

Пловдивски университет „Паисий Хилендарски“
Факултет по математика и информатика
Катедра „Математически анализ“

Ивайло Владимиров Боюклиев

**МОДЕЛИРАНЕ И ИЗСЛЕДВАНЕ НА
ВАЛУТНИ ФИНАНСОВИ ПАЗАРИ**

АВТОРЕФЕРАТ

на дисертационен труд
за присъждане на образователна и научна степен “Доктор”
в област на висше образование 4. Природни науки, математика и информатика
Професионално направление 4.5 Математика
Докторска програма: Математическо моделиране и приложение на математиката

Научни ръководители:

проф. д.н. Снежана Георгиева Гочева-Илиева

доц. д-р Христина Николова Кулина

Пловдив 2024 г.

Дисертационният труд е обсъден и насрочен за защита на разширен катедрен съвет на катедра „Математически анализ” при Факултет по математика и информатика на Пловдивския университет „Паисий Хилендарски”, град Пловдив, проведен на 15.01.2024 г.

Дисертационният труд е с общ обем от 160 страници, в които 46 таблици и 68 фигури, с увод, четири глави, заключение и библиография, състояща се от 135 източника. Списъкът на авторските публикации включва 4 заглавия.

Защитата на дисертационния труд ще се състои на 12.04.2024 от часа в Заседателна зала на Нова сграда на Пловдивския университет „Паисий Хилендарски” на открито заседание на научно жури в състав:

1. Проф. д-р Христо Стефанов Кискинов - председател
2. Проф. д-р Михаил Димов Тодоров
3. Проф. д.н. Леда Димитрова Минкова
4. Проф. д.н. Миглена Николаева Колева
5. Доц. д.н. Иванка Миткова Желева

Номерацията на формулите, цитиранията, примерите, таблиците и фигурите съвпада с тяхната номерация в дисертационния труд.

Материалите по защитата са на разположение на интересуващите се в секретариата на ФМИ, нова сграда на ПУ, каб. 330, всеки работен ден от 8:30 до 17:00 часа.

Автор на дисертационния труд: Ивайло Владимиров Боюклиев

Заглавие: Моделиране и изследване на валутни финансови пазари

Съдържание

ОБЩА ХАРАКТЕРИСТИКА НА ДИСЕРТАЦИОННИЯ ТРУД	4
Актуалност на проблемите	4
Цел и задачи на дисертационния труд	4
Структура на дисертационния труд	5
КРАТКО СЪДЪРЖАНИЕ НА ДИСЕРТАЦИОННИЯ ТРУД	6
ГЛАВА 1. ВЪВЕДЕНИЕ	6
1.1. Основни понятия за временен ред	6
1.2. Използвани методи за моделиране и изследване на валутни финансови пазари	7
ГЛАВА 2. МОДЕЛИРАНЕ НА ДЕПОЗИТИ В ЧУЖДЕСТРАННА ВАЛУТА С АНСАМБЛОВИ МЕТОДИ	9
2.1. Обща рамка на изследванията в дисертационния труд	9
2.2. Описание на данните, начална обработка и построяване на референтен ARIMA модел за идентифициране на предиктори	9
2.3. Приложение на ансамблов CART метод за данни за валутни депозити на български граждани	10
2.4. Приложение на EBag метода за Depos204 без трансформиране на данните	11
ГЛАВА 3. МОДЕЛИРАНЕ НА ВАЛУТНИЯ КУРС EUR/USD	13
3.1. Използвани данни за анализ и прогнозиране на валутния курс EUR/USD и начална обработка	13
3.2. Приложение на CART Ensembles and Bagging за моделиране на валутния курс	15
3.3. Приложение на Arcing за моделиране на данните за валутния курс	15
3.4. Моделиране на валутния курс с Random Forests	17
ГЛАВА 4. МОДЕЛИРАНЕ НА БАНКОВИ ДЕПОЗИТИ С ХИБРИДНИ МЕТОДИ	18
4.1. Описание на проблема и данните	18
4.2. Построяване и анализ на референтни едномерни ARIMA модели на Tr_DeposA и Tr_DeposB, и идентифициране на предиктори	19
4.3. Хибриден Arcing -ARIMA метод за прогнозиране на банковите депозити	19
4.4. Хибридно Arcing-ARIMA моделиране на депозити без трансформация на данните	20
4.5. Моделиране на банковите депозити с метода Random Forests без трансформация на данните	22
ЗАКЛЮЧЕНИЕ	23
Претенции за научни и научно-приложни приноси	23
Апробация на научните и научно-приложни резултати	24
Списък на публикациите по дисертационния труд	25
Д Е К Л А Р А Ц И Я за оригиналност	26
Благодарности	27
Библиография	27

ОБЩА ХАРАКТЕРИСТИКА НА ДИСЕРТАЦИОННИЯ ТРУД

Актуалност на проблемите

В настоящия дисертационен труд се развиват и прилагат както класически, така и най-съвременни интелигентни методи с машинно обучение (МО) за статистическо моделиране и анализ на данни в приложната област на валутните и финансови пазари.

Резултатите от изследването, моделирането, анализа и прогнозирането на финансовите и валутните пазари намират широко приложение в ежедневната дейност на редица участници както на тези пазари, така и на всички останали икономически области. С цел повишаване на качеството на вземаните решения, тяхната научна обосновааност, системност и достоверност през последните години се наблюдава все по-интензивно прилагане на статистическо моделиране предимно с интелигентни методи.

Заедно с това, в научно-приложен аспект съществуват редица актуални задачи за по-нататъшното развитие на моделирането и изследването на временни редове. Те произтичат от факта, че за получаване на адекватни прогнози на изследваната зависима променлива напред във времето е необходимо да разполагаме със стойности на предикторите в същия бъдещ времеви период. Освен това, често изцяло липсват предиктори или намирането им е силно затруднено. Друг труден, но важен за практиката въпрос е идентифицирането на факторите, които имат най-силно влияние в процеса на прогнозирането.

Цел и задачи на дисертационния труд

Основната цел на настоящия дисертационен труд е развитие и приложение на нови подходи за моделиране на временни редове, базирани на стохастични и мощни интелигентни методи за изследване, анализ и прогнозиране на валутни финансови пазари.

Обект на изследванията са реално съществуващи исторически данни от финансовите пазари и системи, използвани като основа за прогнозиране на бъдещите им движения, стойности и динамика, както и за определяне на действащи инвестиционни стратегии в актуалните финансови системи и софтуер.

За постигане на поставената цел на дисертационния труд се формулираха следните **задачи**:

1. Разработка на обща рамка за моделиране на едномерни и многомерни временни редове с интелигентни методи, базирани на машинно обучение.
2. Изследване на временни редове от сферата на валутните финансови пазари и развитие на методи за избор на предиктори с цел кратковременни бъдещи прогнози.
3. Приложение на ансамблови методи за моделиране и прогнозиране на едномерни временни редове за депозитите в чуждестранна валута на българските граждани.
4. Приложение на ансамблови методи за моделиране на многомерни временни редове за валутни курсове и макроикономически показатели.
5. Развитие на хибриден подход с ARIMA корекция за моделиране на колебанията в банковите депозити на граждани.

Научни хипотези

1. Приложение на разработената рамка за анализ и моделиране на едномерни временни редове с данни за банкови депозити;
2. Получаване на предварителна оценка и предсказване на динамиката на валутния курс EUR/USD напред във времето чрез моделиране, анализиране и прогнозиране на многомерни времеви редове за финансови данни от стохастичен характер;
3. Установяване на зависимостта на валутните курсове от различни основни макроикономически показатели като основен лихвен процент (ОЛП), брутен вътрешен продукт (БВП), инфлация, доходност на 10г. държавен дълг и други;

4. Демонстриране на производителността и мощта на приложенияте за първи път от нас ансамблови методи за данни за финансовия и банковия сектор;
5. Постигане на голяма точност на прогнозите при изготвяне на моделите, съизмерима с точността на наблюденията.

Структура на дисертационния труд

Дисертационният труд съдържа увод, 4 глави, заключение, декларация и библиография с общ обем от 160 печатни страници.

В Глава 1 са представени основните елементи и етапи на статистическото моделиране, дефинирана е приложната област, направен е подробен литературен обзор по темата на дисертационния труд, обяснени са използваните методи, посочен е помощният софтуер и са формулирани целите и задачите на изследванията.

В Глава 2 е описана разработената обща рамка на изследванията, развит е нов подход за моделиране и прогнозиране на едномерни времеви редове от финансовия сектор с официални банкови данни от Българската национална банка за краткосрочните депозити на български граждани в USD валута. Подходът се състои в детайлно изследване на авторегресионните и частично авторегресионните функции (ACF и PACF) на началния временен ред, разглеждан като зависима променлива, с цел идентифициране на възможни предикторни редове. Изследва се наличието на мултиколинеарност между намерените предиктори и вида на разпределението. С помощта на кандидатите за предиктори се построяват модели с ансамбовия метод с машинно обучение CART Ensembles and Bagging (EBag). Построяват се и се оценяват EBag модели в два случая – с предварителна трансформация на данните за подобряване на разпределението към нормалното и без трансформация на данните. Изследването на реда, ACF и PACF показва възможност за използване на 6 лагирани променливи и тренд от първи ред като кандидат-предиктори. За да се удовлетворят условията за ARIMA, началният времеви ред беше трансформиран с формулата на Йео-Джонсън. Построен и изследван е референтен модел ARIMA(6,1,1). За моделирането с CART Ensembles and Bagging са построени и анализирани множество модели чрез калибриране на хиперпараметрите за обучение на моделите. Най-добрият модел с трансформирани данни е ансамбъл с 60 CART дървета W60a, който постигна коефициент на детерминация $R^2=94.3\%$, RMSE=28.8 млн. долара и MAPE=13.8%. От EBag моделите без трансформация на данните моделите са с по-добри статистики. Избраният най-добър EBag модел EYU30, с $R^2=97.7\%$, RMSE=18.2 млн. долара и MAPE=10.2%.

В Глава 3 е развит нов подход за моделиране и краткосрочно прогнозиране на динамиката на валутния курс евро/американски долар (EUR/USD) като функция на 8 основни макроикономически показатели. Моделирането се провежда по общата рамка на изследванията. В първия етап са анализирани компонентите на многомерните временни редове с едномерен ARIMA метод, статистически тестове и други анализи. На втори етап са приложени и сравнени резултатите от три различни ансамбови метода с MO: CART-Ensembles and bagging (EBag), Arcing и Random Forests (RF). Получените резултати са такива, че се определят предиктори от стойностите на зависимата променлива $Y=EUR/USD_FX_RATE$ и лагираните променливи на осемте макроикономически фактори X_1, X_2, \dots, X_8 . Прилагат се трите ансамбови метода EBag, Arcing и RF за моделиране и прогнозиране на валутния курс с един месец напред. С всеки от ансамбловите методи са построени и анализирани множество модели. От тях са представени по четири избрани модела, с използване на четири групи предиктори. За EBag моделите с най-добри статистически показатели са получени с модел G4, който съвпада с данните за валутния курс с $R^2=98.8\%$ и има статистики за точност RMSE=0.0174, MAPE=1.03%. От Arcing моделите с най-добрите показатели е модел AG4 с $R^2=99.1\%$, RMSE=0.0152, MAPE=1.00%. От RF моделите, най-добър е модел RG4 с: $R^2=98.6\%$, RMSE=0.0194, MAPE=1.20%.

Адекватността на моделите е установена с диагностика на резидуумите им с теста на Люнг-Бокс. Статистиките MAPE на описаните модели варират от 1% до 1.2%, т.е. са под 10%, с което можем да класифицираме получените модели като много точни. .

В Глава 4 са развити и изследвани хибридни методи. Изследван е едномерен динамичен ред с месечни данни за депозити на български лица в щатски долари за период от 227 месеца. Моделираните данни са разделени на две части, от които по-големите части с наблюдения са използвани за изграждане на моделите, а последните три стойности са оставени за тест и проверка на точността на моделите да предсказват „нови данни“. Най-напред са построени хибридни Arcing-ARIMA модели с трансформирани данни поотделно за две извадки от данни – извадка с всички налични данни (A) и редуцирана извадка (B), включваща 70% от началните данни. След трансформиране на данните са изградени параметрични ARIMA модели за сравнение, които показват съответствие на данните с до $R^2=58\%$ и $R^2=49\%$, съответно за двете извадки A и B. Впоследствие, като предиктори са идентифицирани четири лагирани променливи (съответно за лагове 1, 2, 3, 6), тренд от първи ред и една времева променлива (Month). Изградени са няколко Arcing модела, от които A50 е избраният модел за извадка A и B60 – за извадка B. За да се елиминира серийната корелация, грешките на тези модели са моделирани допълнително с едномерен ARIMA модел и са конструирани хибридни Arcing-ARIMA модели. Показателите им след ретрансформация за пълната извадка A са оценени на $R^2=98.5\%$ и MAPE=12.5%, а за редуцираната извадка B – съответно на $R^2=98.1\%$ и MAPE=16.22%. Разработени са Arcing-ARIMA хибридни модели с нетрансформирани данни. Избраният най-добър хибриден модел HA_AL120 за извадката с 224 данни има статистики $R^2=99.6\%$ и MAPE=12.9% и показва най-точни прогнози за трите тестови „бъдещи“ месеца. Накрая за сравнение са построени и модели с Random Forests с нетрансформирани данни. Чрез вариране на хиперпараметрите бяха идентифицирани три модела, от които за най-добър модел бе избран RF Model 2. Неговите резидууми не съдържат автокорелация и са с близко до нормалното разпределение, което не налага хибридизация. Статистиките на модела RF Model 2 са: $R^2=98.2\%$ и MAPE=17.7%. От трите типа моделиране в тази глава, с най-добри статистики и най-точни прогнози за тестовите данни е хибридният Arcing-ARIMA модел HA_AL120.

Заклучението представя кратко резюме, списък на публикациите по дисертационния труд, апробации и са систематизирани претенциите за научни приноси на докторанта.

Библиографията включва 135 източника.

Дисертационният труд е разработен към катедра “Математически анализ” на Факултета по математика и информатика на Пловдивския университет “Паисий Хилендарски” в периода 2020-2023 година.

КРАТКО СЪДЪРЖАНИЕ НА ДИСЕРТАЦИОННИЯ ТРУД

ГЛАВА 1. ВЪВЕДЕНИЕ

1.1. Основни понятия за временен ред

Изследването на валутните финансови пазари се свежда до обработка, анализ и моделиране на временни редове. Ще разгледаме задачата за моделиране с регресионен тип методи. Целта на регресионния тип модел е както да опише и предскаже наблюдаваните данни, така и да даде възможност за получаване на прогнози за следващите във времето моменти. Общият вид на временния ред $Y(t)$ е последователност от числа от вида

$$Y(t) = \{y_1, y_2, \dots, y_t, \dots, y_N\}, y_t \in \square \quad (1.1)$$

където t е времеви индекс, N е брой наблюдения (размер на извадката), \square е множеството на реалните числа. За построяване на регресионен модел, в едномерния случай се търси

зависимост между всяко наблюдение y_t и няколко негови предишни y_{t-1}, y_{t-2}, \dots . Този процес се нарича авторегресионен и е характерен само за динамични данни. При наличие на други, свързани с (1.1) редове в същата скала и времеви период задачата става многомерна. В реална ситуация, за регресионен тип модел, ще могат да се отчетат както авторегресионни компоненти на $Y(t)$, така и влиянията на другите редове с техни авторегресионни членове. Ще считаме, че са зададени r независими редове $X_{1,t}, X_{2,t}, \dots, X_{r,t}$ във вектор за всяко t $\mathbf{X}_t = (X_{1,t}, X_{2,t}, \dots, X_{r,t})$. Търсим явен вид на зависимостта (модел) на Y от \mathbf{X}_t от вида:

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

Грешката (резидиум, остатък) на модела $\hat{Y}(t)$ за всяко t означаваме с $\varepsilon(t) = Y(t) - f(\mathbf{X}_t)$. Обикновено се приема, че тази грешка е некорелирана с нулева средна стойност и постоянна дисперсия σ^2 (бял шум), т.е. $\varepsilon(t) \in N(0, \sigma^2)$.

Друго основно понятие за временен ред е понятието стационарност. Казваме, че редът е стационарен, ако неговата средна стойност и стандартно отклонение не се променят с течение на времето. В противен случай, той е нестационарен и има тренд. Тренд от първи ред се изчислява с числено диференциране по формулата $Trend(t) = Y(t) - Y(t-1)$ за всяко t .

1.2. Използвани методи за моделиране и изследване на валутни финансови пазари

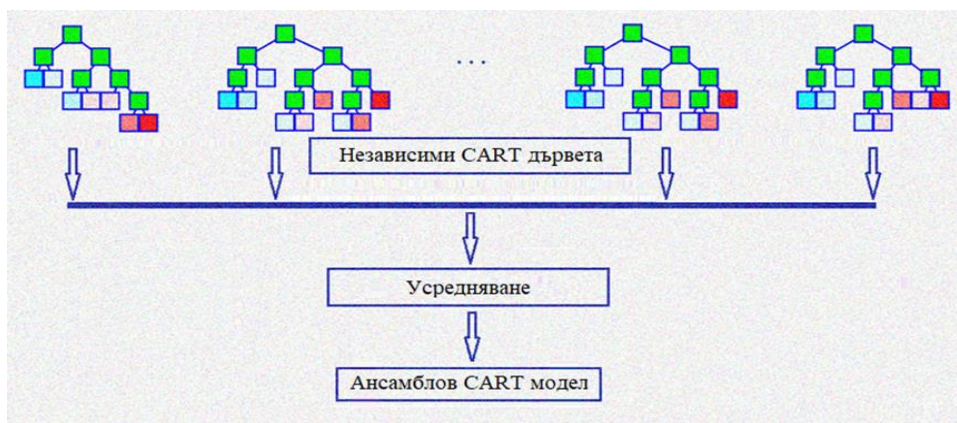
Като общо правило при всички проучвания в приложната тематика се забелязва, че за източник най-често се използват база-данните на централните банки или статистически институти в изследваните държави [22, 42], а за статистическата им обработка се прибегва до помощта на определени специализирани и автоматизирани софтуерни продукти и статистически софтуер [57, 103].

Метод на авторегресионна интегрирана с подвижна средна (Box Jenkins ARIMA)

Методологията за прогнозиране с Box-Jenkins използва итеративен подход за определяне на валиден линеен тип параметричен модел. Общият вид на модела задава с обща формула стойността на реда във всеки времеви момент t . Моделът се записва като ARIMA(p,d,q)_s, където параметрите са: p - брой предходни стойности за текущо t (лагирани членове), d - брой трендове, q - брой стохастични подвижни средни, s - сезонност. Избраният модел се сравнява с исторически данни за да се провери доколко точно описва временния ред. Моделът се счита за приемлив, ако остатъците му са малки, разпределени на случаен начин и не съдържат полезна информация. Ако даден модел не е задоволителен, процесът се повтаря с актуализирани параметри. Тази процедура се повтаря, докато се намери задоволителен модел. Полученият модел може да се използва за прогнозни цели.

Ансамблов метод на класификационните и регресионни дървета с багинг (CART Ensembles and Bagging)

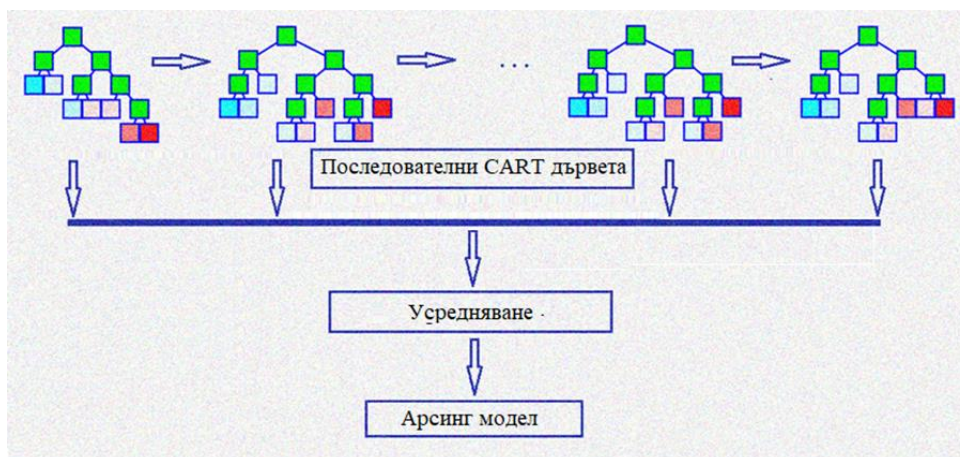
EBag е интелигентен метод с машинно обучение (МО) с CART дървета [19]. Той се получава с усредняване на предсказанията на множество модели, генерирани с еднакъв алгоритъм за различни преразпределени (бутстрап) извадки от началните данни. Обучава се стандартно с кросвалидация. EBag предоставя допълнителни функции за относителното тегло по средна стойност и медиана на предикторите в регресионен предсказващ модел, диференцирани тегла при многокласова класификация, запълване на липсващи стойности, резултатите от статистиките на всички отделни дървета в ансамбъла, сравнения на статистиките на всички дървета и др. (Фиг. 1.3). EBag е слабо изучен за данни от финансовата сфера.



Фигура 1.3. Идеята на ансамблов CART модел.

Ансамблов метод с адаптивен бустинг (Adaptively resampling and combine – Arcing)

Arcing е алгоритъм, базиран на CART разрешаващи дървета от класа на градиентното усилване [18]. Той комбинира последователно моделите, като всеки следващ отчита и подобрява качествата на предходните модели. Идеята на метода е илюстрирана на Фиг. 1.4.



Фигура 1.4. Arcing модел.

Arcing дава възможност за по-ефективно обучение на единичните модели, по-добро прогнозиране на тестови извадки и оттам получаване на алтернативен ансамблов модел. Предоставя информация както EBag. Досега този метод не е добре изследван за емпирични данни и не е бил използван в сферата на валутните финансови пазари.

Метод на случайните гори (Random Forests, RF)

Използването на RF е като основен „бенчмарк“ за прилагане на машинно обучение [20]. Той е един от най-предпочитаните методи за прогнозиране и многокласова класификация. Начинът за обучение и прогнозиране на нови данни с OOB предоставя голяма възможност за получаване на високоефективни модели и прогнози в сложната сфера на валутните финансови пазари. Резултатите от RF моделите ще се сравнят с останалите подходи.

Хибридизация на модели

Понякога се налага допълнително доуточняване на резултатите за получения модел. В случая на временни редове, за статистическата адекватност и възможност за приложение на моделите се изискват проверки, включени в диагностиката на грешките на модела или в други изисквания. За целта може да се прилага моделиране на грешките и извличане от тях на остатъчна зависимост. В настоящия дисертационен труд грешките (остатъците) се моделират допълнително с ARIMA модели. Хибридният модел се получава след сумиране на изходния модел с получения ARIMA модел.

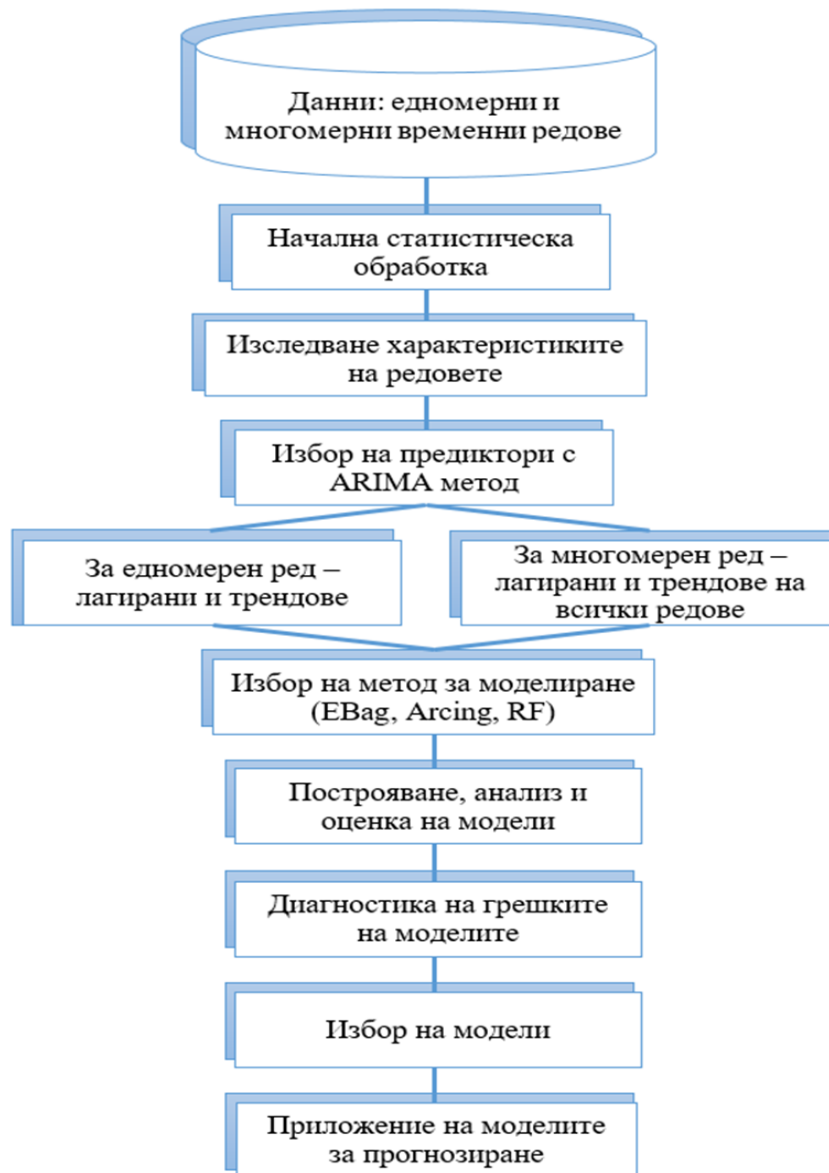
Теоретичните основи на ансамбловите методи са развити в [45, 66, 110, 119].

ГЛАВА 2. МОДЕЛИРАНЕ НА ДЕПОЗИТИ В ЧУЖДЕСТРАННА ВАЛУТА С АНСАМБЛОВИ МЕТОДИ

В тази глава е представена обща рамка на изследванията в дисертационния труд. Прилага се методът EBag за предсказване на месечни данни за депозити в чуждестранна валута.

2.1. Обща рамка на изследванията в дисертационния труд

За построяване и анализ на предсказващи регресионни модели предлагаме и следваме в дисертационния труд следната обща рамка, представена на Фиг. 2.1.



Фигура 2.1. Схема на разработената обща рамка на изследванията.

2.2. Описание на данните, начална обработка и построяване на референтен ARIMA модел за идентифициране на предиктори

Изследването се основава на официални данни за депозитите на български граждани в периода февруари 2004 г. - април 2021 г., общо $N=207$ записа. Данните са в милиони щатски долари (USD), усреднени по месеци. Техен източник е Българската народна банка [22].

Променливата на съответния едномерен динамичен ред се означава с *Depos*. Описателна статистика на началните стойности на депозитите е дадена в Таблица 2.2.

Таблица 2.2. Описателна статистика на зависимата променлива Depos

Променлива	Средна стойност	Медиана	Стд. отклонение	Дисперсия	Асиметрия	Стд. грешка на асиметрия	Ексцес	Стд. грешка на ексцес
<i>Depos, USD</i>	158.28	139.44	111.55	12444.39	1.009	0.169	0.983	0.337

Тези показатели предполагат, че разпределението на Depos не е близко до нормалното. Тестовите на Колмогоров-Смирнов и Шапиро-Уилк са значими със стойности Sig.=0.001 и Sig.=0.000. Заключаваме, че началната променлива Depos не е нормално разпределена. За стабилизиране на дисперсията и подобряване разпределението на Depos, прилагаме следната трансформация на Йео-Джонсън:

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

където Tr_Y е трансформираната променлива, а параметърът λ се намира от изследователя.

След трансформацията получаваме променливата Tr_Depos с нормално разпределение и използваме метода ARIMA за построяване на референтен модел. Моделът е:

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

Статистическите показатели на модел (2.2) са: $R^2 = 51.7\%$ и $RMSE=3.837$ млн. долара.

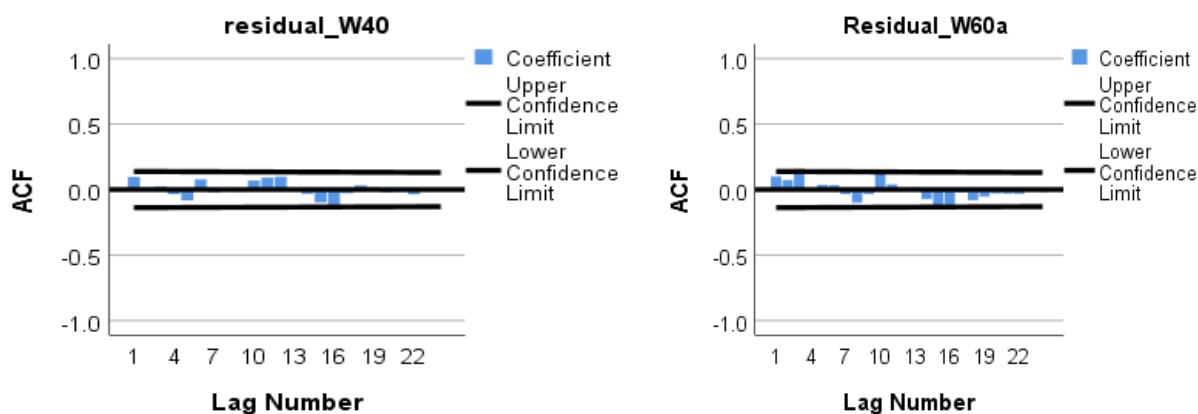
2.3. Приложение на ансамблов CART метод за данни за валутни депозити на български граждани

Построеният ARIMA модел (2.2) предполага да изберем предиктори от лагираните променливи с до $p=6$ закъснения (лагове назад от всеки текущ). За нашия случай ще вземем предвид и диференцирания времеви ред от първи ред, описващ наличието на тренд. Прилагаме метода EBag с различен брой дървета в ансамбъла и с различен брой лагирани променливи. За да се избегне евентуално преопределяне на моделите EBag, те се обучават чрез 10-кратно кръстосано валидиране. От всички получени модели за най-добри избираме моделите W40 и W60a (съответно с 40 и 60 дървета в ансамбъла). След ретрансформиране на прогнозираните стойности и връщане към началната променлива Depos се изчисляват прогнозираните стойности от моделите W40 и W60a. Получените променливи се означават съответно с Predicted_W40 и Predicted_W60a. Техните статистики, изчислени за всички данни (при $N=207$) показват лек спад. Те са дадени в Таблица 2.9.

Таблица 2.9. Статистика на модели W40 и W60a след ретрансформация на данните.

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

За да се анализира адекватността на избраните модели W40 и W60a бяха изследвани техните остатъци. Остатъците са в границите на теоретичната грешка $\pm 2/\sqrt{N}$ както може да се види от диаграмата на техния ACF на Фигура 2.13.



Фигура 2.13. Автокорелационна функция (ACF) на резидуумите на модели W40 и W60a.

Конструираният EBag модел W40 и W60a може да се счита за статистически валиден. Числовите резултати за последните три месеца, прогнозираните от тези модели са дадени в Таблица 2.12. Отчита се по-добро прогнозиране на W60a, особено за първите два месеца.

Таблица 2.12. Сравнение на прогнозни стойности от EBag моделите на Tr_Depos204 за тестовите три месеца.

Променлива	Февруари, 2021	Март, 2021	Април, 2021
Измерени Depos	30.562	48.952	16.083
ARIMA(6,1,1)	18.642	6.878	8.779
W40	18.184	19.587	19.392
W60a	27.711	36.409	35.323

2.4. Приложение на EBag метода за Depos204 без трансформиране на данните

Повтаряме изследването използвайки същите данни, описани в параграф 2.2. Тук отново сме отделили последните три данни (месеци февруари, март и април 2021 г.) за тестване. Както вече бе отбелязано в 2.3, тъй като данните не са нормално разпределени, не се препоръчва използването на класически параметрични методи. В нашия случай ще използваме EBag, който няма ограничения за разпределението на променливите. Изследването провеждаме без трансформация на изходните данни.

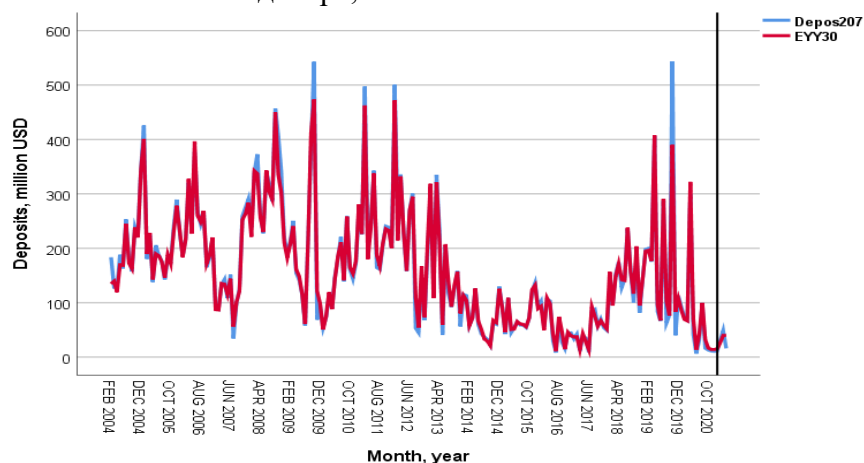
Следвайки параграф 2.3 възможни предиктори на Depos204 са неговите лагирани променливи Depos <1, 2, 3, 6> и Trend204. Бяха построени и оценени голям брой модели. За калибриране на моделите са вариран следните хиперпараметри: брой CART дървета T в ансамбъла от 40, 50, 60 и 70; тип на обучението – кросвалидация CV 5 и CV 10; отношение между минимум наблюдения в родителски към наследствен възел m1:m2 – 5:5, 10:5. В Таблица 2.14 са показани статистиките на избрани EBag модели. От първата група модели с m1:m2=10:5 най-добри показатели има модел EY30. Но той отстъпва на модел EYY30 от останалата група с m1:m2=5:5, който ще изберем за най-добър.

Таблица 2.14. Статистики на избрани EBag модели на Depos204 без трансформиране на данни.

EBag модел	Брой дървета	m1:m2	Кросвалидация CV	Предиктори	R ²	RMSE	MAPE
EY30	30	10:5	CV10	Depos204<1>, <6>, Depos204_d1	0.9516	25.623	0.163
EY50	50	10:5	CV10	Depos204<1>, <6>, Depos204_d1	0.9512	25.948	0.166

<i>EY60</i>	60	10:5	CV10	<i>Depos204</i> <1>, <6>, <i>Depos204_d1</i>	0.9513	25.971	0.164
<i>EY60a</i>	60	10:5	CV10	<i>Depos204</i> <1>, <3>, <6>, <i>Depos204_d1</i>	0.9560	25.074	0.155
<i>EY60b</i>	60	10:5	CV10	<i>Depos204</i> <1>, <2>, <3>, <6>, <i>Depos204_d1</i>	0.9559	25.178	0.164
<i>EYY30</i>	30	5:5	CV10	<i>Depos204</i> <1>, <6>, <i>Depos204_d1</i>	0.9765	18.218	0.102
<i>EYY30a</i>	30	5:5	CV5	<i>Depos204</i> <1>, <6>, <i>Depos204_d1</i>	0.9765	18.195	0.102
<i>EYY40</i>	40	5:5	CV10	<i>Depos204</i> <1>, <6>, <i>Depos204_d1</i>	0.9767	17.966	0.103
<i>EYY50</i>	50	5:5	CV10	<i>Depos204</i> <1>, <6>, <i>Depos204_d1</i>	0.9777	18.032	0.104
<i>EYY60</i>	60	5:5	CV10	<i>Depos204</i> <1>, <6>, <i>Depos204_d1</i>	0.9777	18.162	0.103

На Фигура 2.17 е дадена графика за сравнение на оригиналните данни с предсказаните стойности на EYY30. Наблюдава се много добро приближение. След преизчисляване на статистиките за N=207 данни, получаваме: $R^2=0.9766$ или 97.7% съвпадение с данните и грешки RMSE=18.174 млн. долара, MAPE=0.109 или 10.9%.

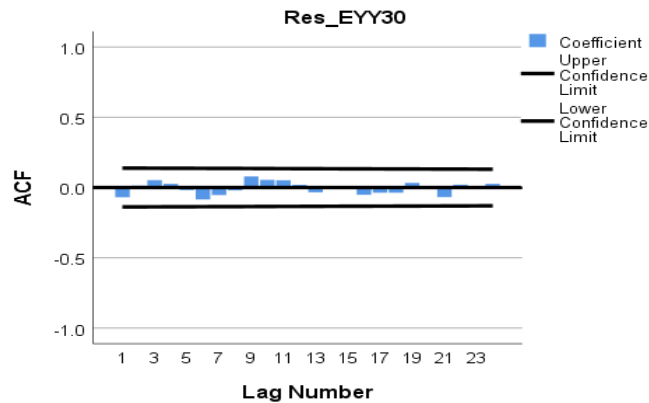


Фигура 2.17. Наблюдавани стойности на депозитния времеви ред *Depos* (синя линия) и прогнозираните от модел EYY30 (с червена линия).

За демонстрация на липсата на автокорелация в остатъците на избрания модел EYY30, на Фигура 2.20 представена автокорелационната функция на резидуумите (ACF). Вижда се, че всички ACF коефициенти на резидуумите до лаг 24 са в доверителните интервали.

Можем да заключим, че модел с EBag без трансформации на данните е адекватен и може да се прилага за предсказване на реда *Depos207*.

Прогнозните стойности за последните три месеца, неучаствали в моделирането, за модели EY30 и EYY30 са дадени на следващата Таблица 2.15. Може да се отчете доброто прогнозиране на повечето модели за първия и втория месец, докато за третия месец има големи неточности. Това се дължи на използването на малкия брой предиктори.



Фигура 2.20. ACF на остатъците на модел EYY30

Таблица 2.15. Прогнозни стойности от EBag моделите на Depos204 без трансформация на данни за тестовите три месеца.

Променлива	Февруари, 2021	Март, 2021	Април, 2021
Измерени Depos	30.562	48.952	16.083
EY30	31.284	39.643	39.643
EYY30	26.811	40.000	40.000

ГЛАВА 3. МОДЕЛИРАНЕ НА ВАЛУТНИЯ КУРС EUR/USD

В тази глава, за да постигнем възможно най-точна прогноза за развитието на валутния курс EUR/USD, ще се фокусираме върху зависимостта на валутната търговия, като вземем предвид влиянието на ключовите макроикономически показатели.

Моделирането се провежда на два основни етапа, а именно избор на предиктори и изграждане на модели по общата рамка. В първия етап използваме анализ на временните редове и едномерен ARIMA подход на Бокс-Дженкинс [15, 87, 48], като и статистически тестове. На втори етап, след като сме определили съответните предикторни временни редове от макроикономическите фактори за моделиране и прогнозиране на валутния курс EUR/USD прилагаме и сравняваме резултатите от три различни ансамблови методи с MO: CART-Ensembles and bagging (EBag), Arcing и Random Forests (RF).

3.1. Използвани данни за анализ и прогнозиране на валутния курс EUR/USD и начална обработка

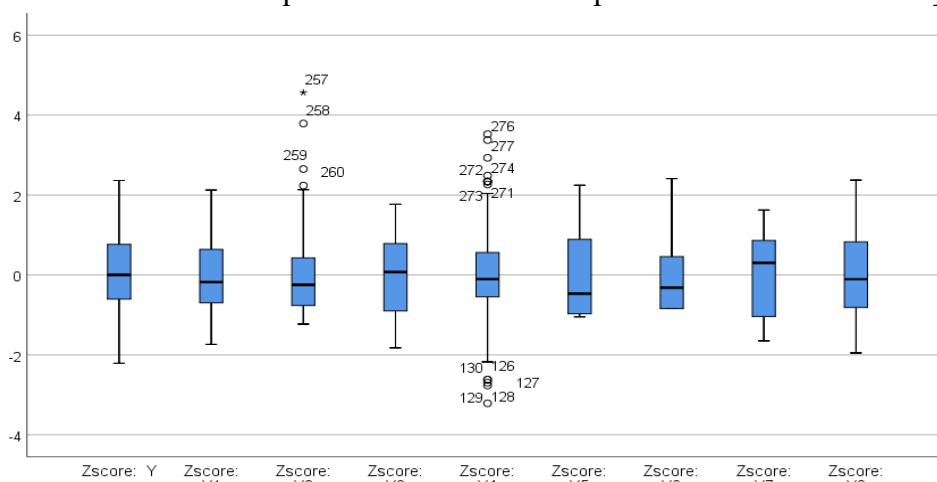
В тази глава са използвани официални данни, предоставени от професионалния терминал на Bloomberg. Данните са подредени по месеци, като динамичният ред обхваща периода декември 1998 г. - декември 2021 г. или общо $N=277$ случая. Променливите и техните описателни статистики са дадени в Таблица 3.1.

Таблица 3.1. Описателна статистика на използваните данни за моделиране на валутния курс EUR/USD.

Променлива	Описание	Мерна единица	Средна ст.	Медиана	Стд. откл.
Y	EUR/USD_FX_RATE	Отн.ед.	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	индекс	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

Величината на валутния курс EUR/USD, означена с $Y=EUR/USD_FX_RATE$ се счита за зависима променлива. Спецификата на данните от динамичните редове в Таблица 3.1 е, че те се измерват едновременно, което затруднява прогнозирането в реално време за бъдещите им стойности. На Фигура 3.2 са показани бокс-графиките на стандартизираните променливи на Y . От тях може да се заключи, че разпределенията на всички променливи са близки до нормалното. Най-ясно това е изразено за зависимата променлива $Y=EUR/USD_FX_RATE$.



Фигура 3.2. Бокс-плот на разпределенията на временните редове, използвани за моделиране на валутния курс.

След проверка на необходимото предположение за нормалност на първоначалните независими времеви редове се изследваха техните автокорелационни функции и се извърши моделиране с едномерен ARIMA анализ поотделно за всеки от тях. Част от построените модели са показани в Таблица 3.3.

Таблица 3.3. Статистика на получените едномерни ARIMA модели за избор на предиктори за моделиране на валутния курс EUR/USD

Променлива	Описание	ARIMA (p,d,q) ₁₂	Значимост на теста на Ljung-Box	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

Моделите имат отлични предсказващи качества с голям коефициент на детерминация, показващ съвпадение с моделираните данни от 87% за USA_UNEMPLOYMENT до 100% за EUROZONE_INFLATION. Най-важната информация за нас от Таблица 3.3 е, че във всички модели е идентифициран тренд от първи ред, т.е. $d=1$ за всички променливи и нулев авторегресионен (AR) член (т.е. $p=0$).

3.2. Приложение на CART Ensembles and Bagging за моделиране на валутния курс

За да изградим EBag модели, използваме лагираните променливи на X_1, X_2, \dots, X_8 , разделени на четири предикторни групи според вида на тяхната информация, а именно:

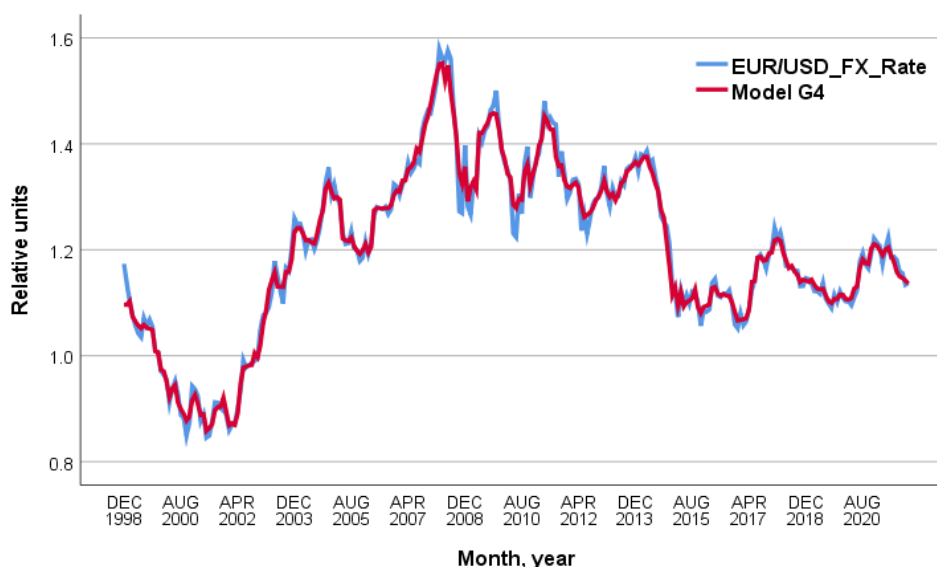
- Данни с индексите на безработицата в ЕС и САЩ;
- Променливите с данните за инфлацията в ЕС и САЩ;
- Основните лихвени проценти на ЕЦБ и FED;
- Доходността на 10-годишните ценни книжа на ЕС и САЩ.

В процеса на моделиране бяха построени голям брой модели, чрез вариране на техните хиперпараметри, а моделите с най-добри качества се оказаха EBag моделите с 40 дървета, стандартна 10-кратна кросвалидация и минимум наблюдения в краен възел 5:5. Таблица 3.6 показва постигнатите резултати, където моделите са означени с G1, G2, G3 и G4.

Таблица 3.6. Статистика на избраните EBag модели за валутния курс

Модел	Предиктори	RMSE	MAPE	R ²
G1	LX_1, LX_2	0.0408	0.0222	0.9377
G2	LX_1, LX_2, LX_3, LX_4	0.0182	0.0108	0.9874
G3	$LX_1, LX_2, LX_3, LX_4, LX_5, LX_6$	0.0185	0.0108	0.9868
G4	$LX_1, LX_2, LX_3, LX_4, LX_5, LX_6, LX_7, LX_8$	0.0174	0.0103	0.9884

Фигура 3.3 илюстрира линейна графика на реда с оригиналните данни за EUR/USD_FX_RATE сравнени с тези, предсказани от най-добрия модел G4.



Фигура 3.3. Сравнение на оригиналните данни за $Y = \text{EUR/USD_FX_RATE}$ с предсказаните стойности от EBag модела G4.

3.3. Приложение на ArcSine за моделиране на данните за валутния курс

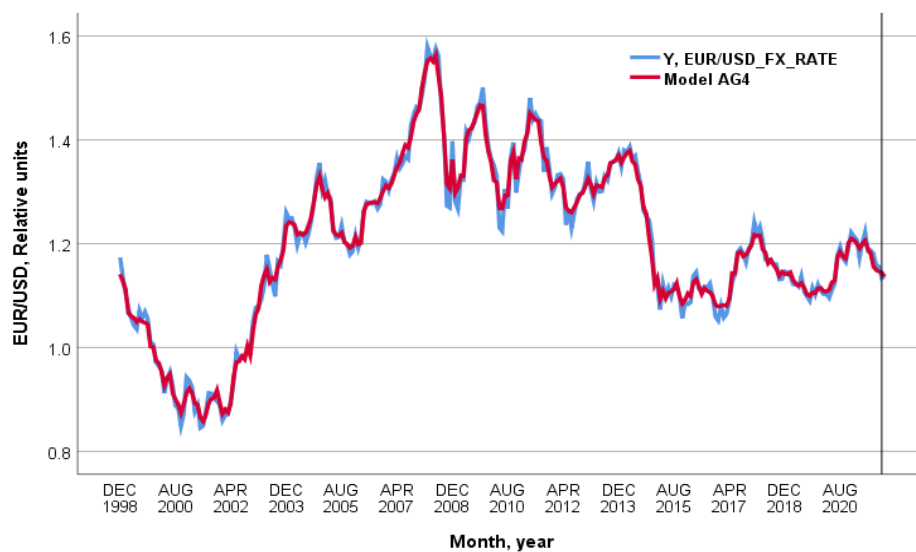
Повтаряме изследването като за предиктори използваме същите променливи, както за EBag моделите, разделени в четирите групи според вида на тяхната информация (виж параграф 3.2). Зависимата променлива е $Y = \text{EUR/USD_FX_RATE}$, от която последната стойност (за декември 2021 г.), е премахната и беше използвана за проверка.

В процеса на моделиране бяха построени голям брой модели, чрез избор и калибриране на основните хиперпараметри. Таблица 3.7 показва постигнатите резултати, където моделите са означени с AG1, AG2, AG3 и AG4.

Таблица 3.7. Статистики на получените Arcsine модели на валутния курс.

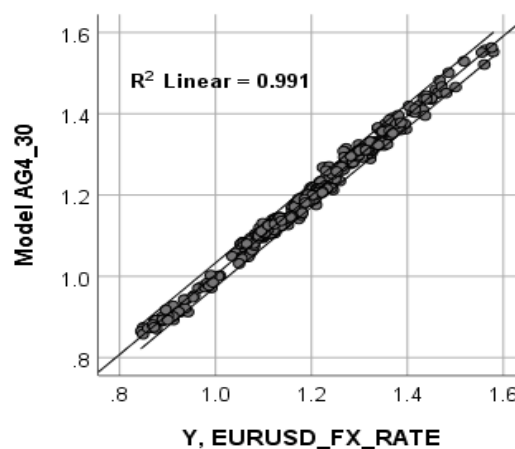
Модел	Предиктори	RMSE	MAPE	R ²
AG1	LX_1, LX_2	0.0349	0.0227	0.9580
AG2	LX_1, LX_2, LX_3, LX_4	0.0164	0.0107	0.9897
AG3	$LX_1, LX_2, LX_3, LX_4, LX_5, LX_6$	0.0160	0.0104	0.9901
AG4	$LX_1, LX_2, LX_3, LX_4, LX_5, LX_6, LX_7, LX_8$	0.0152	0.0100	0.9912

Фигура 3.9 показва графика на всички оригинални данни за EUR/USD_FX_RATE, сравнени с предсказаните стойности от най-добрия модел AG4. Наблюдава се много добро съвпадение, с изключение на някои екстремни стойности, което е характерно за ансамбловите методи, чиито модели се изчисляват с усредняване.



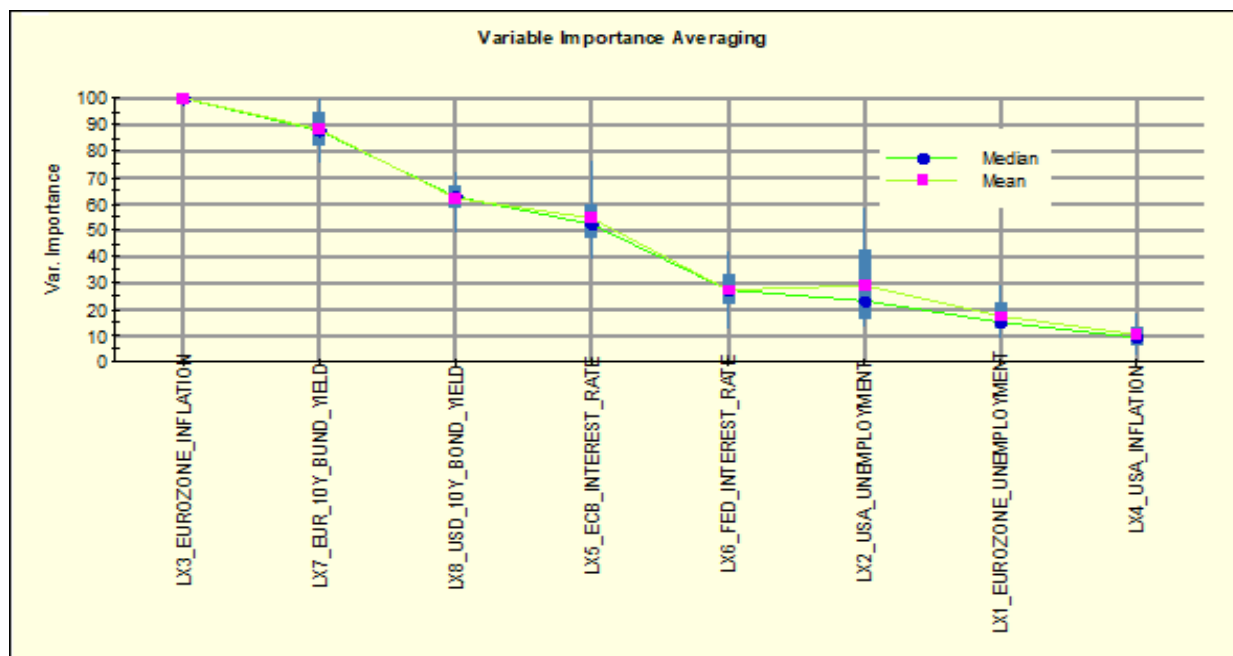
Фигура 3.9. Сравнение на оригиналните данни за $Y = \text{EUR/USD_FX_RATE}$ с предсказаните стойности от Arcsine модела AG4.

Качеството на приближение на оригиналните данни за EUR/USD_FX_RATE с избрания най-добър Arcsine модел AG4 е илюстрирано на Фигура 3.11 с 95% доверителни интервали.



Фигура 3.11. Точкова диаграма на EUR/USD_FX_RATE срещу предсказаните стойности от Arcsine модел AG4.

Важно преимущество на изследваните ансамблови модели е, че се извлича информация за тежката на използваните предиктори. За модел AG4 на Фигура 3.12 е дадена графиката със средните стойности и медианите на относителния принос на предикторите. Най-голямо относително влияние има инфлацията в ЕС (100), следвана от EUR_10Y доходност от бунд (88), USD_10Y доходност от бунд (63) и лихвен процент на ЕЦБ (57).



Фигура 3.12. Относителна тежест на предикторите в модел AG4.

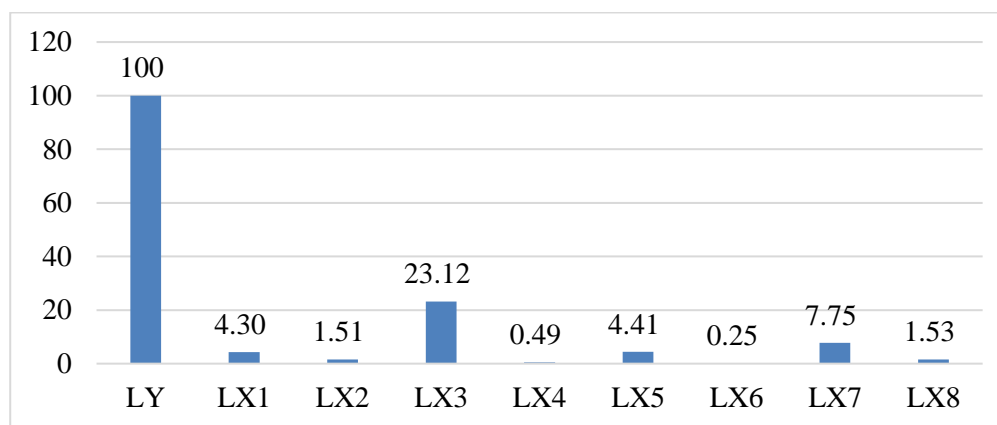
3.4. Моделиране на валутния курс с Random Forests

Повтаряме изследването като за предиктори използваме същите променливи, както за EBag и ArcSine моделите, следвайки общата рамка на изследванията. За RF моделиране към осемте предиктори се оказва подходящо и включването на първата лагирана променлива на зависимата $Y=EUR/USD_FX_RATE$. Т.е. тук предикторните променливи са девет. Бяха построени множество модели с вариране на хиперпараметрите на RF, като най-добри статистически резултати с RF модели се получиха при $T=350$ дървета, минимум наблюдения в краен възел $Atm=5$ и $mtry=3$. Избраните модели са означени с RG1, RG2, RG3 и RG4, а съответните им статистики са дадени Таблица 3.9. От Тези таблици се вижда, че последните три модела са с много близки мерки, като най-добър е модел RG4.

Таблица 3.9. Статистики на избраните RF модели на валутния курс

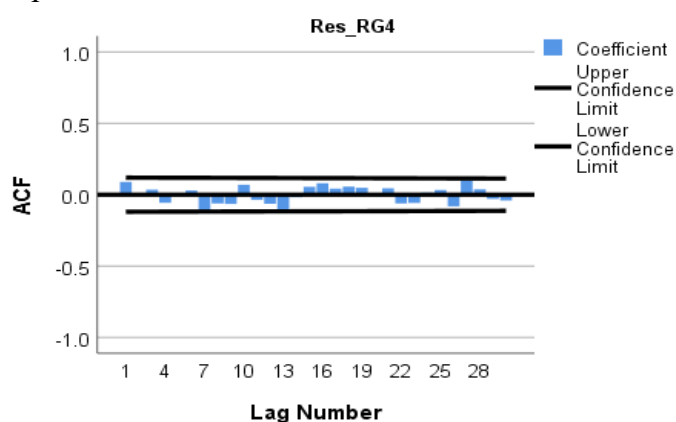
Модел	Предиктори	RMSE	MAPE	R ²
RG1	LY, LX_1, LX_2	0.0208	0.0131	0.9837
RG2	$LY, LX_1, LX_2, LX_3, LX_4$	0.0196	0.0120	0.9858
RG3	$LY, LX_1, LX_2, LX_3, LX_4, LX_5, LX_6$	0.0194	0.0121	0.9861
RG4	$LY, LX_1, LX_2, LX_3, LX_4, LX_5, LX_6, LX_7, LX_8$	0.0194	0.0120	0.9862

Относителното влияние на отделните предиктори в модел RG4 е визуализирано на Фигура 3.17.



Фигура 3.17. Влияние на предикторите в RF модел RG4.

Изследването за автокорелация на резидуумите на модел RG4 с Фигура 3.18 показва, че те не съдържат автокорелация.



Фигура 3.18. Автокорелационна функция на остатъците на RF модел RG4

За последната тестова стойност за последния месец за курса EUR/USD, равна на 1.1370 прогнозите на RF моделите са както следва:

1.12445 (RG1), 1.12640 (RG2), 1.13115 (RG3), 1.13674 (RG4). Прогнозираната стойност на модел RG4 е много точна, с абсолютна грешка, по-малка от 0.0003.

ГЛАВА 4. МОДЕЛИРАНЕ НА БАНКОВИ ДЕПОЗИТИ С ХИБРИДНИ МЕТОДИ

В тази глава са получени резултати за моделиране на едномерни временни редове с хибриден подход, базиран на интелигентни методи с МО и едномерен сезонен ARIMA. Методът Arcsing с последваща хибридизация не е използван досега от други автори в сферата на финансовите пазари. Тъй като в научната литература по този проблем все още не е намерено официално такова приложение, ние приехме предизвикателството да демонстрираме възможностите на този подход.

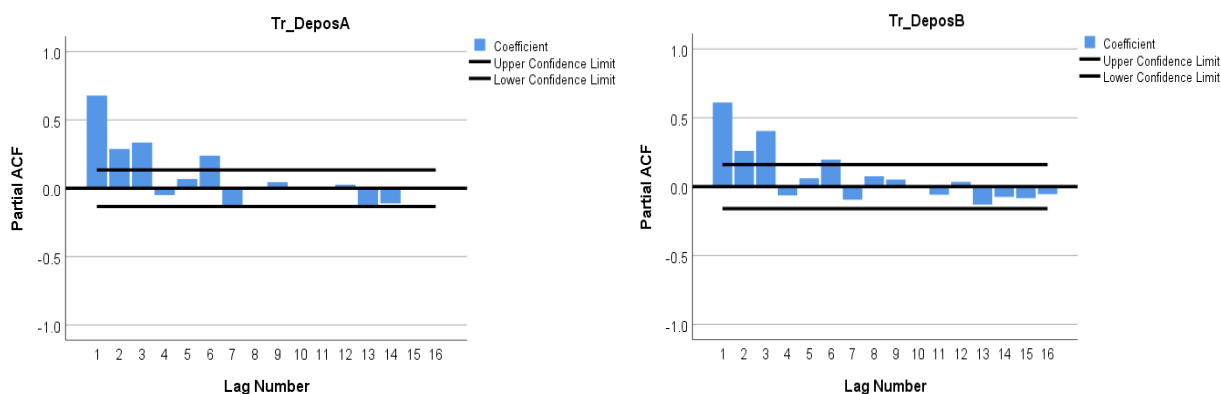
4.1. Описание на проблема и данните

В изследването се моделира едномерен времеви ред с данни от банкови депозити, с използване на разширена извадка, спрямо данните от Глава 2. Прилага се разработената обща рамка, използвана в Глави 2 и 3 с идентифициране на лагирани променливи и трендове за конструиране на предиктори. В изследването използваме официално публикуваните данни на БНБ [22] за банковите депозити в милиони щатски долари на физически лица на

средномесечна база. Данните обхващат периода от февруари 2004 г. до октомври 2022 г., (N=227 записа) и са допълнен вариант на данните от Глава 2, с нови 20 месеца.

4.2. Построяване и анализ на референтни едномерни ARIMA модели на Tr_DeposA и Tr_DeposB, и идентифициране на предиктори

За предварително определяне на възможните стойности на параметрите изследваме ACF и PACF на съответните редове. Съответните графики са дадени на Фигура 4.6. Плавното намаляване на ACF коефициентите с времето е характерно за времеви редове без цикли и сезонност в данните. PACF показват по няколко стойности извън доверителните интервали. Това е индикатор за определяне на параметрите на съответния ARIMA модел. В нашия случай p, q са в рамките от 1 до 6 и за двата реда Tr_DeposA, Tr_DeposB, с по-ясно изразен лаг 6 за първия ред.



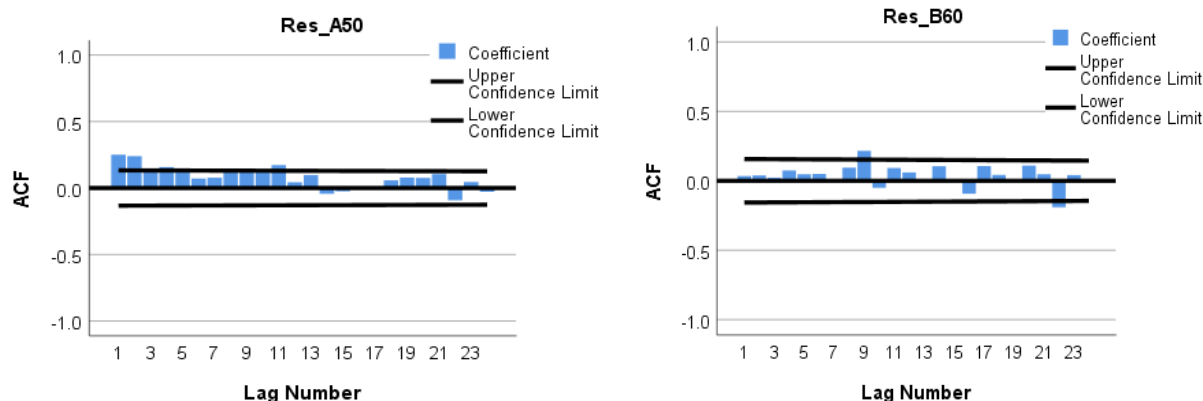
Построени и изследвани бяха различни ARIMA модели. От тях бяха избрани модели за Tr_DeposA: (4.4) ARIMA(2,1,3) и за Tr_DeposB: (4.5) ARIMA(2,1,0). Статистическите оценки на модел (4.4) са $R^2=0.578$, MAPE=29.62%, а за модел (4.5) $R^2=0.492$, MAPE=29.59%. Получените статистически оценки и за двата модела (4.4) и (4.5) не са задоволителни. Въпреки това, получените ARIMA модели дават възможност за някакъв първоначален избор на предиктори за провеждане на регресия с по-мощни методи.

4.3. Хибриден Arcing -ARIMA метод за прогнозиране на банковите депозити

В предния параграф за променливите Tr_DeposA и Tr_DeposB идентифицирахме възможни предиктори от лагираните им променливи – до p, q равни на 6. Също така, наличието на тренд (параметър d) в моделите (4.4), (4.5) ни позволява да използваме данните с трендове за предиктори. Допълнително, за да отразим, че данните зависят от времето, ние можем да включим и номинална променлива за месец за отчитане на времето.

Прилагаме метода Arcing с различен брой дървета в ансамбъла. Обучението на моделите беше извършено с помощта на стандартна 10-кратна CV. Съотношението на минималния брой случаи в родителски възел ($m1$) към минималния брой случаи в дъщерен възел е $m1=5, m2=5$. Броят на дърветата, посочени в моделите, е означен с T . Таблица 4.10 представя основните индикатори на избрани Arcing модели за прогнозиране на времевите редове Tr_DeposA и Tr_DeposB. Моделите на Tr_DeposA са означени с A, тези на Tr_DeposB съответно с B. За най-добри модели определяме A50 и B60.

За да се анализира адекватността на избраните модели A50 и B60 бяха изследвани техните остатъци. Остатъците не са в рамките на очакваните граници, както може да се установи от графиките на техните ACF на Фигура 4.11. По-съществени отклонения се наблюдават за първия модел.



Фигура 4.11. Графики на ACF на остатъците на модели A50 и B60.

Поради тази причина ние допълнително ще моделираме остатъците на A50 и B60 като използваме едномерен ARIMA метод. Получените модели са означени съответно с ResA50_ARIMA(1,0,9) и ResB60_ARIMA(0,0,22). Накрая добавяме предсказаните стойности на Res_ARIMA моделите към Arcing моделите A50 и B60 по (1.26) и получаваме хибридни модели, означени като hybrid Arcing-ARIMA (HA50, HB60). Статистиките на сравненията на тези хибридни модели след ретрансформация към началните мерни единици означаваме с retr_HA50 и retr_HB60. Сравнението на тези модели с всички начални данни Depos227 и Depos160, съответно, са дадени в Таблица 4.12.

Таблица 4.12. Сравнение на приближението на ретрансформирани хибридни Arcing-ARIMA модели с пълните извадки Depos227 и Depos160.

Модел	RMSE	MAPE	R^2
<i>retr_HA50</i>	15.9538	12.5%	0.9854
<i>retr_HB60</i>	18.8841	16.22%	0.9813

Моделите бяха приложени за прогнозиране на извадка от Depos три месеца напред. Резултатите са както следва: точни стойности: (80, 79, 170) млн. долара, изчислени от retr_HA50 (123, 76.5, 104). За намалената извадка retr_HB60 прогнозите са съответно (114, 74, 129). Тези резултати не са удовлетворителни, независимо от добрите статистически показатели на моделите.

4.4. Хибридно Arcing-ARIMA моделиране на депозити без трансформация на данните

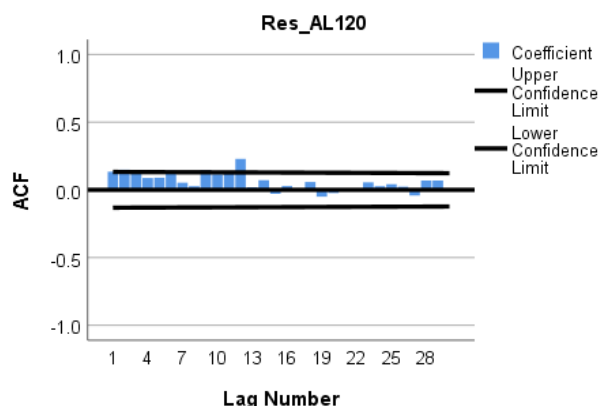
Тъй като методите с МО не зависят от типа на разпределението, в този параграф ще построим и изследваме Arcing модели без трансформация на данните. Целта е чрез калибриране на моделите да получим по-точни прогнози. Използва се същият подход с тренд и лагирани променливи като предиктори. За обучение на моделите хиперпараметрите са варираны с брой дървета в ансамбъл $T=30, 40, \dots, 150$; кросвалидация CV10 и CV5; $m1:m2=5:5$, и $10:5$.

Най-добри резултати се получиха с три предиктора, CV10-кратна и $m1:m2=5:5$. Основните статистики на избрани модели са дадени в Таблица 4.13. Наблюдава се плавно подобрене на статистическите показатели. Избрахме като най-добър модел AL120, тъй като за по-голям брой дървета R^2 и другите статистики се подобряват много слабо, а прогнозите за тестовите данни се влошават.

За проверка на адекватността на модела AL120 построяваме графиката на ACF на резидуумите му на Фигура 4.19. Стойности на ACF са критични, а някои от тях са извън доверителните интервали.

Таблица 4.13. Статистика на Arcing модели на *Depos224* без трансформация на данни.

ARC модел	Брой дървета T	Предиктори	R^2	RMSE	MAPE
AL30	30	<i>Depos224</i> <1>, <2>, <i>Trend224</i>	0.9939	13.662	0.355
AL50	50	<i>Depos224</i> <1>, <2>, <i>Trend224</i>	0.9957	10.328	0.246
AL70	70	<i>Depos224</i> <1>, <2>, <i>Trend224</i>	0.9963	9.118	0.234
AL100	100	<i>Depos224</i> <1>, <2>, <i>Trend224</i>	0.9968	8.208	0.212
AL120	120	<i>Depos224</i> <1>, <2>, <i>Trend224</i>	0.9970	7.865	0.202
AL150	150	<i>Depos224</i> <1>, <2>, <i>Trend224</i>	0.9970	7.567	0.1931



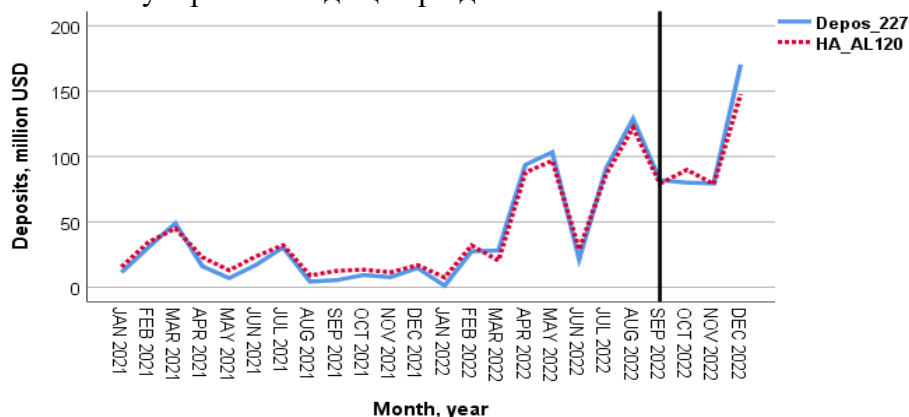
Фигура 4.19. ACF на остатъците на модел AL120.

За да отстраним този недостатък построяваме хибриден модел, като моделираме остатъците с ARIMA. Редът с резидуумите е означен с Res_AL120. За него построихме модел ARIMA(1,0,1). Хибридният модел означаваме с HA_AL120. Резидуумите на този хибриден модел вече не съдържат автокорелация, а статистиките на хибридният модел HA_AL120, сравнени с цялата извадка Depos227 са дадени на Таблица 4.16.

Таблица 4.16. Статистики на хибридният модел HA_AL120.

Модел	R^2	RMSE	MAPE	DW
HA_AL120	0.9963	7.377	0.129	1.936

Точността на предсказването с хибридният Arcing-ARIMA модел HA_AL120 за последните 24 месеца е показана на Фигура 4.21. Отдясно на вертикалната линия са прогнозите за тестовите последни три месеца. Отчита се много добро качество на прогнозирането за симулираните бъдещи три данни.



Фигура 4.21. Прогнози на хибридният Arcing-ARIMA модел HA_AL120 за тестови данни.

4.5. Моделиране на банкови депозити с метода Random Forests без трансформация на данните

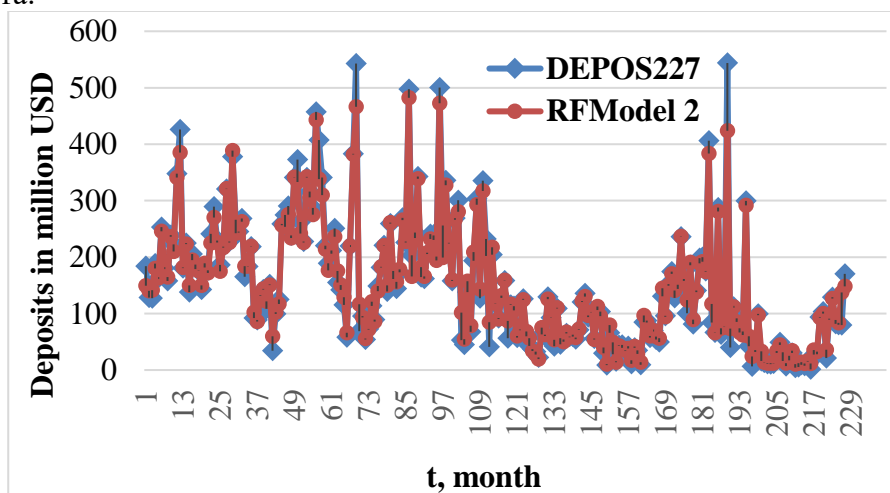
Прилагаме метода Random Forests (RF) за моделиране и прогнозиране на едномерен ред за данните с банкови депозити в чуждестранна валута. Използват се данните, описани в параграф 4.1. Всичките налични данни за 227 месеца са означени с променливата Depos227. Тъй като данните не са нормално разпределени, не се препоръчва използването на класически параметрични методи като множествена линейна регресия, ARIMA и други. В нашия случай ще използваме RF, който няма ограничения за разпределението на променливите. Изследването провеждаме без трансформация на изходните данни. Следвайки резултатите и заключенията от параграф 4.2 ще използваме следните времеви редове като предиктори: тренд от първи ред и лагираните променливи Depos224<1>, Depos224<2>, ..., Depos224<6>. Хиперпараметрите на RF са: брой дървета (Mtree), минимални наблюдения в терминален възел на дървото (Atm, атоми), брой предиктори, произволно избрани от общия брой предиктори (Mtry) отделно за всяко разделяне на всеки възел на всяко RF дърво. Бяха изградени множество модели с вариране на предикторите. Хиперпараметрите бяха зададени Mtree = 500, 700, и 1000; Atm = 2, 5, 10, и 20; Mtry=3, а моделите бяха обучени със стандартната OOB тестова извадка.

Най-добрите резултати от обучението и прогнозирането са получени със следните хиперпараметри: Mtree = 1000, Atm = 5, Mtry = 3. Окончателните статистически показатели за три най-добри предсказващи RF модела на Depos224, са дадени в Таблица 4.20. Най-малки грешки и най-висок $R^2=98,2\%$ са постигнати с RFModel 2. Той се генерира с променливите Trend и първите две лагирани променливи Depos224<1>, <2>. След него е RFModel 3, в който участва и Depos224<6>.

Таблица 4.20. Статистика на най-добрите предсказващи RF модели, сравнени с пълната извадка Depos227.

Модел	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%

Фигура 4.23 показва приближението на RFModel 2 за целия времеви ред на Depos227. Наблюдава се много добро съвпадение, с изключение на някои пикови стойности. Това явление е характерно за ансамбловите методи, поради усредняването на прогнозите на много дървета.



Фигура 4.23. Сравнение на началните времеви редове на депозитни данни с прогнози от RFModel 2.

Таблица 4.21. Относително тегло на променливите в моделите с RF и прогнозни стойности на най-добрите RF модели.

Модел	Trend	Depos2 24<1>	<2>	<3>	<6>	Прогнози за последните три месеца
<i>RFModel 1</i>	100	71.3	4.7	5.8	2.6	(85.4, 124.5, 118.7)
<i>RFModel 2</i>	100	92.6	3.7	-	-	(83.6, 137.6, 148.3)
<i>RFModel 3</i>	100	81.8	6.0	-	2.5	(84.5, 121.8, 134.1)

Проверката на адекватността на модела показва, че всички ACF стойности на резидуумите са в доверителните граници, следователно можем да заключим, че избраният най-добър RF модел *RFModel 2* е адекватен и получените прогнози в последната колона на Таблица 4.20 могат да бъдат използвани за интерпретация.

ЗАКЛЮЧЕНИЕ

Главният извод, който можем да направим е, че целта на настоящия дисертационен труд и поставените задачи са постигнати. Основната част от получените резултати са публикувана в 4 научни публикации. Изнесени са 4 доклада на научни форуми и семинари.

Претенции за научни и научно-приложни приноси

Получените научни и научно-приложни резултати на дисертационния труд могат да се систематизират както следва:

1. Разработена и приложена е обща рамка на изследванията за статистическо моделиране на едномерни и многомерни временни редове в областта на валутните и финансови пазари. Рамката включва подход за идентифициране на предиктори и приложение на слабоизучени интелигентни методи с машинно обучение за предсказване и краткосрочно прогнозиране.

2. Построени и анализирани са ефективни предсказващи модели на едномерни временни редове за депозитите в чуждестранна валута на българските граждани. За целта за първи път в тази област е приложен и изследван ансамбловият метод CART Ensembles and Bagging (EBag). Установено е, че моделите без трансформация на данните са по-точни и достигат съвпадение с реалните стойности до 97.7%. Моделите са приложени за получаване на прогнози за депозити от 3 месеца напред и показват съществено по-добри резултати от референтния ARIMA модел.

3. Построени и анализирани са високоефективни предсказващи модели на многомерни временни редове за валутния курс EUR/USD в зависимост от макроикономически фактори с три метода - EBag, Arcing и Random Forests. За първи път в областта на финансовите и валутни пазари е приложен и изследван ансамбловият метод Arcing от класа на градиентното усилване. Определено е влиянието на отделните макроикономически фактори върху валутния курс. Получените резултати от моделирането показват преимущество по точност на Arcing моделите със средна абсолютна процентна грешка (MAPE), равна на 1% и достигащи съвпадение с данните до 99.1%. Прогнозите с един месец напред и за трите метода са сравними с точността на наблюденията.

4. Развит и приложен е подход за хибридно Arcing-ARIMA моделиране на данни за депозитите на български граждани с трансформирани и нетрансформирани данни. Установено е, че получените хибридни модели с нетрансформирани данни са по-точни и достигат статистически показатели MAPE=12.9%, съвпадение с данните до 99.6%, и най-точни прогнози за трите тестови „бъдещи“ месеца, невключени в моделирането. Демонстрирано е, че тези резултати превъзхождат предсказващите възможности на стандартните ARIMA и Random Forests методи.

В следващата таблица е показана връзката между основните получени резултати, поставената цел и задачи, приносите и публикациите на дисертационния труд.

Таблица. Връзка на основните резултати с целите и публикациите.

Принос	Цел	Задача	Параграф	Публикации
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]

Апробация на научните и научно-приложни резултати

А) Участия в проекти

Част от резултатите са апробирани и използвани в следните научни проекти:

1. МУ21-ФМИ-015 (2021-2022 г.): „Високоэффективни предсказващи алгоритми с машинно обучение за моделиране на данни от екологията, индустрията и образованието” с ръководител доц. д-р Христина Кулина, финансиран от Фонд „Научни изследвания“ при ПУ „Паисий Хилендарски“ (приключил).

2. КП-06-Н52/9 (2021-2024 г.): „Изследване на математически и статистически методи с машинно обучение за интелигентна обработка на информация“ с ръководител доц. д-р Христина Кулина, финансиран от Фонд Научни изследвания към МОН (действащ).

Б) Доклади , изнесени на научни форуми

Част от резултатите, получени в дисертационния труд са докладвани на следните национални и международни конференции, научни сесии и семинари:

- 13th International Hybrid Conference for Promoting the Application of Mathematics in Technical and Natural Sciences (AMiTaNS'21), 24–29 June 2021, Albena, Bulgaria.
- Доклад на тема: „Статистически модели на спестяванията на граждани в чуждестранна валута“, семинар на катедра Математически анализ по Математическо моделиране и приложения, 26.06.2021 г. <https://www.fmi-plovdiv.org/GetResource?id=3908>
- Доклад на тема: „Ансамблови методи за моделиране и предсказване на валутния курс EUR/USD“, семинар на катедра Математически анализ по Математическо моделиране и приложения, 26-27 ноември 2022 г., гр. Хисар, <https://www.fmi-plovdiv.org/GetResource?id=4372>
- Доклад на тема: „Използване на хибридни методи за прогнозиране на банковите депозити на физически лица“, семинар на катедра Математически анализ по Математическо моделиране и приложения, 12.05.2023 г., <https://www.fmi-plovdiv.org/GetResource?id=4480>
- ACM-2023 6th International Conference on Mathematics and Statistics (ICoMS 2023), July 14-16, 2023, Leipzig, Germany.

Перспективи за бъдеща работа

По-нататъшни изследвания и публикации в областта на тематиката на настоящия дисертационен труд могат да бъдат насочени в следните направления:

1. По-нататъшно развитие на рамката за идентифициране и избор на предиктори, базирани на класически и интелигентни методи за обработка на данни от валутните пазари и финансовата сфера.
2. Приложение на други видове методи с машинно обучение като невронни мрежи и др. с цел прогнозиране на временни редове в реално време.
3. По-нататъшно развитие и получаване на резултати с хибридно моделиране.
4. Прилагане на разработения подход за изследване и краткосрочно прогнозиране на други валутни курсове.
5. Публикуване на дисертационния труд в разширен вариант като книга.

Списък на публикациите по дисертационния труд

- [P1] Boyouklev, 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] Boyouklev, 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] Boyouklev, 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] Boyouklev, 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**

Д Е К Л А Р А Ц И Я за оригиналност

по чл. 27, ал. 2 от ППЗРАСРБ

от Ивайло Владимиров Боюклиев

редовен докторант към катедра „Математически анализ”
на Факултет по математика и информатика
при Пловдивския университет „Паисий Хилендарски“

Във връзка с провеждането на процедура за придобиване на образователната и научна степен „Доктор” в Пловдивския университет „Паисий Хилендарски” и защита на представения от мен дисертационен труд,

декларирам:

Резултатите и приносите на проведеното дисертационно изследване, представени в дисертационния ми труд на тема **„Моделиране и изследване на валутни финансови пазари“** са оригинални.

12.01.2024

гр. Пловдив

ДЕКЛАРАТОР :.....

/Ивайло Боюклиев/

Благодарности

Издавам сърдечна благодарност и признателност към научните си ръководители проф. д.н. Снежана Георгиева Гочева-Илиева и доц. д-р Христина Николова Кулина за техните полезни професионални съвети, препоръки и добронамерено отношение по време на докторантурата и подготовката на дисертационния ми труд.

Благодаря на преподавателите от обучаващата катедра по Математически анализ и на ръководството на Факултета по математика и информатика на Пловдивския университет „Паисий Хилендарски“, за съдействието, съпричастността и административната подкрепа при обучението ми във ФМИ.

Благодарност издавам на търговските представители на „Bloomberg“ за България, както и на всички колеги, които по един или друг начин са били съпричастни към работата ми и обучението ми като докторант.

Библиография

- [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 classifiers. *The North Amer. J. Econ. Finance*, 47, 552–567, 2019.
- [10] Batmaz İ, Danişoğlu 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, Bencharaf 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, Porrás 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 Access*, 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.