasonal ARIMA. The Arcing method with subsequent hybridization has not been used so far by other authors in the field of financial markets. Since no formal such application has yet been found in the scientific literature on this problem, we accepted the challenge to demonstrate the capabilities of this approach.

4.1. Problem description and data

B The study models a univariate time series of bank deposit data, using an extended sample, against the data of Chapter 2. The developed general framework used in Chapters 2 and 3 is applied to identify lagged variables and trends to construct predictors. In the study, we use the officially published data of the BNB [22] on bank deposits in millions of US dollars of individuals on an average monthly basis. The data covers the period from February 2004 to October 2022, (N=227 records) and is an extended version of the data from Chapter 2, with a new 20 months.

4.2. Building and analyzing reference univariate ARIMA models of Tr_DeposA and Tr_DeposB, and identifying predictors

To predetermine the possible values of the parameters, we examine the ACF and PACF of the respective rows. The corresponding graphs are given in Figure 4.6. The smooth decrease of ACF coefficients with time is characteristic of time series without cycles and seasonality in the data. PACFs show several values outside the confidence intervals. This is an indicator for determining the parameters of the corresponding ARIMA model. In our case, p, q are within 1 to 6 for both rows Tr_DeposA, Tr_DeposB. with a more pronounced lag 6 for the first row.

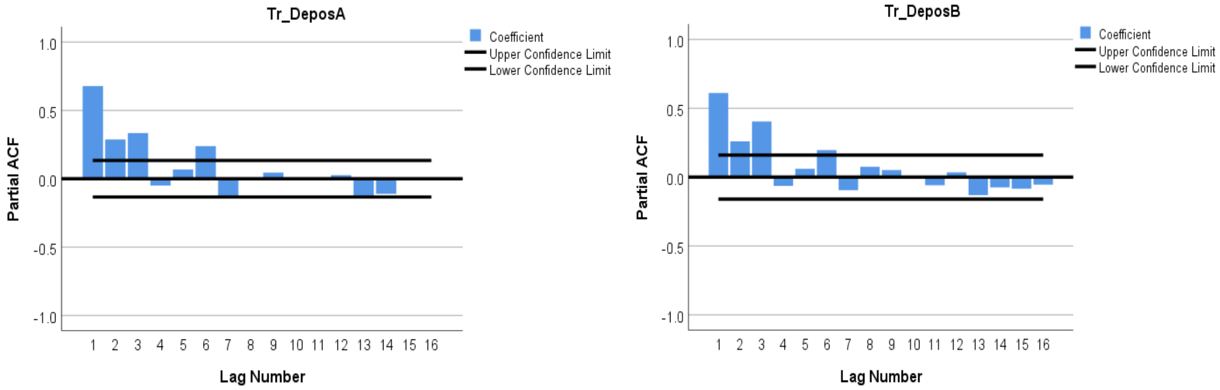

Figure 4.6. ACF and PACF plots of time series Tr_DeposB.

Various ARIMA models were built and tested. From these, models were selected for Tr_DeposA: (4.4) ARIMA(2,1,3) and for Tr_DeposB: (4.5) ARIMA(2,1,0). The statistical estimates of model (4.4) are R2=0.578, MAPE=29.62%, and for model (4.5) R2 =0.492, MAPE=29.59%. The obtained statistical estimates for both models (4.4) and (4.5) are not satisfactory. However, the resulting ARIMA models allow for some initial selection of predictors for conducting regression with more powerful methods.

4.3. A hybrid Arcing-ARIMA method for forecasting bank deposits

In the previous paragraph for the variables Tr_DeposA and Tr_DeposB we identified possible predictors from their lagged variables – up to p, q equal to 6. Also, the presence of a trend (parameter d) in the models (4.4), (4.5) allows us to use the data with trends for predictors. Additionally, to reflect that the data is time dependent, we can also include a nominal month variable to account for the time.

We apply the Arcing method with different numbers of trees in the ensemble. Model training was performed using a standard 10-fold CV. The ratio of the minimum number of cases in a parent node (m1) to the minimum number of cases in a child node is m1=5, m2=5. The number of trees specified in the models is denoted by T. Table 4.10 presents the main indicators of selected Arcing models for forecasting the Tr_DeposA and Tr_DeposB time series. The models of Tr_DeposA are marked with A, those of Tr_DeposB with B, respectively. We define A50 and B60 as the best models. To analyze the adequacy of the selected models A50 and B60, their residuals were examined. The residuals are not within the expected limits, as can be seen from their ACF plots in

Figure 4.11. More significant deviations are observed for the first model.

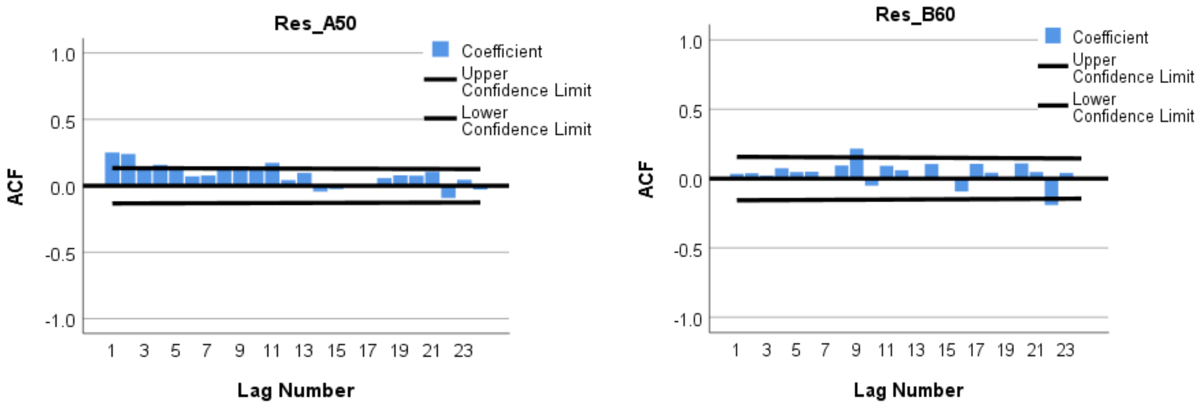

Figure 4.11. ACF plots of the residuals of models A50 and B60.

Due to this reason, we will further model the residuals of A50 and B60 using a univariate ARIMA method. The resulting models are labeled ResA50_ARIMA(1,0,9) and ResB60_ARIMA(0,0,22), respectively. Finally, we add the predicted values of the Res_ARIMA models to the Arcing models A50 and B60 according to (1.26) and obtain hybrid models, denoted as hybrid Arcing-ARIMA (HA50, HB60). We denote the comparison statistics of these hybrid models after retransformation to the initial units by retr_HA50 and retr_HB60. The comparison of these models with all initial data Depos227 and Depos160, respectively, are given in Table 4.12.

Table 4.12. Comparison of the approximation of the retransformed hybrid Arcing-ARIMA models to the full samples Depos227 and Depos160.

Model	RMSE	MAPE	R^2
*retr_HA*50	15.9538	12.5%	0.9854
*retr_HB*60	18.8841	16.22%	0.9813

The models were applied to forecast a sample of Depos three months ahead. The results are as follows: exact values: (80, 79, 170) million dollars calculated by retr_HA50 (123, 76.5, 104). For the reduced retr_HB60 sample, the predictions are (114, 74, 129), respectively. These results are not satisfactory, despite the good statistical performance of the models.

4.4. Hybrid Arcing-ARIMA deposit modeling without data transformation

Since ML methods do not depend on the type of distribution, in this paragraph we will build and investigate Arcing models without data transformation. The goal is to calibrate the models to get more accurate predictions. The same approach is used with trend and lagged variables as predictors. For model training, hyperparameters were varied with number of trees in ensemble T= 30, 40, ..., 150; CV10 and CV5 cross-validation; m1:m2=5:5, and 10:5.

The best results were obtained with three predictors, CV10-fold and m1:m2=5:5. Basic statistics of selected models are given in Table 4.13. A smooth improvement of statistical indicators is observed. We chose AL120 as the best model because for a larger number of trees, R2 and the other statistics improve very little, and the predictions for the test data get worse.

To check the adequacy of the AL120 model, we plot the ACF of its residuals in Figure 4.19. ACF values are critical, and some of them are outside the confidence intervals..

Table 4.13. Arcing Model Depos 224 Statistics without data transformation.

ARC model	Number of trees T	Predictors	R^2	RMSE	MAPE
*AL*30	30	*Depos*224 <1>, <2>, *Trend*224	0.9939	13.662	0.355
*AL*50	50	*Depos*224 <1>, <2>, *Trend*224	0.9957	10.328	0.246
*AL*70	70	*Depos*224 <1>, <2>, *Trend*224	0.9963	9.118	0.234
*AL*100	100	*Depos*224 <1>, <2>, *Trend*224	0.9968	8.208	0.212
***AL*120**	**120**	***Depos*224 <1>, <2>, *Trend*224**	**0.9970**	**7.865**	**0.202**
AL150	150	*Depos*224 <1>, <2>, *Trend*224	0.9970	7.567	0.1931

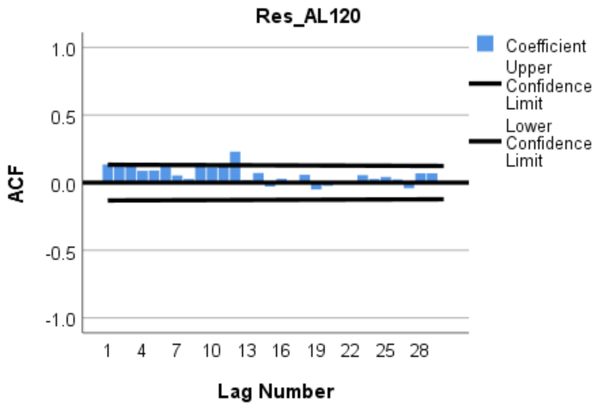

Figure 4.19. ACF of the residiums of the model *AL*120.

To overcome this shortcoming we build a hybrid model by modeling the residuals with ARIMA. The order with the residues is labeled Res_AL120. We built an ARIMA(1,0,1) model for it. We denote the hybrid model by HA_AL120. The residuals of this hybrid model no longer contain autocorrelation, and the statistics of the HA_AL120 hybrid model compared to the entire Depos227 sample are given in Table 4.16.

Table 4.16. *HA_AL120* hybrid model statistics

Model	R^2	RMSE	MAPE	DW
*HA_AL*120	0.9963	7.377	0.129	1.936

The forecast accuracy with the hybrid Arcing-ARIMA model HA_AL120 for the past 24 months is shown in Figure 4.21. To the right of the vertical line are the predictions for the last three months of testing. A very good quality of prediction is reported for the simulated future three data.

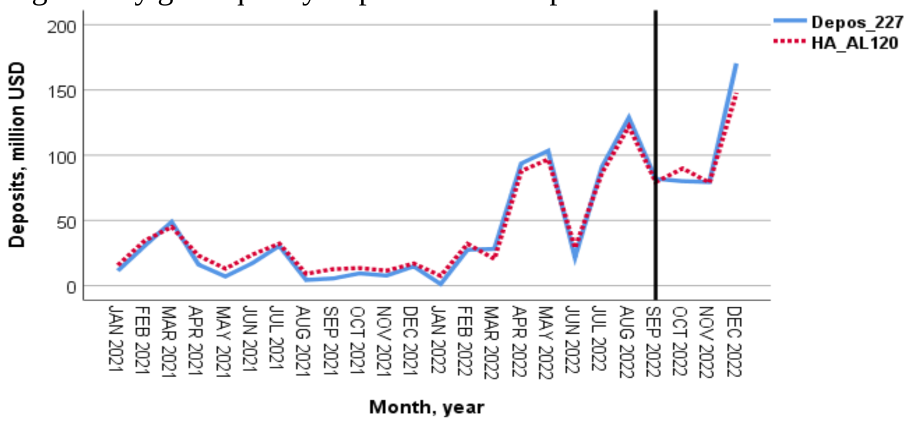

Figure 4.21. Predictions of the hybrid Arcing-ARIMA model *HA_AL120* for test data.

4.5. Modeling bank deposits with the Random Forests method without data transformation

We apply the Random Forests (RF) method for univariate modeling and forecasting to the foreign currency bank deposit data. The data described in paragraph 4.1 are used. All available data for 227 months is denoted by the Depos227 variable. Since the data is not normally distributed, the use of classical parametric methods such as multiple linear regression, ARIMA and others is not recommended. In our case, we will use RF, which has no restrictions on the distribution of variables. We conduct the research without transformation of the source data. Following the results and conclusions of paragraph 4.2, we will use the following time series as predictors: first-order trend and the lagged variables Depos224<1>, Depos224<2>,…, Depos224<6>. The RF hyperparameters are: number of trees (Mtree), minimum observations at a terminal node of the tree (Atm, atoms), number of predictors randomly selected from the total number of predictors (Mtry) separately for each split at each node of each RF tree. Multiple models were built varying the predictors. The hyperparameters were set to Mtree = 500, 700, and 1000; Atm = 2, 5, 10, and 20; Mtry=3, and the models were trained with the standard OOB test sample.

The best training and prediction results were obtained with the following hyperparameters: Mtree = 1000, Atm = 5, Mtry = 3. The final statistics for the three best predictive RF models of Depos224 are given in Table 4.20. Smallest errors and highest R2=98.2% are achieved with RFModel 2. It is generated with the Trend variables and the first two lagged variables Depos224<1>, <2>. Next is RFModel 3, in which Depos224<6> also participates.

Table 4.20. Statistics of the best predictive RF models compared to the full sample *Depos*227.

Model	RMSE	MAPE	R^2
RFModel 1	22.841	25.4%	97.03%
RFModel 2	**17.172**	**17.7%**	**98.22%**
RFModel 3	19.855	22.0%	97.68%

Figure 4.23 shows the RFModel 2 approximation for the entire Depos227 time series. A very good match is observed, except for some peak values. This phenomenon is characteristic of ensemble vitr methods, due to the averaging of the predictions of many trees.

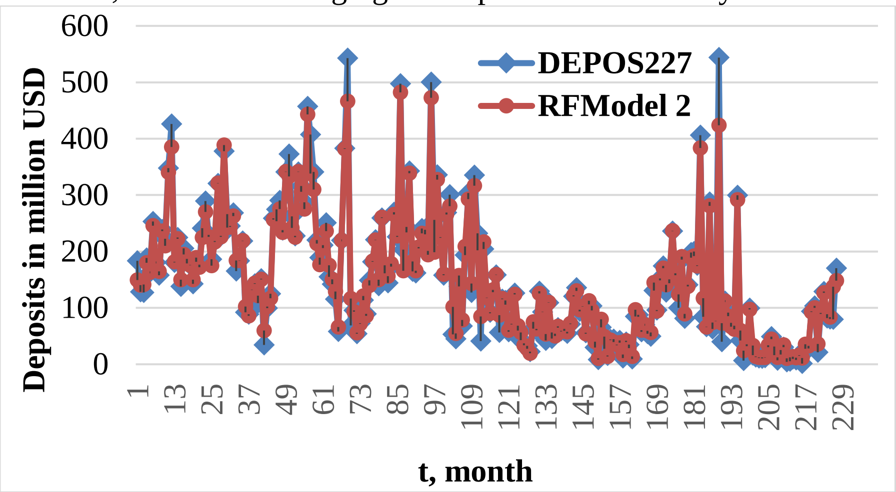

Figure 4.23. Comparison of initial time series of deposit data with forecasts from RFModel 2.

Table 4.21. Relative weights of variables in the RF models and predicted values of the best RF models.

Model	Trend	Depos2 24<1>	<2>	<3>	<6>	Forecast for the last 3 months
RFModel 1	100	71.3	4.7	5.8	2.6	(85.4, 124.5, 118.7)
RFModel 2	100	92.6	3.7	-	-	(83.6, 137.6, 148.3)
RFModel 3	100	81.8	6.0	-	2.5	(84.5, 121.8, 134.1)

The check-up of the adequacy of the model shows that all ACF values of the residuals are within the confidence limits, therefore we can conclude that the selected best RF model RFModel 2 is adequate and the predictions obtained in the last column of Table 4.20 can be used for interpretation.

CONCLUSION

The main conclusion we can make is that the aim of the present dissertation work and the set tasks have been achieved. The main part of the obtained results were published in 4 scientific publications. 4 reports were presented at scientific forums and seminars.

Claims for scientific and scientific-applied contributions

Получените The obtained scientific and scientific-applied results of the dissertation work can be systematized as follows:

1. A general research framework for statistical modeling of univariate and multivariate time series in the field of currency and financial markets has been developed and applied. The framework includes a predictor identification approach and the application of understudied intelligent machine learning methods for forecasting and short-term forecasting.

2. Effective predictive models of one-dimensional time series for foreign currency deposits of Bulgarian citizens have been built and analyzed. For this purpose, the CART Ensembles and Baggigning (EBag) ensemble method has been applied and studied for the first time in this field. The models without data transformation were found to be more accurate and reach a match with the real values up to 97.7%. The models have been applied to obtain forecasts for deposits 3 months ahead and show significantly better results than the reference ARIMA model.

3. Highly efficient multivariate time series forecasting models for the EUR/USD exchange rate depending on macroeconomic factors were built and analyzed using three methods - EBag, Arcing and Random Forests. For the first time in the field of financial and currency markets, the Arcing ensemble method of the gradient amplification class has been applied and studied. The influence of individual macroeconomic factors on the exchange rate has been determined. The obtained modeling results show an accuracy advantage of the Arcing models with a mean absolute percentage error (MAPE) equal to 1% and reaching a data match of up to 99.1%. The one-month-ahead forecasts for all three methods are comparable to the accuracy of the observations.

4. An approach for hybrid Arcing-ARIMA modeling of data for deposits of Bulgarian citizens with transformed and untransformed data was developed and implemented. The resulting hybrid models with untransformed data were found to be more accurate and achieve statistical indicators MAPE=12.9%, matching data up to 99.6%, and most accurate forecasts for the three test "future" months not included in the modeling. These results have been shown to outperform the predictive capabilities of standard ARIMA and Random Forests methods..

The following table shows the relationship between the main results obtained, the set goal and tasks, the contributions and the publications of the dissertation work.

Table. Relation of main results to objectives and publications..

Contribution	Purpose	Task	Paragraph	Publications
1)	1	1	2.1	[P1]
1)	1	2	2.3, 2.5.1, 3.1.3, 4.2, 4.3.1	[P1, P2, P3]
2)	1	3	2.4, 2.5, 4.5	[P1, P4]
3)	1	4	3.2, 3.3, 3.4	[P2]
4)	1	5	4.3, 4.4	[P3]

Approbation of the scientific and scientific-applied results

A) Participation in projects

Some of the results have been tested and used in the following scientific projects:

1. MU21-FMI-015 (2021-2022): "High-performance predictive algorithms with machine learning for modeling data from ecology, industry and education" led by Assoc. Dr. Hristina Kulina, funded by the "Scientific Research" Fund at PU "Paisii Hilendarski" (finished).
2. KP-06-H52/9 (2021-2024): "Research of mathematical and statistical methods with machine learning for intelligent information processing" headed by Assoc. Dr. Hristina Kulina, financed by the Scientific Research Fund at the Ministry of Education and Science (acting).

B) Papers presented at scientific forums

Part of the results obtained in the dissertation work have been reported at the following national and international conferences, scientific sessions and seminars:

1. 13th International Hybrid Conference for Promoting the Application of Mathematics in Technical and Natural Sciences (AMiTaNS'21), 24–29 June 2021, Albena, Bulgaria.
2. Report on the topic: "Statistical models of citizens' savings in foreign currency", seminar of the Department of Mathematical Analysis of Mathematical Modeling and Applications, 06/26/2021 https://www.fmi-plovdiv.org/GetResource?id= 3908
3. Report on the topic: "Ensemble methods for modeling and forecasting the EUR/USD exchange rate", seminar of the Department of Mathematical Analysis of Mathematical Modeling and Applications, November 26-27, 2022, Hisar, https://www. fmi-plovdiv.org/GetResource?id=4372
4. Report on the topic: "Using hybrid methods for forecasting bank deposits of individuals", seminar of the Department of Mathematical Analysis of Mathematical Modeling and Applications, 05/12/2023, https://www.fmi-plovdiv.org/ GetResource?id=4480
5. ACM-2023 6th International Conference on Mathematics and Statistics (ICoMS 2023), July 14-16, 2023, Leipzig, Germany..

Prospects for future work

Further research and publications in the field of the subject of this dissertation can be directed in the following directions:

1. Further development of the framework for identification and selection of predictors based on classical and intelligent methods of processing data from currency markets and the financial sphere.
2. Application of other types of methods with machine learning such as neural networks, etc. in order to forecast temporal lines in real time.
3. Further development and obtaining results with hybrid modeling.
4. Application of the developed approach for research and short-term forecasting of other exchange rates.
5. Publication of the dissertation work in an expanded version as a book.

List of publications related to the dissertation

[P1] Boyoukliev, I.V., Gocheva-Ilieva, S.G., Kulina, H.N., Time series modeling and forecasting of deposits in foreign currency using CART ensemble and bagging, In: Todorov, M.D. (ed.), 13th International Hybrid Conference for Promoting the Application of Mathematics in Technical and Natural Sciences - AMiTaNS'21, 24–29 June 2021, Albena, Bulgaria. AIP Conference Proceedings, vol. 2522, 050003-1–050003-12, 2022; AIP Publishing (American Institute of Physics, Melville, NY), ISBN: 978-0-7354-4361-7. https://doi.org/10.1063/5.0101185. **WoS, Scopus, SJR 0.164**

[P2] Boyoukliev, I.V., Kulina, H.N., Gocheva-Ilieva, S.G., Modelling and forecasting of EUR/USD exchange rate using ensemble learning approach, Cybernetics and Information Technologies, vol. 22(4), pp. 142-151, 2022. Print ISSN: 1311-9702; Online ISSN: 1314-4081. https://doi.org/10.2478/cait-2022-0044. **IF 1.2 (Q3), Scopus (Q2), SJR 0.464**.

[P3] Boyoukliev, I.V., Kulina, H.N., Gocheva-Ilieva, S.G. Forecasting volatility of bank deposits of individuals using hybrid Arcing-ARIMA approach, ACM ICoMS '23: Proceedings of the 2023 6th International Conference on Mathematics and Statistics, pp. 56–62, 2023. https://doi.org/10.1145/3613347.3613356, ISBN: 979-8-4007-0018-7. ACM igital Library. **Scopus**

[P4] Boyoukliev, I.V., Gocheva-Ilieva, S.G. Statistical modeling and forecasting bank deposit data using Random Forests. Sciences of Europe, vol. 129, pp. 124-130, 2023. ISSN: 3162-2364. https://doi.org/10.5281/zenodo.10209391. **Index Copernicus**

DECLARATION OF ORIGINALITY

according to Art. 27, para. 2 of PPZRASRB

by Ivaylo Vladimirov Boyoukliev

a doctoral student at the Department of "Mathematical Analysis"
of the Faculty of Mathematics and Informatics
at "Paisii Hilendarski" University of Plovdiv

In connection with the procedure for obtaining the educational and scientific degree "Doctor" at the "Paisii Hilendarski" University of Plovdiv and the defense of the dissertation work presented by me,

I declare:

The results and contributions of the dissertation research conducted, presented in my dissertation on "**Modeling and Research of Foreign Exchange Financial Markets**" are original.

19.01.2024 DECLARANT :…..……………….
Plovdiv /Ivaylo Boyoukliev/

Acknowledgments

I express my heartfelt gratitude and appreciation to my scientific supervisors Prof. Ph.D. Snezhana Georgieva Gocheva-Ilieva and Associate Professor Dr. Hristina Nikolova Kulina for their useful professional advice, recommendations and benevolent attitude during my doctoral studies and the preparation of my dissertation work.

I would like to thank the professors of the Department of Mathematical Analysis and the management of the Faculty of Mathematics and Informatics of "Paisii Hilendarski" University of Plovdiv for their cooperation, empathy and administrative support during my studies at FMI.

I express my gratitude to the sales representatives of "Bloomberg" for Bulgaria, as well as to all colleagues who in one way or another were sympathetic to my work and my studies as a doctoral student.

Bibliography

[1] Abreu RJ, Souza RM, Oliveira JG. Applying singular spectrum analysis and ARIMA-GARCH for Forecasting: Box-Jenkins (ARIMA) Testing Approach. Econ. Manag. Res., 20(2), 188–198, 2019.

[2] Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716–723, 1974.

[3] Akyildirim E, Nguyen DK, Sensoy A, Šikić M. Forecasting high-frequency excess stock returns via data analytics and machine learning. Eur. Financ. Manag., 29(1), 22–75, 2023.

[4] Al-Gounmeein RS, Ismail MT. Forecasting the exchange rate of the Jordanian dinar versus the US dollar using a Box-Jenkins seasonal ARIMA model, Int. J. Math. Comput. Sci., 15(1), 27–40, 2020.

[5] Alkhatib K, Najadat H, Hmeidi I, Shatnawi MKA. Stock price prediction using k-nearest neighbor (kNN) algorithm. Int. J. Business, Human. Technol., 3(3), 32–44, 2013.

[6] Asare-Frempong J, Jayabalan M. Predicting customer response to bank direct telemarketing campaign. Proc. Int. Conf. Eng. Technol. Technopreneurship (ICE2T), pp. 1–4, 2017, IEEE.

[7] Babu AS, Reddy SK. Exchange rate forecasting using ARIMA, neural network and fuzzy neuron. Journal of Stock & Forex Trading, 3(4), 1–5, 2015.

[8] Baestaens DE, Van den Bergh WM, Vaudrey H. Market inefficiencies, technical trading and neural networks. In Forecasting Financial Markets, Financial Economics and Quantitative Analysis, Christian D (ed.). Wiley: Chichester; 254–260, 1996.

[9] Basak S, Kar S, Saha S, Khaidem L, Dey SR. Predicting the direction of stock market prices using tree-based classifers. The North Amer. J. Econ. Finance, 47, 552–567, 2019.

[10] Batmaz İ, Danışoğlu S, Yazıcı C, Kartal-Koç E. A data mining application to deposit pricing: Main determinants and prediction models. Appl. Soft Comput. 60, 808–819, 2017.

[11] Bauer E, Kohavi R. An empirical comparison of voting classification algorithms: bagging, boosting, and variants. Mach. Learn., 36, 105–139, 1999.

[12] Bikker A, Gerritsen DF. Determinants of interest rates on time deposits and saving accounts: Macro factors bank risk and account features. Int. Rev. Finance, 18, 169–216, 2018.

[13] BIS Quarterly Review, International banking and financial market developments. December 2022. https://www.bis.org/publ/qtrpdf/r_qt2212.pdf

[14] Bordo M, Owen H, Anna JS. U.S. Intervention during the Bretton Wood Era: 1962-1973, Working Paper 11-08, Federal Reserve Bank of Cleveland, Cleveland, Ohio, April 2011.

[15] Box GEP, Jenkins GM, Reinsel GS, Ljung GM. Time Series Analysis, Forecasting and Control, 5th edn., John Wiley & Sons. Inc, New Jersey 2016.
[16] Boye P, Ziggah YY. A short-term stock exchange prediction model using Box-Jenkins approach, J. Appl. Math. Phys., 08, 05, Art. no 99774, 1–14, 2020.
[17] Breiman L, Friedman JH, Stone CJ, Olshen RA. Classification and Regression Trees. Chapman & Hall Wadsworth, Inc., New York, 1984.
[18] Breiman L. Arcing classifiers. Ann. Stat., 26 (3), 801-849, 1998
[19] Breiman L. Bagging predictors. Mach. Learn., 24(2), 123–140, 1996.
[20] Breiman L. Random forests. Mach. Learn., 45(1), 5–32, 2001.
[21] Breiman L. Stacked regressions. Mach. Learn., 24, 49–64, 1996.
[22] Bulgarian National Bank, www.bnb.bg.
[23] Byrne JP, Korobilis D, Ribeir PJ. Exchange rate predictability in a changing world. Journal of International Money and Finance, 62, 1–24, 2016.
[24] Chatfield C. Time-series forecasting. Chapman & Hall CRC, Boca Raton, 2000.
[25] Cheung Y-W, Chinn MD, Pascual AG. Empirical exchange rate models of the nineties: Are any fit to survive? Journal of International Money and Finance, 24, 1150–75, 2005.
[26] Chihab Y, Bousbaa Z, Chihab M, Bencharef O, Ziti S. Algo-trading strategy for intraweek foreign exchange speculation based on random forest and probit regression. Applied Computational Intelligence and Soft Computing, Art. no. 8342461, 1–13, 2019.
[27] Chortareas G, Jiang Y, Nankervis JC. Forecasting exchange rate volatility using high-frequency data: Is the euro different? Int. J. Forecasting, 27(4), 1089–1107, 2011.
[28] Cutler A, Cutler DR, Stevens JR. Random forests. In: Zhang C, Ma Y (eds) Ensemble Machine Learning: Methods and Applications, pp.157–175, 2012.
[29] De Souza Vasconcelos C, Hadad Júnior E. Forecasting exchange rate: A bibliometric and content analysis. International Review of Economics & Finance, 83(1), 607–628, 2023.
[30] Deka A, Resatoglu NG. Forecasting Foreign Exchange Rate and Consumer Price Index with ARIMA Model: The Case of Turkey. Int. J. Sci. Res. Manag., 7(8), 1254–1275, 2019.
[31] Dhakar DS, Shiwani S. Stock market prediction employing discrete wavelet transform and moving average gradient descent. Int. Conf. Cogn. Sci. Technol. ICCIC, 617–629, 2023.
[32] Dickey DA, Fuller WA. Distribution of the estimators for autoregressive time series with a unit root. Journal of American Statistical Association, 74(366a), 427–431, 1979.
[33] Ding J, Tarokh V, Yang Y. Model selection techniques: An overview. In IEEE Signal Processing Magazine, 35, no. 6, pp. 16–34, 2018.
[34] Doumpos M, Zopounidis C, Gounopoulos D, Platanakis E, Zhang W. Operational research and artificial intelligence methods in banking. Eur. J. Oper. Res., 306, 1, 1–16, 2023.
[35] Dunis CL, Laws J, Sermpinis G. Higher order and recurrent neural architectures for trading the EUR/USD exchange rate. Quantitative Finance, 11(4) 615-629, 2011.
[36] Effendy MH, Anggraeni D, Dewi YS, Hadi AF. Classification of bank deposit using Naïve Bayes classifier (NBC) and K–Nearest Neighbor (K-NN). Proc. Int. Conf. Math., Geometry, Statistics, and Comput. (IC-MaGeStiC 2021), pp. 163–166, 2022.
[37] Einzig P. The history of Foreign Exchange. Hardcover, 1970.
[38] El Morr C, Jammal M, Ali-Hassan H, El-Hallak W. Overview of machine learning algorithms. In Machine Learning for Practical Decision Making. 334, 61–115, 2022.
[39] Elliott G, Timmermann A. Handbook of Economic Forecasting 2. North Holland, 2013.
[40] Fabozzi FJ, Mann SV. The Handbook of Fixed Income Securities-McGraw. Hill Education, 2012.
[41] Farhan AK, Fakhir MR. Forecasting the Exchange Rates of the Iraqi Dinar against the US Dollar using the Time Series model (ARIMA). Int. J. Eng. Manag. Res., 9(6), 51–55, 2019.
[42] Federal Reserve History. Available online at https://www.federalreservehistory.org/ 2020.
[43] Field A. Discovering Statistics Using SPSS, 3rd ed. London: Sage Publications, 2009.

[44] Galeshchuk S. Neural networks performance in exchange rate prediction. Neurocomputing, 172, 446–452, 2016.
[45] Geman S, Bienenstock E, Doursat R. Neural networks and the bias/variance dilemma. Neural computation, 4(1), 1–58, 1992.
[46] Ghoalamian E, Davoodi SMR. Predicting the direction of stock market prices using random forest. Financ. Eng. Secur. Manag., 1397, 9(35), 301–322, 2018.
[47] Gocheva-Ilieva S, Yordanova A, Kulina H. Predicting the 305-day milk yield of Holstein-Friesian cows depending on the conformation traits and farm using simplified selective ensembles. Mathematics, 10(8), art. 1254, 2022.
[48] Gocheva-Ilieva SG, Ivanov AV, Voynikova DS, Boyadzhiev DT. Time series analysis and forecasting for air pollution in small urban area – an SARIMA and factor analysis approach. Stoch. Env. Res. Risk Assess., 28 (4), 1045–1060, 2014.
[49] Gocheva-Ilieva SG, Voynikova DS, Stoimenova MP, Ivanov AV, Iliev IP. Regression trees modeling of time series for air pollution analysis and forecasting. Neural Computing and Applications, 31(12), 9023–9039, 2019.
[50] Gurav U, Sidnal N. Predict stock market behavior: Role of machine learning algorithms. Intelligent Comput. Inform. Commun., Springer, 673, 383–394, 2018.
[51] Hajirahimi Z, Khashei M. Hybridization of hybrid structures for time series forecasting: a review. Artificial Intelligence Review, 56, 1201–1261, 2022.
[52] Hastie T, Tibshirani R, Friedman JH. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., New York, Springer, 2009.
[53] Henrique B, Sobreiro V, Kimura H. Literature review: Machine learning techniques applied to financial market prediction. Exp. Syst. Appl., 124, 226–251, 2019.
[54] Hota S, Mohanty AK, Satapathy P, Mishra D. Currency exchange rate forecasting using FLANN-WCA model. Proc. Int. Conf. Appl. Mach. Learn. (ICAML), 2019.
[55] Htun HH, Biehl M, Petkov N. Survey of feature selection and extraction techniques for stock market prediction. Financial Innovation, 9(26), 2023.
[56] Humphrey F, Mercy FM, Wilson F. Forecasting ZMK/USD Exchange Rate with ARIMA Model, Amer. Int. J. Res. Human. Arts and Soc. Sci. , 5(15), 5–10, 2015.
[57] IBM SPSS Software. https://www.ibm.com/products/spss-statistics
[58] Imece S, Cetinkaya M, Koroglu B. A comparative study of machine learning algorithms for time deposit pricing. Proc. 3rd Int. Inform. Soft. Eng. Conf. (IISEC), IEEE, pp. 1–6, 2022.
[59] Ince H, Trafalis TB. A hybrid model for exchange rate prediction. Decision Support Systems, 42, 2, 1054–1062, 2006.
[60] Islam MS, Hossain E, Rahman A, Hossain, MS, Andersson K. A review on recent advancements in FOREX currency prediction. Algorithms, 13(8), 186, 2020.
[61] Izenman J. Modern Multivariate Statistical Techniques. Regression, Classification, and Manifold Learning. Springer, New York, 2008.
[62] Jian ZP, Deng K, Luo Z, Zhu. The effect of market quality on the causality between returns and volatilities: Evidence from CSI 300 index futures. Journal of Management Science and Engineering, 3(1), 16-38, 2018.
[63] Joshi VK, Band G, Naidu K, Ghangare A. Modeling exchange rate in India–empirical analysis using ARIMA model. Studia Rosenthaliana, 12(3), 13–26, 2020.
[64] Kim JM, Kim DH, Jung H. Applications of machine learning for corporate bond yield spread forecasting. The North Amer. J. Economics and Finance, 58, 101540, 2021.
[65] Krauss C, Do XA, Huck N. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. Eur. J. Oper. Res., 259(2), 689–702, 2017.
[66] Krogh A, Sollich P. Statistical mechanics of ensemble learning. Phys. Rev. E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics, 55(1), 811–825, 1997.
[67] Kumbure MM, Lohrmann C, Luukka P, Porras J. Machine learning techniques and data for stock market forecasting: A literature review. Expert Systems with Applications 197, 2022.

[68] Kuncheva LI, Combining Pattern Classifiers: Methods and Algorithms. John Wiley and Sons, Inc., Hoboken, 2004.
[69] Lewis CD. Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting. Butterworth Scientific-Heinemann, London, 1982.
[70] Lin Y, Liu S, Yang H, Wu H. Stock trend prediction using candlestick charting and ensemble machine learning techniques with a novelty feature engineering scheme. IEEE Accses, 9, 9481924, 101433–101446, 2021.
[71] Liu C, Hou D, Liu D. Foreign exchange rates forecasting with convolutional neural network. Neural Processing Letters, 46(3), 1095–1119, 2017.
[72] Liu F, Pantelous AA, von Mettenheim HJ. Forecasting and trading high frequency volatility on large indices. Quantitative Finance, 18(5), 737–748, 2018.
[73] Livieris I, Kotsilieris T, Stavroyiannis S, Pintelas P. Forecasting stock price index movement using a constrained deep neural network-training algorithm. Intelligent Decision Technologies, IOS Press 14(3), 313–323, 2020.
[74] Ljung GM, Box GEP. On a measure of a lack of fit in time series models. Biometrika, 65(2), 297–303, 1978.
[75] Mienye ID, Sun YA. Survey of ensemble learning: Concepts, algorithms, applications, and prospects. IEEE Access. Sep 16(10), 99129–99149, 2022.
[76] Mills TC. Applied Time Series Analysis: A Practical Guide to Modeling and Forecasting. Academic press, 2019.
[77] Mishkin FS. Economics of Money, Banking, and Financial Markets, 10th ed. Prentice Hall, 2012.
[78] Moosa IA, Vaz JJ. Cointegration, error correction and exchange rate forecasting. Journal of International Financial Markets. Institutions and Money, 44, 21–34, 2016.
[79] Mussa M. A model of exchange rate dynamics. Scand. J. Econ., 78, 229–48, 1976.
[80] Mustapa FH, Ismail MT. Modelling and forecasting S&P 500 stock prices using hybrid Arima-Garch Model. Journal of Physics: Conference Series, 1366, 012130, 2019.
[81] Nakagawa K, Yoshida K. Time-series gradient boosting tree for stock price prediction. International Journal of Data Mining, Modelling and Management, 14(2), 110–125, 2022.
[82] Nesreen KA, Amir FA, Neamat El G, Hisham El S. An empirical comparison of machine learning models for time series forecasting. Econom. Reviews 29, (5-6), 594–621, 2010.
[83] Neter J, Kutner MH, Nachtsheim CJ, Wasserman W. Applied Linear Statistical Models. WCB/McGraw-Hill, 1996.
[84] Opitz D, Maclin R. Popular ensemble methods: An empirical study. Journal of Artificial Intelligence Research, 11, 169–198, 1999.
[85] Padhi DK, Padhy N, Bhoi AK, Shafi J, Ijaz MF. A Fusion Framework for Forecasting Financial Market Direction Using Enhanced Ensemble Models and Technical Indicators. Mathematics 9(21), 2021.
[86] Pandey TN, Jagadev AK, Dehuri S, Cho S-B. A review and empirical analysis of neural networks based exchange rate prediction. Intelligent Decision Technologies, 12(4), 423–439, 2019.
[87] Pankratz A. Forecasting with univariate Box-Jenkins models: concepts and cases. Wiley, New York, 1983.
[88] Paravee M, Wilawan S. Forecasting foreign exchange markets: further evidence using machine-learning models. Soft Computing 25(12), 7887–7898, 2021.
[89] Parkin M. Economics, 10th ed. Prentice Hall, 2011.
[90] Patwary MJ, Akter S, Alam MB, Karim AR. Bank deposit prediction using ensemble learning. Artificial Intelligence Evolution, 2(2), 42–51, 2021.
[91] Petropoulos F, Apiletti D, Assimakopoulos V, Babai MZ, Barrow DK, Taieb SB, Bergmeir C, Bessa RJ, Bijak J, Boylan JE, Browell J et al. Forecasting: theory and practice. Int. J. Forecasting, 38(3), 705–871, 2022.

[92] Pinelis M, Ruppert D. Machine learning portfolio allocation. The Journal of Finance and Data Science, 8, 35–54, 2022.
[93] Plakandaras V, Papadimitriou T., Gogas P. Forecasting Daily and Monthly Exchange Rates with Machine Learning Techniques. Journal of Forecasting, 34(7), 560–573, 2022.
[94] Pradeepkumar D, Ravi V. Forecasting financial time series volatility using Particle Swarm Optimization trained Quantile Regression Neural Network. Applied Soft Computing Journal, 58, 35–52, 2017.
[95] Pradeepkumar D, Ravi V. Forex rate prediction: A hybrid approach using chaos theory and multivariate adaptive regression splines, Proc. 5th Int. Conf. Frontiers in Intelligent Computing: Theory and Applications, Springer, Berlin, pp, 515, 219–227, 2017.
[96] Pradeepkumar D, Ravi V. Soft computing hybrids for FOREX rate prediction: A comprehensive review. Computers & Operations Research, 99, 262–284, 2018.
[97] Psaradellis I, Sermpinis G. Modelling and trading the U.S. implied volatility indices. Evidence from the VIX, VXN and VXD indices. Int. J. Forecast., 32 (4), 1268–1283, 2016.
[98] Radwan M, Drissi S, Secinaro S. Machine learning in the fields of accounting, economics and finance The emergence of new strategies. The Essentials of Machine Learning in Finance and Accounting, 181–198, 2021.
[99] Ratchagit M, Xu H. A Two-Delay Combination Model for Stock Price Prediction. Mathematics, 10(19), 3447, 2022.
[100] Rossi B. Exchange Rate Predictability. J. Econ. Literature 51, 1063–119, 2013.
[101] Sadeghi A. Daneshvar A, Zaj M. Combined ensemble multi-class SVM and fuzzy NSGA-II for trend forecasting and trading in Forex markets. Exp. Syst. Appl., 185, 115566, 2021.
[102] Sagi O., Rokach L. Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1249, 2018.
[103] Salford Predictive Modeler (SPM), Machine Learning and Predictive Analytics Software, https://www.minitab.com/en-us/products/spm/.
[104] Santuci A, Sbruzzi E, Filho LA, Leles M. Evaluation of FOREX trading strategies based in random forest and support vector machines. IEEE Latin America Trans., 20, 9, 2022.
[105] Schwarz G, Estimating the dimension of a model. Annals of Statistics, 6, 461–464, 1978.
[106] Sermpinis G, Dunis C, Laws J, Stasinakis C. Forecasting and trading the EUR/USD exchange rate with stochastic Neural Network combination and time-varying leverage. Decision Support Systems, 54(1), 316–329, 2012.
[107] Sermpinis G, Stasinakis C, Theofilatos K, Karathanasopoulos A. Modeling, forecasting and trading the EUR exchange rates with hybrid rolling genetic algorithms - Support vector regression forecast combinations. European Journal of Operational Research, 247 (3), art. no. 13059, 831–846, 2015.
[108] Shen F, Chao J, Zhao J. Forecasting exchange rate using deep belief networks and conjugate gradient method. Neurocomputing, 167, 243–253, 2015.
[109] Sheth D, Shah M. Predicting stock market using machine learning: best and accurate way to know future stock prices. Int. J. Syst. Assur. Eng. Manag. 14, 1–18, 2023.
[110] Sollich P, Krogh A. Learning with ensembles: How overfitting can be useful, Advances in Neural Information Processing Systems, 8, 190–196, 1995.
[111] Steinberg D. CART: Classification and regression trees, in The Top Ten Algorithms in Data Mining, Chapman and Hall/CRC, Boca Raton, pp. 179–202, 2009.
[112] Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. BMC Bioinformat. 2008, 9(1), 307.
[113] Sun S, Wang S, Wei Y. A new ensemble deep learning approach for exchange rates forecasting and trading. Advanced Engineering Informatics, 46, art. no 101160, 2020.
[114] Sun SL, Wang SY, Wei YJ, Zhang GW. A clustering-based nonlinear ensemble approach for exchange rates forecasting. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 50(6), 2284–2292, 2018.

[115] Tak ALC, Logeswaran R. Foreign Currency Exchange Market Prediction using Machine Learning Techniques. Proc. IEEE Int. Conf. Distributed Computing and Electrical Circuits and Electronics (ICDCECE), pp. 1–5, 2022.
[116] Thu TNT, Xuan VD. Using support vector machine in FoRex predicting. Proc. IEEE Int. Conf. Innovative Research and Development (ICIRD), Bangkok, Thailand, pp. 1–5, 2018.
[117] Ting K, Witten IH. Issues in stacked generalization. Journal of Artificial Intelligence Research, 10, 271–289, 1999.
[118] Tran MUN. Forecasting foreign exchange rate by using ARIMA model: A Case of VND/USD exchange rate. Res.J. Finance Accounting, 7(12), 38–44, 2016.
[119] Ueda N, Nakano R. Generalization error of ensemble estimators. Proc. of IEEE Int. Conf. on Neural Networks (ICNN'96), Washington, DC, USA, 1, pp. 90–95, 1996.
[120] Umar S, Abubakar SS, Salihu AM, Umar Z. Modelling naira/pounds exchange rate volatility: application of Arima and garch models. Int. J. Eng. Appl. Sci. Technol., 4(8), 238–242, 2019.
[121] Wei WWS. Time Series Analysis: Univariate and Multivariate Methods. 2nd edn., Pearson Education, Boston. 2006.
[122] Wei, L,Wan, S., Guo, J., Wong, K.K. A novel hierarchical selective ensemble classifier with bioinformatics application. Artif. Intel. Med. 2017, 83, 82–90
[123] Wolfram Mathematica, https://www.wolfram.com/mathematica/.
[124] Wolpert D, Stacked generalization. Neural Networks 1992, 5, 241–259.
[125] Wolpert DH, Macready WG. No free lunch theorems for optimization. IEEE Transactions on Evolutionary Computation, 1(1), 67–82, 1997.
[126] Yan H, Ouyang H. Financial time series prediction based on deep learning, Wireless Personal Communications, 102(2), 683–700, 2018.
[127] Yang Z, Yang Y, Yin D, Yang M, Li L. Research on data analysis for time deposit of bank customers based on ensemble learning, IEEE 3rd Int. Conf. on Electronic Information Technology and Computer Engineering, EITCE, 9094858, pp. 1325–1330, 2019.
[128] Yeo IK, Johnson RA. A new family of power transformations to improve normality or symmetry. Biometrika, 87(4), 954–959, 2000.
[129] Yordanova L, Kiryakova G, Veleva P, Angelova N, Yordanova A. Criteria for selection of statistical data processing software. IOP Conference Series: Materials Science and Engineering, 1031(1), 012067. IOP Publishing, 2021.
[130] Yu L, Lai KK, Wang SY. Multistage RBF neural network ensemble learning for exchange rates forecasting. Neurocomputing, 71(16-18), 3295–3302, 2008.
[131] Zhang Y, Hamori S. The predictability of the exchange rate when combining machine learning and fundamental models. J Risk Financ. Manag., 13(3), 48, 2020.
[132] Zhou ZH, Tang W. Selective ensemble of decision trees. In Proc.Int. Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing, RSFDGrC 2003. Lecture Notes in Computer Science, 2639, 476–483, 2003. Springer, Berlin, Heidelberg.
[133] Zhou ZH. Ensemble Methods: Foundations and Algorithms. CRC Press, Boca Raton, 2012.
[134] Норт М. Парите и тяхната история. ISBN 954-607-601-5, 160-169. ЛИК, 2003.
[135] Петранов С. Финансови пазари - функции, организация, инфраструктура и инвеститори. Изд. “СУ Св. Климент Охридски”, 2012.