



БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ

ИНСТИТУТ ПО РОБОТИКА



Магистър. инж. Екатерина Попова

**МАТЕМАТИЧЕСКИ МЕТОДИ ЗА ИЗСЛЕДВАНЕ, МОДЕЛИРАНЕ, АНАЛИЗ И
ПРОГНОЗИРАНЕ В ЕНЕРГЕТИКАТА И ЕНЕРГИЙНИТЕ ПАЗАРИ**

АВТОРЕФЕРАТ

**на дисертационен труд за придобиване на образователната и научна степен
"Доктор"**

Научна област: Технически науки

Професионално направление:

5.2 Електротехника, електроника и автоматика

Научна специалност:

Приложение на принципите и методите на кибернетиката в различни области на науката (02.21.10)

София, 2023 г.



БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ

ИНСТИТУТ ПО РОБОТИКА



Магистър. инж. Екатерина Поповска

**МАТЕМАТИЧЕСКИ МЕТОДИ ЗА ИЗСЛЕДВАНЕ, МОДЕЛИРАНЕ, АНАЛИЗ И
ПРОГНОЗИРАНЕ В ЕНЕРГЕТИКАТА И ЕНЕРГИЙНИТЕ ПАЗАРИ**

АВТОРЕФЕРАТ

**на дисертационен труд
за придобиване на образователната и научна степен
"Доктор"**

***Научна област: Технически науки
Професионално направление:
5.2 Електротехника, електроника и автоматика***

***Научна специалност:
Приложение на принципите и методите на кибернетиката в различни области на
науката (02.21.10)***

**Научен консултант: Доц. Д-р Галя Георгиева - Цанева
София, 2023 г.**

Дисертантът е докторант в Института по Роботика при Българската Академия на Науките - София.

Дисертационният труд се състои от увод, 4 глави, заключение, приноси и списък на използваната литература. Библиографската справка съдържа 139 заглавия, от които всички на латиница и 2 адреса в интернет. Обемът на дисертацията е 183 страници. Изложението съдържа 9 таблици и 37 фигури, чиито номера в автореферата съвпадат с тези в дисертацията. Номерацията на точките и подточките също съответства на тази от дисертационния труд. Приложенията са оформени на 22 страници, включващи списък с забелязани цитирания и кодове на софтуерни програми.

Дисертационният труд е обсъден и насрочен за защита съгласно изискванията на ЗРАС в РБ и Заповед № 110/_07.09.2023_г. на Директора на ИР-БАН на заседание на Разширен секционен семинар на учените от секция МР, проведено на 14.09.2023 г.

Защитата на дисертационния труд ще се състои на _____ г. от _____ часа в зала _____ на ИР-БАН, ул. „Акад. Георги Бончев“, блок 2.
Материалите по защитата са на разположение на интересувашите се в канцеларията на ИР-БАН, на адрес: ул. „Акад.Георги Бончев “, блок 2, етаж 4.

Автор: маг. инж. Екатерина Поповска

Заглавие: Математически методи за изследване, моделиране, анализ и прогнозиране в енергетиката и енергийните пазари

Тираж: 10 броя

Излиза от печат на: ноември 2023 г.

Печатна база: АВАНГАРД ПРИМА, гр. София

ОБЩА ХАРАКТЕРИСТИКА НА ДИСЕРТАЦИОННИЯ ТРУД

Актуалност на проблема

Прогнозирането на цените на електроенергията стана ключов фактор при вземането на стратегически решения от страна на енергийните компании и държавните политики в енергетиката. Въвеждането на интелигентни мрежи и възобновяеми източници на енергия увеличи несигурността в бъдещите цени. Поради тази причина, вероятностното прогнозиране на цените, производството и потреблението на електроенергия стана от решаващо значение за планирането и операциите на енергийните системи и пазари.

В условията на либерализирания пазар за електроенергия, лошото планиране на договорите за покупка и продажба на електроенергия с физическа доставка и балансиращият пазар могат да доведат до финансови затруднения в сектора. Това се отнася и за договорите с финансов сетълмент. Затова прогнозирането на цените и количествата стана от съществено значение за енергийните компании и операторите на пазара. Този процес играе ключова роля при вземането на стратегически решения и планирането на регионалните и национални енергийни стратегии.

В дисертацията се разглеждат методи за оценка на волатилността на цените на електроенергията, предоставяйки средства за подобрене на планирането и операциите на енергийните компании, системните оператори и други участници в пазара. Тези изследвания са от решаващо значение във връзка с изразяващата се нестабилност на цените на електроенергията породена от волатилността на производството от ВЕИ, потреблението и необходимостта от намаляване риска при търговията и управлението в енергийния сектор. Разработват се математически методи за точно прогнозиране на цените на електроенергията и за оценка на волатилността, важни за избягване на оперативни и финансови рискове.

Цел и задачи на дисертационния труд

Целта на настоящата дисертационната работа е създаване на методика и алгоритми за изследване, анализ, моделиране и прогнозиране на цените на пазара на електроенергия “Ден напред” част от организирания борсов пазар чрез използване методите за анализ на времеви серии от данни.

За изпълнение на поставената цел се дефинират следните **задачи**:

1. Създаване на методика за изследване и анализ на данните за цените на електроенергията. Избор на подходящи методи, за анализ на реални продължителни (дългосрочни) данни, получени от енергийния пазар в България и региона в зависимост от условията;
2. Създаване на методика за изследване на данните за цените на електроенергията при анализ на реални краткосрочни данни на борсовите цени на електроенергията. Изследване на техники за подобряване на енергийното прогнозиране;
3. Изготвяне на демонстрационни софтуерни процедури за анализ, симулация и прогнозиране на цените на електроенергията чрез прилагане на избраните научни методи и модели за анализ;

4. Определяне на най-ефективните симулационни модели за прогнозиране, с които да могат да се правят бъдещи реалистични анализи за поведението на изследваните данни в зависимост от пазарните фундаменти.

Практическа полезност и приложимост

Дисертацията представя важни научни изследвания и методи, свързани с прогнозирането на цените, потреблението и производството на електроенергия. Тя разглежда как математическите модели и техники за вероятно прогнозиране могат да бъдат приложени в енергийния сектор и на пазара за електроенергия. Този вид прогнозиране е от голямо значение за енергийни компании, операторите на енергийната система и пазара и регулаторни органи, тъй като помага във вземането на стратегически и бизнес-решения. Освен това, дисертацията разглежда и методи за оценка на волатилността на цените на електроенергията, което е от решаващо значение за справянето с финансовите рискове в енергийния сектор. Така, дисертацията предоставя ценни инструменти и знания, които могат да бъдат използвани за подобряване на ефективността и устойчивостта на енергийния пазар и сектор, както и за насърчаване на прехода към без-въглеродна икономика. Получените резултати ще намерят приложение в областта на планирането, изграждането и управлението на нови мощности, управлението на търговски портфолия и баланиране на производството и потреблението при обработка и анализ на данни от електроенергийната система. Също така могат да бъдат използвани в учебния процес на студенти по енергетика, икономика и информационни технологии.

Апробация на резултатите от дисертационния труд

Дисертационният труд е апробиран в основните си части на следните научни форуми:

- Участие пред научен семинар от секция Медицинска роботика, зала "Макс Планк", Квазар – ноември 2022г., София, България;
- Scientific conference with international participation Innovative- STEMEDU-20220, Mart, 2022, Veliko Tarnovo Bulgaria;
- Scientific conference with international participation Innovative STEMEDU-2020, June, 2020, Veliko Tarnovo Bulgaria;
- 21st International Conference on Computer Systems and Technologies (CompSysTech '20). Association for Computing Machinery, June 2020, Ruse, Bulgaria;
- 20th International Conference on Computer Systems and Technologies (CompSysTech '19). Association for Computing Machinery, June 2019, Ruse, Bulgaria;
- XXIII International Conference 'Robotics&Mechatronics, and Social Implementations, June, 2018, Varna, Bulgaria.

Научни публикации

Основните резултати от дисертационния труд са публикувани в 6 статии, издадени в периода 2018-2022 г. Те са докладвани на международни конференции и са издадени в томовете на тези конференции в пълен текст. Публикациите са в авторски колектив. Списък на публикациите е даден в края на автореферата.

Забелязани цитирания

По представените публикации има направени 7 цитирания. Списък на цитиранията е представен в края на автореферата.

Обем и структура на дисертационния труд

Дисертационният труд се състои от увод, 4 глави, заключение, приноси, списък на използваната литература и приложения. Дисертацията е разработена в общ обем от 183 страници. Изложението съдържа 9 таблици и 37 фигури, чиито номера в автореферата съвпадат с тези в дисертацията. Номерацията на точките и подточките също съответства на тази от дисертационния труд. Приложенията са оформени на 22 страници, включващи списък със забелязани цитирания и кодове на софтуерни програми.

КРАТКО СЪДЪРЖАНИЕ НА ДИСЕРТАЦИОННИЯ ТРУД

ГЛАВА ПЪРВА

Обзор на методите за прогнозиране, моделиране и анализ в енергетиката и енергийните пазари

В тази глава на базата на проучване на публикувани научни статии и изследвания на водещи световни и български автори, са показани новостите и тенденциите в развитието на методите за прогнозиране на времевите серии от цени на електроенергия. Обхванати са най-актуалните тенденции за максимизиране на точността и надеждността на прогнозите, като се представени необходимите насоки за използване на методична рамка в съответствие с парадигмата на "максимизиране на надеждността на вероятностните разпределения, предмет на прогнозите". Направен е обзор на съществуващите методи на научното познание за прогнозиране на цените на електроенергията по определени критерии, като е обсъдено функционирането на различни по природа системи, представени чрез модели в резултат на разработката на методите на имитационното моделиране. Изследвани са както краткосрочната, така и дългосрочната времева устойчивост, концепциите и методите, както и техните специфични характеристики и възможности за приложение и моделиране в рамките на проблемите свързани с прогнозиране и анализ в сложни системи, каквито са енергийните.

ГЛАВА ВТОРА

Методи за изследване на дългосрочна устойчивост на времеви серии от данни на борсови цени на електроенергия

Електроенергията е продукт, търгуван на организирания пазари, който се е превърнал в стока с динамични търговски характеристики, нетипични за други пазарни продукти. Динамиката на търговските параметри се дължи на факта, че този продукт не притежава физически възможности да се съхранява и електроенергийната система винаги трябва да поддържа постоянен баланс между производството и потреблението. Основните характеристики на пазарните цени на електроенергията имат сложен характер, променливост и функционална зависимост от времето, функционирането на пазара и енергийната система, търсенето и предлагането, разходите за пренос по мрежата, регионалните специфики като интензивност на индустриалните и държавни интервенции. Предвид това, целта на настоящата глава е да представи научен подход за прогнозиране на цените на електроенергията чрез анализ на времевите редове, използвайки математическия метод на фрактален анализ на времевите редове за пресмятане на експонентата на Хърст чрез R/S и DFA метод.

2.1. Използване на фрактален анализ на времевите редове за пресмятане на експонента на Хърст чрез R/S метода

2.1.1. Дългосрочна памет в речните потоци

Статистиката $(R/S)_n$ се изчислява, като първо се взема диапазонът на частичните суми на отклоненията на даден времеви ред от неговата средна стойност и след това тази величина се преоразмерява чрез оценката на максималната вероятност за стандартното отклонение на времевия ред. Това може да се опише чрез следната формула:

$$(R/S)_n = \frac{1}{s_n} \left[\max_{1 \leq k \leq n} \sum_{j=1}^k (X_j - \mu_n) - \min_{1 \leq k \leq n} \sum_{j=1}^k (X_j - \mu_n) \right] \quad (2.1)$$

където:

$(R/S)_n$ представлява статистиката R/S за дадения период n . Това е мярка за разпределението на отклоненията на времевия ред от неговата средна стойност.

s_n е оценка на максималната вероятност за стандартното отклонение на времевия ред.

X_j са стойностите във времевия ред на позиция j .

μ_n е средната стойност за времеви ред до позиция n с наблюдения $x_1, x_2, \dots, x_n$.

Стандартното отклонение на времевия ред до позиция n е:

$$s_n = \left[\frac{1}{n} \sum_{j=1}^n (X_j - \mu_n)^2 \right]^{1/2} \quad (2.2)$$

Хърст използва своята $(R/S)_n$ статистика, за да анализира времевите редове на 75 различни геофизични явления като валежи, температура, слънчеви петна, цени на пшеницата, както и речни потоци. Мащабирането по s_n дава безразмерна стойност. Важността на мащабирането не е била оценена отначало от няколко автори, които са избрали да изследват свойствата на обикновения коригиран диапазон Rn (формула 2.1).

2.1.2. Стабилни паретиански разпределения

Луис Башелие формулира следния модел за цените на рисковите активи: ако $X(t)$ е цената на рисков актив в края на период от време t , тогава последователните разлики $\Delta X(t) = X(t+T) - X(t)$ се приемат, че са независими и нормално (гаусово) разпределени случайни променливи с нулева средна стойност и дисперсия, пропорционална на интервала на разлика T . Стабилното разпределение на Парето има следната характеристична функция:

$$\varphi(t; \alpha, \beta, \gamma, \delta) = \exp [it\delta - |\gamma t|^\alpha (1 - i\beta \operatorname{sgn}(t)\Phi)], \quad (2.7)$$

където:

$$\Phi = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right) & \alpha \neq 1 \\ -\frac{2}{\pi} \log |\gamma t| & \alpha = 1 \end{cases} \quad (2.8)$$

и $\operatorname{sgn}(t)$ е знаковата функция на t . Задаването на параметрите по подходящ начин (т.е. $\alpha = 2, \beta = 0, \gamma = \delta = 1$) води до характеристичната функция на нормалното разпределение. При $1 \leq \alpha < 2$ дисперсията става безкрайна. По-късно е разработена идеята за фрактално брауново движение (fBm) за моделиране на стохастични процеси, които се характеризират с α в този диапазон.

2.1.3. Измерване на дългосрочна зависимост

Както е споменато по-горе, експонентата на Хърст, чрез която се измерва дългосрочната зависимост, може да бъде оценена, като се използва $(R/S)_n$. Създаден е алгоритъм за

определяне на експонентата за самоподобие на Хърст (H) с помощта на анализ на мащабиран диапазон чрез **R/S** метода.

Основните стъпки на предложения алгоритъм са следните:

1. Разделя се времеви ред с дължина **L** на **d** съседни под-серии с дължина **n**.
2. Намира се средната стойност на пробата $\mu_{n,m}$ и стандартното отклонение на пробата $s_{n,m}$ от всяка подсерия $m = 1 \dots d$;
3. Нормализират се данните $x_{j,m}$ чрез изваждане на средната стойност $Z_{k,m} = x_{k,m} - \mu_{n,m}$ на пробата от всеки елемент $k = 1 \dots n$ във всяка под-серия.
4. Създава се кумулативен времеви ред от всяка подсерия:
 $Y_{k,m} = \sum_{j=1}^k Z_{j,m}$ за всеки елемент $k = 1 \dots n$ във всяка подсерия.
5. Изчислява се диапазона $R_m = \max[Y_{k,m}] - \min[Y_{k,m}]$ и се мащабира диапазона със стандартното отклонение: Rm/Sm .
6. Изчислява се средната стойност на измерения мащаб за всички подсерии:

$$(R/S)_n = \frac{1}{d} \sum_{m=1}^d \left(\frac{R_m}{s_m} \right) \quad (2.9)$$

7. С увеличаване на дължината на подсерията **n**, статистиката $(R/S)_n$ асимптоматично следва отношението:

$$(R/S)_n \sim cn^H, \quad (2.10)$$

където **H** е оценка на експонентата на Хърст, а **c** е константа.

8. Тогава стойността на H може да се получи чрез изчисляване $(R = S)$ n при увеличаване на стойностите на **n** и след това линейно регресиране на стойности на $\log (R/S)_n$ за $\log (n)$. Разбира се, най-голямата стойност, която **n** може да приеме, е **L**, дължината на времевите редове, за които се търси да се определи експонента на Хърст. При логаритмуването на двете страни се получава:

$$\log \left\{ E \left[\frac{R(n)}{S(n)} \right] \right\} \rightarrow \log(C_H) + H \cdot \log(n), \text{ когато } n \rightarrow \infty.$$

Предвид получената зависимост, **H** експонентата може да бъде оценена посредством наклона на регресионната линия в графичната зависимост:

$$\log (E[R(n)/S(n)]) \text{ от } \log (n)$$

$$\hat{H} = \hat{\beta}_1,$$

където:

$\hat{\beta}_1$ - наклон на регресионната линия;

$\hat{H}$ - определена стойност на хърст експонентата.

9. След като се изчисли H, чрез наблюдения за времевите редове остава да се характеризират стохастичните процеси:

- Ако стохастичният процес е гаусов бял шум, тогава $H=0.5$.
- Ако процесът показва дългосрочна памет, тогава $H>0.5$.
- При стойности на $H<0.5$, процесът е антиперсистентен.

2.1.4. Изчисляване на експонента на Хърст чрез R/S анализ

Изчисляването на експонентата с помощта на R/S анализ е реализирано в среда за програмиране MATLAB. Кодът има три основни функции, които са описани по-долу:

Функция делител: Функцията намира стойността, близка до дължината на времевия ред, която има най-голям брой делители. За всеки делител времевите редове са разделени на d подсерии, всяка с дължина n , равна на дадения делител. Функцията делител връща матрица с колони n (дължина на подсерия), d (брой подсерии) и $\log_{10}(n)$.

Функция Хърст: Функцията създава списък с брой на членовете, равен на броя на делителите, изчислен в предишната стъпка. Всеки член е $d \times 4$ матрица. 4-те колони се попълват като:

$$[,1] = d[,2] = \mu_m[,3] = s_m[,4] = R/S \quad (2.13)$$

За всеки член на списъка се изчислява средният мащабиран диапазон $(R/S)n$. Поредицата от стойности $\log_{10}(R/S)$ n от всеки член на списъка се добавя като допълнителна колона към матрицата, генерирана първо от функцията делител. След това функцията Хърст извиква Models Function за изчисляване на експонентата на Хърст.

Функция на модела: Функцията изпълнява линейна регресия на $\log_{10}(R = S)n$ на $\log_{10}(n)$. Наклонът на тази регресия е приблизителна оценка за H . Функцията може да бъде пригодена за оценка на експонентата на Хърст в рамките на периода и между периода.

Алгоритъм за реализация на определяне на експонентата на Хърст чрез метода R/S

Алгоритъмът използва програмната среда MATLAB за изчисляване на експонентата на Хърст чрез метода R/S анализ. Този параметър е важен инструмент за оценка на стационарността и наличието на дългосрочни зависимости във времеви редове или данни. Функцията включва различни стъпки, които позволяват изчисляването на експонентата на Хърст и предоставя широк спектър от статистически данни и доверителни интервали..

2.2. Изследване на времевите редове чрез метода на флуктуационен анализ с елиминиране на хода (Detrended Fluctuation Analysis, DFA)

DFA е метод за изследване на ефектите от дългосрочни корелации във времеви редове. Методът, който ще бъде използван за измерване на зависимостта на дълги разстояния в този раздел, е флуктуационен анализ с елиминиране на тренда. DFA може да се окаже идеалния метод за анализ на дългосрочната памет на цените на електроенергията, тъй като свойства му съответстват на специфичните свойства на спот цените на електроенергията. Метода DFA може да се приложи, като се използват следните стъпки:

- (1) Дефинира се серия от данни $s(i)$, която да бъде анализирана с DFA, където $1 \leq i \leq N$ и N е размерът на серията от данни.
- (2) Изчислява се произволното движение на серията от данни $s(i)$:

$$y(i) = \sum_{k=1}^i [s(k) - \langle s \rangle] \quad (2.14)$$

където $y(i)$ представлява произволното движение на данните в позиция i , $s(k)$ са стойностите в серията на позиция k , а $\langle s \rangle$ е средната стойност на серията данни $s(i)$, която се изчислява по формулата:

$$\langle s \rangle = \frac{1}{N} \sum_{i=1}^N s(i). \quad (2.15)$$

(3) Серията $y(i)$ разделена на $[N/n]$ интервали от данни с еднакъв размер n , където $[\cdot]$ е функцията, която връща най-големия цял числов резултат, който е по-малък или равен на даденото число. Това означава, че съществува разделение на серията $y(i)$ на $[N/n]$ равни интервала с по n стойности във всеки интервал.

(4) Във всеки интервал от данни, данните се обработват чрез използване на полиномна функция на реда l за получаване $y_l(i)$, която се нарича локална тенденция (l -DFA).

(5) Интегрираният сигнал $y(i)$ се детрендира чрез изваждане на локалния тренд $y_l(i)$ във всеки интервал от данни, генериращ детрендирания флукуационен сигнал:

$$Y_l(i) = y(i) - y_l(i). \quad (2.16)$$

(6) За размера на интервала n се изчислява средната квадратична стойност от детрендирания флукуационен сигнал:

$$F_l(n) = \sqrt{\frac{1}{N} \sum_{i=1}^N [Y_l(i)]^2}. \quad (2.17)$$

Това дава информация за вариациите или флукуациите в данните след премахване на локалния тренд и помага при анализа на тези флукуации.

(7) За да се намери връзката между степента на мащабиране и флукуацията на данните, предишните стъпки се повтарят за различни размери на интервала на данните $4 \leq n \leq N/4$.

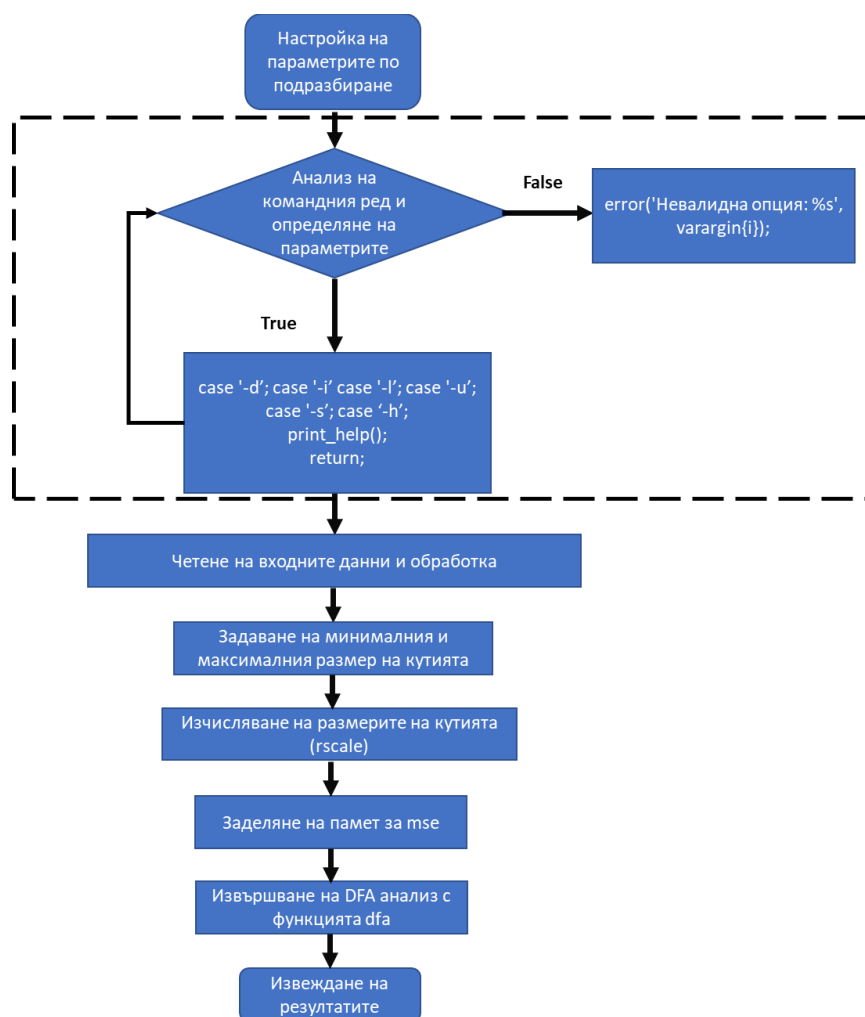
Корелационните свойства на сигнала се характеризират въз основа на степента на мащабирането α . Стойността на α се определя от наклона на правата линия, която се коригира от точките. За линейния случай ($l=1$), ако $\alpha = 0.5$, тогава анализираният сигнал е некорелиран (произволен процес). В случай, когато $\alpha < 0.5$, сигналът е антикорелиран и за $\alpha > 0.5$ сигналът съдържа положителни корелации. Степента на мащабиране α може да се определи с помощта на следната формула, както следва:

$$\alpha = \frac{\log_p(F(n))}{\log_p(n)}, \quad (2.19)$$

Стойността на експонентата α в случай на некорелирани сигнали варира в диапазона $[0, 0.5]$, докато за корелираните сигнали експонентата α е по-голяма от 0.5. Ако стойността на експонентата α е равна на единица, тогава сигналът е $1/f$ шум, но ако тази експонента е равна на 1.5, тогава сигналът се определя като браунов шум. Недостатъкът на този метод е, че е необходима голяма времева серия за надеждно изчисляване на експонентата на Хърст и фракталното измерение.

Предложен алгоритъм за реализация

Предложен нов алгоритъм за изчисление на експонентата на Хърст чрез прилагане на метода за анализ на DFA. Този алгоритъм е реализиран в среда за програмиране MATLAB и изчислява експонентата на Хърст чрез прилагане на DFA върху времеви редове от данни за цени на електрическа енергия. Този код като входни данни използва времевите редове от данни на цените на електроенергия (x) и, при необходимост, опционалния параметър за размер на шрифта (fontsize). Когато се извиква функцията, тя връща стойностите на експонентата на Хърст (H), емпиричните 95% доверителни интервали (pval95) и стандартните отклонения (p). Резултатите също така се изобразяват на графика.



Фигура 2.1: Стъпки на алгоритъма за изчисление на експонентата на Хърст чрез DFA метод.

Сравнение на R/S метода и DFA метода според точността за определяне на H

Относителната грешка на изчислената хърст експонентата се определя по формулата:

$$\Delta H = \frac{|\hat{H} - H|}{H} \cdot 100\%, \quad (2.20)$$

където:

$\hat{H}$ - определена емперична стойност на хърст експонентата; H – теоретична (входна) стойност на хърст експонентата.

Проведени са изследвания за определяне на точността при изчисляване на експонентата на Хърст с 2 метода: R/S метод и DFA метод. Изследванията са проведени с предложените в тази глава алгоритми за намиране на H . За целта са симулирани времеви серии с дължина 100 000 точки. Общо са симулирани 10 такива времеви серии с входна стойност за хърст експонентата 0.6 и получените за тях резултати са усреднени.

Определената стойност на експонентата на Хърст чрез R/S метода за 10 реализирани времеви серии е 0.604. Изчислената относителна грешка ΔH по формула (2.20) в проценти

(усреднена за всички реализации) е 6%. Чрез прилагане на DFA метода получената стойност за 10 реализирани времеви серии е 0.601, което се доближава значително до стойността на входната стойност 0.6. Изчислената относителна грешка ΔH по формула (2.20) в проценти (усреднена за всички реализации) е 1.2%. R/S методът показва положителни относителни грешки при входните стойности на експонентата на Хърст в диапазона от 0.5 до 0.65 и отрицателни относителни грешки при $H > 0.65$, с максимална относителна грешка от 8%. DFA има по-малката относителна грешка, която е по-малка от 2% за всички изследвани стойности на експонентата на Хърст. Това показва, че DFA метода е точен метод и е подходящ за изследване на хърст експонентата.

Методика за изследване на краткосрочна устойчивост на времеви серии от данни на борсови цени на електроенергия и прогнозно моделиране

Създадена е методика за прогнозиране на цените на електроенергия в дългосрочен план. Основни стъпки на методиката са:

Стъпка 1: Събиране на данни за цената на електроенергията.

Стъпка 2: Избор на модел.

Стъпка 3: Предварителна подготовка на данни.

Стъпка 4: Валидация на модела.

Стъпка 5: Прогнозиране.

Стъпка 6. Оценка и актуализация.

Методиката се отнася към метода DFA (като може да бъде приложена и за R/S метода), и трябва да бъде следвана при решаването на поставените задачи. Тя определя стъпките и стратегията, които трябва да се използват за постигане на целта. Методиките за прогнозиране на цените за електроенергия могат да варират в зависимост от конкретните изисквания и наличността на данни, но горепосочените стъпки представляват обща рамка за процеса на прогнозиране.

2.3. Изводи по втора глава

1. Цените на електроенергията са колебливи и значителна част от нестабилността се дължи на дневната и седмичната периодичност и предвид това, чрез използване на R/S метода би могло да се докаже, че данните имат фрактални характеристики и свойства.
2. Изводът, който може да се направи за времевите редове анализирани за период от време е, че във всички разглеждани случаи експонентата на Хърст е по-голяма от 0.5, което е индикатор за фрактална характеристика на наблюдаваните пазари на електроенергия.
3. Създадена е R/S и DFA методика, чрез която са изследвани фракталните свойства на логаритмичните увеличения на спот цената на електроенергията на българския пазар „Ден напред“.
4. Резултатите от прилагането на DFA метода показаха, че поведението на цените са в дългосрочна положителна корелация.
5. Почасовите цени на електроенергията са нестабилни и значителна част от волатилността се дължи на почасовата, дневната, седмичната и сезонна периодичност и метода DFA е подходящ за изследване на цените.

ГЛАВА ТРЕТА

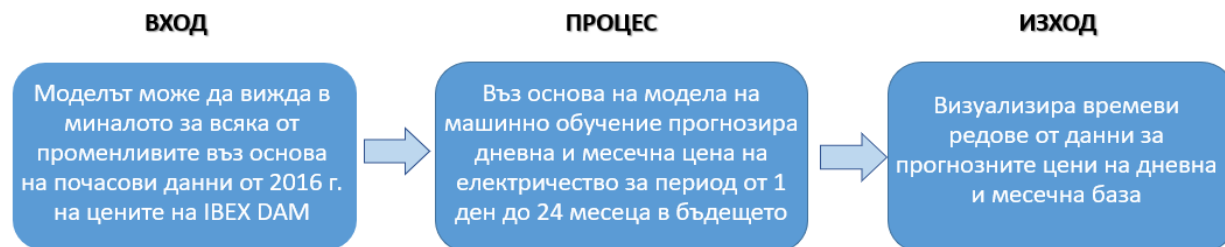
Методи за изследване на краткосрочна устойчивост на времеви серии от данни на борсови цени на електроенергия

3.1. ARIMA метод за прогнозиране на времевите редове от цени на електрическата енергия на пазар „Ден напред“

В тази глава разглеждат няколко подхода за прогнозиране и анализ на динамиката на почасовите цени на електроенергията на пазара за „Ден напред“ в България. В литературата са известни и се използват различни техники за прогнозиране на времевите редове на цените на електроенергията. Тези техники включват динамична регресия и функция на преобразуване, ARIMA (Autoregressive integrated moving average), авторегресивна условна хетероскедастичност (ARCH), хибриден метод на модела ARIMA и модела GARCH (Generalized autoregressive conditional heteroskedasticity), хибриден метод на модела FRWT (Fractional wavelet transform) и модела GARCH. За да се намали грешката на прогнозата и да се увеличи функционалността на невронната мрежа, се развиват непрекъснато нови модели в архитектурата за обучение на невронната мрежа, използвайки различни функции. Моделът на времевите редове и невронната мрежа, които са получени от комбинацията на две линейни и нелинейни системи, се използват за установяване на правилна връзка между входните данни за намаляване на грешката в прогнозата. Техниката за прогнозиране включва моделиране на спот цени на пазар „Ден напред“, като се използва модела ARIMA/SARIMA. Също така се използват и проверки за стационарност, сезонно разлагане, авторегресивно моделиране и автокорелации с цел анализиране и прогнозиране на почасовите данни от времевите редове. За всеки подход са разработени моделни оценки и прогнози, като са използвани данни от почасовите цени, агрегирани данни на дневна и месечна база за българския пазар „Ден напред“.

3.1.1. Методология за анализ и прогнозиране на цените на електроенергията чрез ARIMA

Общият процес за моделиране с ARIMA включва визуализация и агрегиране на данните от времеви редове. Първо, данните, които се използват за прогнозиране, са анализирани с помощта на статистически инструменти, а моделите за сезонност са проверени с помощта на автокорелационен анализ. Данните са декомпозирани и стационарни. Функцията на моделиране е описана на фигура 3.1, като "черна кутия", която има няколко входа и изхода, а връзките се симулират между изходите и входовете.



Фигура 3.1. Процес на моделиране

Моделите ARIMA принадлежат към групата на т.нар. линейни стохастични модели. Чрез тях се моделира развитието на изследваното явление, представено чрез динамичен статистически ред, като се изучават и измерват вътрешните закономерности и връзки,

заложи в самия ред. Най-често се използват следните три вида (според вида на определящите променливи) линейни стохастични модели:

Авторегресивните модели (Autoregressive - AR) изследват дали миналите стойности влияят на текущите стойности.

$$X_t = c + \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t \quad (3.1)$$

Диференцирането може да бъде направено няколко пъти, ако е необходимо, докато данните станат стационарни:

$$y_t^* = y_t' - y_{t-1}' = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2} \quad (3.2)$$

Моделът на плъзгащата се средна (MA) приема, че стойността на зависимата променлива за текущия ден зависи от грешката от предишния ден, както показва формулата:

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (3.3)$$

където μ е средната стойност на серията, $\theta_1, \dots, \theta_q$ са параметрите на модела, а $\varepsilon_t, \varepsilon_{t-1} \dots + \varepsilon_{t-q}$ са грешките на белия шум. Стойността на q е реда на модела MA.

Смесени модели на авторегресия и плъзгащи се средни стойности представляват комбинация на авторегресионни модели и модели на плъзгащи се средни – ARIMA (p,q). Моделът на ARIMA Box-Jenkins добавя диференциации към модел на AIRMA. Диференцирането изважда текущата стойност от предишната стойност и може да се използва за трансформиране на времева серия в стационарни данни:

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} - \theta_1 e_{t-1} \dots - \theta_q e_{t-q} \quad (3.4)$$

Където p е броят на авторегресивните членове (порядък на AR), d е броят на несезонните разлики (ред на диференциране) и q е броят на членовете на подвижната средна стойност (ред на MA).

Ако резултатите от прогнозирането на прогнозираните данни не са добри с помощта на ARIMA, се прилагат методите за сезонно изглаждане на динамични редове с приложение на моделите ARIMA или Seasonal-ARIMA (SARIMA). Моделът SARIMA се използва, когато времевият ред показва сезонност и е подобен на моделите ARIMA, но има няколко параметъра, които трябва да се добавят, за да се отчетат сезоните. За целите на сезонното изглаждане на временни (динамични) редове се появява необходимостта от конструиране на адекватни модели, описващи задоволително точно компонентите на реда. ARIMA включва три взаимосвързани процеса, които биха могли да се разглеждат и поотделно:

- Авторегресионна процедура;
- Процедура на диференциране;
- Процедура на плъзгащите се средни.

За отстраняването на тенденцията на развитие най-често се прилага изчисляването на една или няколко последователни разлики. Процесът на диференциране на реда оказва непосредствено влияние върху точността на конструираните прогнози. Разширеният критерий на Дики-Фулър е тест за единичен корен и като такъв принадлежи към параметричните методи за идентифициране на тенденцията на развитие в динамичните редове. Статистиката на Дики-Фулър (Augmented Dickey Fuller - ADF), използвана в теста,

е отрицателно число. При по-високо отрицателно число, по-силно е отхвърлянето на хипотезата, че има единичен корен на някакво ниво на доверие. Методът на Дики-Фулър за проверка на степента на интегрираност на един динамичен ред е класически тест за стационарност. Той се основава на анализа на единичните корени на характеристичното уравнение. Единичен корен - съответства на границата на областта за стационарност. Ако един времеви ред е от типа единичен корен, то той е със степен на интегрираност от първи порядък. Общият вид на модела на Дики-Фулър е:

$$\Delta Y_t = \alpha + \varphi^* Y_{t-1} + E_t, \quad (3.7)$$

където Y_t представлява реда от първите последователни разлики на Y_t . Тестът на Дики-Фулър проверява статистическата хипотеза за стационарност на конкретен времеви ред. Дефинира следната нулева и алтернативна хипотеза:

H_0 : Динамичният ред е нестационарен, т.е. $\varphi = 1$, ($\varphi^* = 0$);

H_1 : Динамичният ред е стационарен, т.е. $\varphi < 1$, ($\varphi^* < 0$), където връзката между φ и φ^* е следната $\varphi^* = \varphi - 1$.

Когато за един динамичен ред се потвърди нулевата хипотеза и $\varphi = 1$, ($\varphi^* = 0$), тестът доказва наличието на единичен корен в авторегресионната част на модела ARIMA и потвърждава, че изследваният процес е от типа „Unit - root”, като установява, че редът е интегриран от първи порядък - I (1) и обосновава необходимостта от диференцирането му.

Предложен алгоритъм за реализация

Основните стъпки обхващат анализ на времените серии, включително проверка за стационарност, диференциално преобразуване и изчисляване на автокорелации и частични автокорелации. Тези анализи помагат за определение на подходящите параметри за модела SARIMA. След успешния анализ, моделът SARIMA се създава и обучава, използвайки историческите данни за цените на електроенергията. След това се извършват прогнози за бъдещите месечни цени, които са от съществено значение за операторите на енергийни системи, търговци и инвеститори. Тази алгоритъм съчетава теоретичния модел на SARIMA с практични стъпки за анализ на реални времеви серии от данни, което я прави изключително полезна за прогнозиране на цените на електроенергията на пазар "Ден напред", който предоставя мощни библиотеки и инструменти за работа с данни и машинно обучение. Описаният алгоритъм е програмно реализиран на Python.

Основните стъпки на предложения алгоритъм са следните:

1. Зареждане на данни;
2. Четене на данни и подготовка на DataFrame;
3. Проверка за стационарност;
4. Диференциация;
5. Анализ на автокорелация и частична автокорелация;
6. Създаване и обучение на модел SARIMA;
7. Прогнозиране и визуализация;
8. Създаване на бъдещи прогнози.

Този алгоритъм и методиката, върху които се основава, имат критично значение за участниците на пазара, предоставяйки тяхната способност да предвидят ценовите

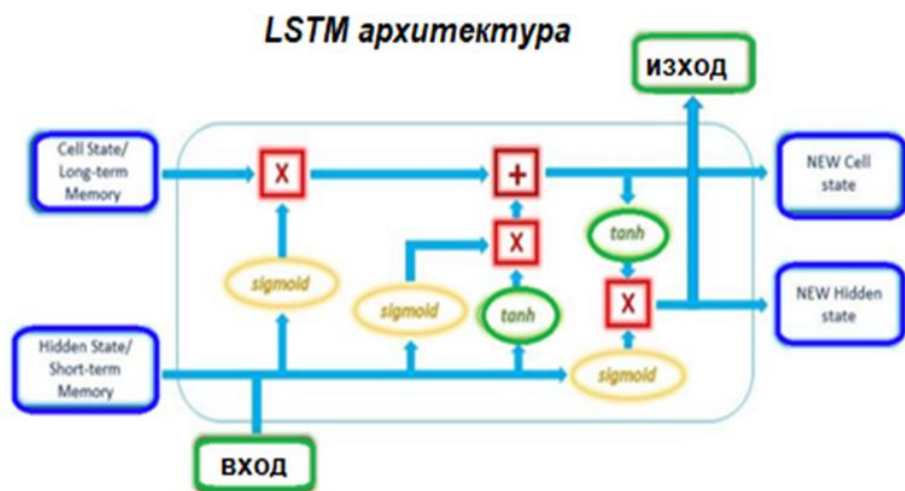
тенденции и да вземат информирани решения относно търговията и инвестициите в енергийния сектор.

3.2. Прогнозиране на цените на електроенергията на пазар „Ден напред“ чрез модел на рекурентна невронна мрежа с дългосрочна-краткосрочна памет (LSTM)

Може да се счита, че архитектурите LSTM и RNN (Recurrent neural network) рекурентните невронни мрежи са добри методи за прогнозиране на цената на електроенергията. В тази глава е използван набор от данни за цените на българския пазар „Ден напред“ с цел прогнозиране на бъдещите цени, използвайки архитектура LSTM за идентифициране на нелинейни краткосрочни времеви зависимости. Това изследване предлага ефективен метод на базата на модел на дългосрочна краткосрочна памет с повтаряща се невронна мрежа LSTM. Като входни данни за модела се използват реални исторически часови цени за периода между 20.01.2016 г. и 05.03.2022 г. Резултат от изследването е прогнозата за цената на електроенергията за часове и дни напред. Бъдещите стойности на цените се прогнозираят рекурсивно.

3.2.1. Методология при анализ чрез рекурентна невронна мрежа с дългосрочна-краткосрочна памет

Рекурентната невронна мрежа с дългосрочна-краткосрочна памет е подобрение спрямо традиционната повтаряща се невронна мрежа RNN в смисъл, че може ефективно да записва дълга последователност от събития в миналото. RNN и LSTM са склонни да извличат информация от исторически събития и да чертаят бъдещи събития. Разликата между RNN и LSTM е, че RNN се използва на места, където запазването на паметта е кратко, докато LSTM е в състояние да свързва събития, които са се случили много по-рано, и събитията, които са ги последвали. Фигура 3.2 визуализира структурата на LSTM, за да дефинира водещите уравнения на LSTM. LSTM има три входа: входна врата, врата за изхвърляне на неважни данни (забраване) и изходна врата, както е показано на фигура 3.2.



Фигура 3.2. Илюстрация на LSTM архитектура

Слоят на входа за изхвърляне на неважни данни (забравяне) показва на клетката LSTM дали да запази историческа информация или напълно да я изхвърли, като при повтарящ цикъл входната врата определя каква нова информация трябва да се добави към състоянието на клетката. Изходната врата най-накрая дава изхода, който е филтрирана версия на състоянието на клетката. Това дава присъщата способност на модела да запазва

информация, която е по-подходяща, и да забравя ненужната информация, която е филтрирана версия на състоянието на клетката. Ето защо LSTM изисква огромен набор от данни по време на процеса на обучение и валидиране за по-добър резултат.

По отношение на входната врата, сигмоидната функция ще преобразува стойностите между 0 и 1, като 0 показва, че част от информацията е маловажна, докато 1 показва, че информацията ще бъде използвана. Това помага да се реши кои стойности да се съхраняват и използват, както и кои стойности да се изхвърлят. Докато слой се обучава чрез обратно разпространение, теглата в сигмоидната функция ще се актуализират така, че да се научи да пропуска само полезното, като същевременно отхвърля по-малко критичните характеристики. Математически това се постига с помощта на 2 слоя:

$$i_1 = \sigma(W_{i_1} \cdot (H_{t-1}, x_t) + \text{bias}_{i_1}), \quad (3.11)$$

Уравнение 3.11 представлява начина, по който се изчислява активацията на неврон i_1 в мрежата, като се умножава входната информация H_{t-1}, x_t с матрицата на теглата и се добавя биасът. След това се прилага функцията за активация "σ" към резултата, за да се генерира изходът на неврона.

$$i_2 = \tanh(W_{i_2} \cdot (H_{t-1}, x_t) + \text{bias}_{i_2}), \quad (3.12)$$

Уравнението 3.12 представлява начина, по който се изчислява активацията на неврон i_2 в мрежата, като се умножава входната информация H_{t-1}, x_t с матрицата на теглата и се добавя биасът. След това се прилага тангенциалната хиперболична функция "*tanh*" към резултата, за да се генерира изходът на неврона. Този процес се използва в обучението на невронни мрежи и може да има различни приложения, включително апроксимиране на функции и анализ на данни.

Резултатите от тези 2 слоя след това се умножават и резултатът представлява информацията, която трябва да се съхранява в дългосрочната памет и да се използва като изход, показан във формулата:

$$i_{\text{input}} = i_1 * i_2, \quad (3.13)$$

i_1, i_2 са изходите или активациите на двата неврона, съответно.

Уравнението 3.13 представлява умножение на изходите на невроните i_1, i_2 .

Резултатите от входната врата и вратата за забравяне ще претърпят поточно добавяне, за да дадат нова версия на дългосрочната памет, която ще бъде предадена на следващата клетка. Тази нова дългосрочна памет също ще бъде използвана в крайната врата на изходния край.

$$C_t = C_{t-1} * f + i_{\text{input}}, \quad (3.15)$$

Формулата представлява начина, по който се обновява състоянието на паметта във времеви ред. Стойността на C_t се изчислява като произведение на предишното състояние на паметта C_{t-1} и стойността f (като резултат от предишното уравнение), след което се добавя входната стойност i_{input} . Този механизъм за обновление на състоянието на паметта се използва в рекурентни невронни мрежи (RNN) и други модели, които трябва да съхраняват

и обновяват информация във времеви редове. Изходът на модела (краткосрочната памет) се изчислява на базата на текущия вход, предишното състояние на краткосрочната памет и новата дългосрочна памет, която се създава. Процесът включва следните стъпки:

Стъпка 1: предишното състояние на краткосрочната памет и текущият вход се подават на сигмоидна функция с различни тегла. Този процес има за цел да създаде третия филтър, който решава каква информация следва да бъде запазена или забравена от предишната краткосрочна памет.

Стъпка 2: поставя се новата дългосрочна памет, като се използва функция за активиране като "tanh". Това позволява моделът да изчисли новата информация, която следва да бъде добавена към дългосрочната памет.

Стъпка 3: резултатът от стъпки 1 и 2, т.е., изходите от сигмоидната функция и новата дългосрочна памет, се умножават, за да се създаде и обнови краткосрочната памет.. Резултатът от тези два процеса ще бъде умножен, за да се създаде новата краткосрочна памет.

$$\begin{aligned} O_1 &= \sigma \left(W_{\text{output}_1} \cdot (H_{t-1}, x_t) + \text{bias}_{\text{output}_1} \right) \\ O_2 &= \tanh \left(W_{\text{output}_2} \cdot C_t + \text{bias}_{\text{output}_2} \right) \\ H_t, O_t &= O_1 * O_2 \end{aligned} \quad (3.16)$$

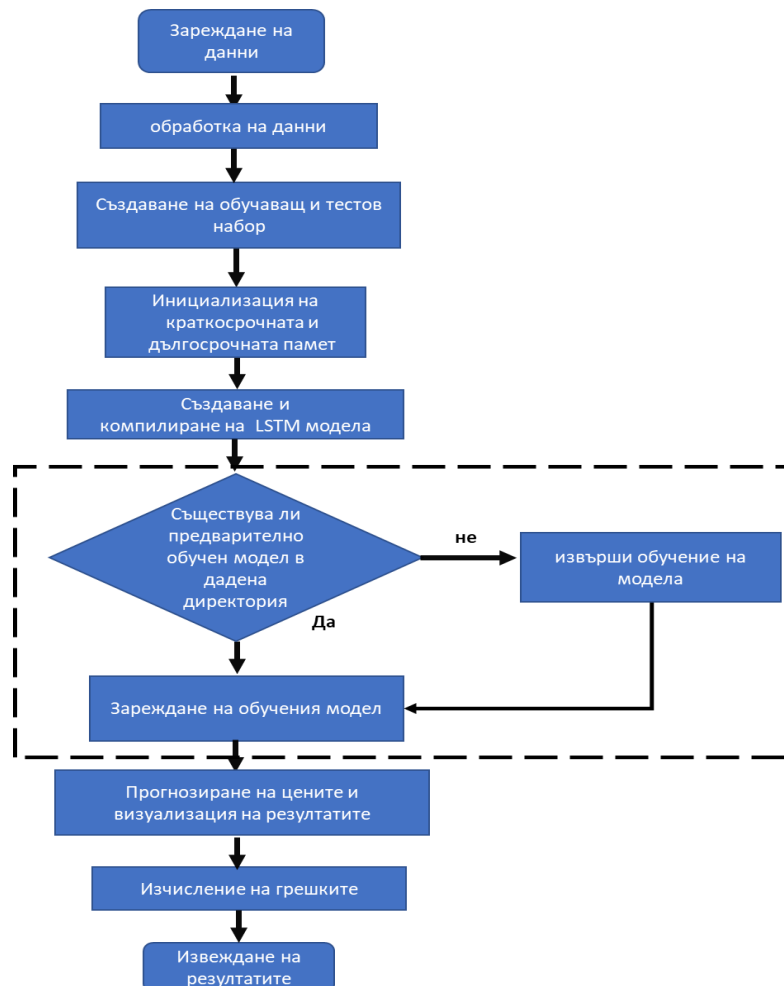
Уравнението 3.16, описва начина, по който се изчисляват два изхода O_1 и O_2 и как те се комбинират, за да се получи H_t и O_t . O_1 и O_2 са изходите на два различни модела или неврона, като " O_1 се изчислява чрез сигмоидна функция, а O_2 чрез хиперболичната тангенциална функция. W_{output_1} и W_{output_2} са матриците на теглата, свързани съответно с O_1 и O_2 . Резултатът от всяка времева стъпка може да бъде получен от краткосрочната памет, известна още като скрито състояние.

Предложен алгоритъм за реализация

С цел справяне с тази сложност и създаване на бъдещи прогнози, изготвен е алгоритъм, базиран на модел на рекурентна невронна мрежа с дългосрочна-краткосрочна памет. Този алгоритъм се основава на обработка и анализ на времевите серии от данни за цените на електроенергията и прогнозирането им с висока точност. Той представлява структуриран и автоматизиран процес за обработка и анализ на времеви редове от данни, които се използват за прогнозиране на бъдещите цени на електроенергия. *Основните стъпки на предложения алгоритъм, показан на Фигура 3.3, са следните:*

- 1. Зареждане на данни:** Зареждане на данни от CSV файл, които се представят като времеви ред на цени.
- 2. Предварителна обработка на данни:** Нормализиране на данните с Min-Max скалиране в интервала $[0, 1]$.
- 3. Създаване на обучаващ и тестов набор:** Разделяне на данните на обучаващ и тестов набор, като се запазват последните 50 стойности за тест.
- 4. Създаване на последователни данни:** Създаване на последователни (sequence) данни с фиксирана дължина от 50 стъпки. Този процес се извършва с цел формиране на данните за обучение и тест.

5. **Създаване на LSTM модел:** Добавяне на слой LSTM с 96 неврона и вход с форма (50,1). Добавяне на Dropout слоеве за намаляване на преобучението. Добавяне на скрит слой LSTM. Завършване на модела с Dense слой.
6. **Компилиране на модела:** Компилиране на модела с функция за загуба (loss function) "mean_squared_error" и оптимизатор "adam".
7. **Обучение на модела:** Проверка дали съществува предварително обучен модел в дадена директория. Ако моделът не съществува, обучаването е с обучаващите данни за 50 епохи.
8. **Зареждане на обученния модел:** Зареждане на предварително обученния модел.
9. **Прогнозиране на цени с модела:** Извършване на прогнози за цените с модела върху тестовите данни. Обратно преобразуване на нормализираните стойности.
10. **Визуализация на резултатите:** Построяване на графики за сравнение между реалните и предвидени стойности на цените.
11. **Изчисление на грешките:** Изчисление на грешките между реалните и предвидените стойности на цените.



Фигура 3.3. Основни стъпки на алгоритъма базиран на модел на рекурентна невронна мрежа с дългосрочна-краткосрочна памет

Предложеният алгоритъм е реализиран на програмния език Python, който е изключително подходящ за обработка на данни и изграждане на невронни мрежи като LSTM модела.

Методика за изследване на краткосрочна устойчивост на времеви серии от данни на борсови цени на електроенергия и прогнозно моделиране

Методиката за изследване на краткосрочна устойчивост на времеви серии от данни на борсови цени на електроенергия и прогнозно моделиране, използвайки ARIMA, SARIMA и LSTM, представлява комплексен аналитичен процес, целящ анализ и прогнозиране на цените на електроенергия на краткосрочна основа. Тази методика включва няколко основни стъпки:

Стъпка 1: Събиране на данни за цената на електроенергията и подготовка на модулите

Стъпка 2: Предварителна обработка на данни

Стъпка 3: Анализ на времевите редове

Стъпка 4: Избор на модел

Стъпка 5: Моделиране и прогнозиране

Стъпка 6: Оценка и тестване

Стъпка 7: Настройка и Оптимизация на Моделит

Тази методика представя общия процес за изследване на краткосрочната устойчивост на времевите серии от данни на борсови цени на електроенергия и прогнозиране на тяхното движение.

3.3. Изводи по трета глава

Разгледани са два метода за изследване и краткосрочно прогнозиране на цените на електроенергия, един базиран на метода SARIMA (Seasonal Autoregressive Integrated Moving Average) и друг, който използва модела LSTM за прогнозиране на времеви редове с цени на електрическа енергия. За целта е използван програмния език Python. Използвайки тези два метода, е разработена методика за краткосрочно прогнозиране и анализ на времевите серии на цените на електроенергията.

1. Прилагането на ARIMA може да се използва като подходящ метод и техника за прогнозиране на стойността на зависима променлива според времето.
2. За наблюдения върху нестационарни времеви редове, които показват сезонни ефекти, тенденции и шумове като пикове и спадове на произволни интервали, които зависят от времето следва да се приложи SARIMA, тъй като цените на електроенергията показват сезонност.
3. Прогнозирането с помощта на SARIMA дава добри резултати при данни, които показват сезонност. Идентифицираният модел SARIMA е подходящ метод за прогнозиране за променливия характер на цените на електроенергията.
4. Прогнозирането с помощта на метода LSTM може да бъде точно за краткосрочна употреба, т.е. за прогнозиране на цените за кратки интервали или почасови данни. Методът LSTM е подходящ за прогнозиране на почасовите и дневните данни от ценни на електроенергия.

ГЛАВА ЧЕТВЪРТА

Отчитане на данните и анализ на резултатите от емпиричното изследване

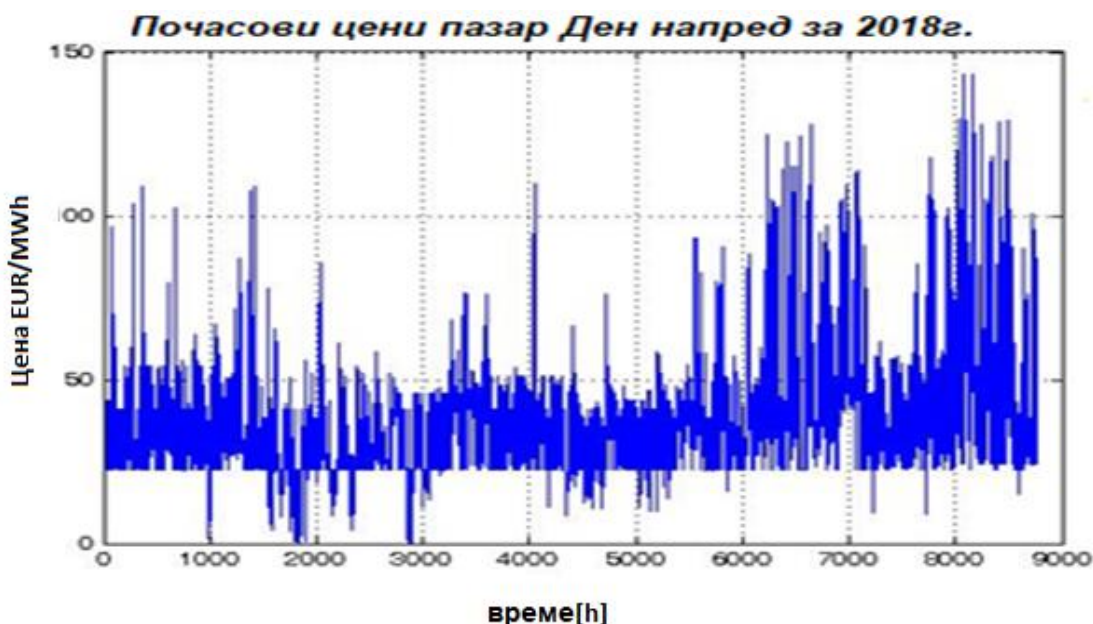
4.1. Фрактален анализ на времевите редове за пресмятане на експонента на Хърст чрез R/S метода

Методът с експонентата на Хърст намира широко приложение при изследването на времеве редове от всякакъв характер и в това число и за анализ на времевите редове на почасовите цени на електроенергията, тъй като е изключително устойчив и позволява да се разграничат случайните редове с напълно независими едно от друго нараствания от неслучайните редове. Изчисленията чрез R/S метода в тази глава са направени чрез разработка на собствени алгоритми и компютърни програми, реализирани на програмен език в среда на MATLAB. За определянето на стойностите на хърст експонентата със статистическите методи са реализирани 10 повторения на симулираните процеси с дължина от около 50 000 – 60 000 точки, предвид почасовите данни за период от около 6 години.

4.1.1. Описание на входните данни

Използвани са набор от данни от българския пазар „Ден напред“, който обхваща периода между 1 януари 2018 г. и 31 декември 2018 г. Този период е избран, защото това е периодът, който бележи преломната точка по отношение на появата на изключително високи цени.

Представените данни на фигура 4.2 са публично достъпни и могат да бъдат изтеглени от уеб страницата на БНЕБ ЕАД. Анализираните данни от времевите редове съдържат информация за почасовите цени на пазар „Ден напред“ за всеки час от 1 януари 2018 г., 00:00 ч. до 31 декември 2018 г., 24:00 ч. Това дава общо 8 761 точки от данни за всеки час и това е илюстрирано на фигура 4.2.



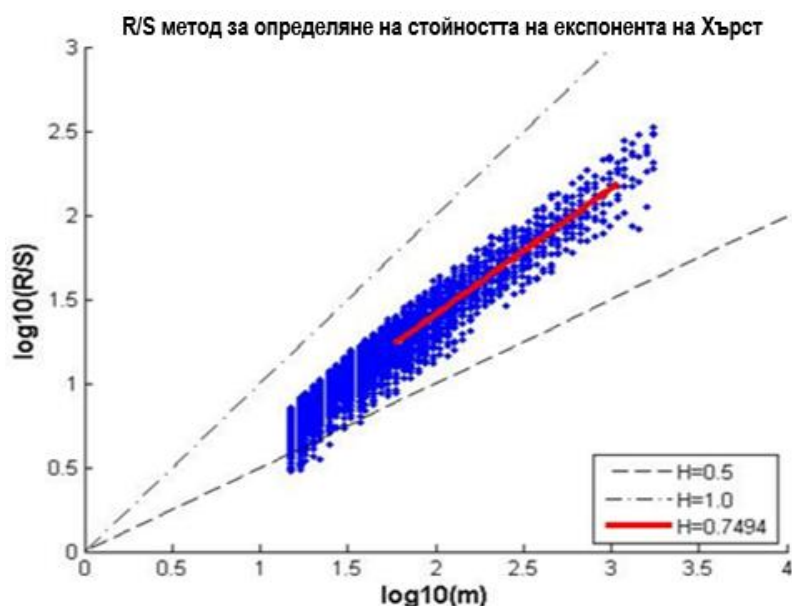
Фигура 4.2: Почасови клирингови цени за пазар „Ден напред“ за 2018 г., източник уеб страница на БНЕБ ЕАД (<https://ibex.bg/>)

Пазарите на електроенергия се характеризират с чести и внезапни скокове и с бързо връщане към средната стойност на цените или по-конкретно към предишните ценови нива.

За справяне с нестационарната средна стойност и за да се направи оценка на дългосрочната зависимост могат да бъдат приложени различни методи като DFA, който се използва, за да се оцени експонентата на Хърст, която количествено определя наличието на дългосрочна зависимост и изучава монофракталните сигнали. Направено е сравнение между видовете поведение на пазарите, както в добре развитите, така и в нововъзникващите пазари със спот цени, включително за българската енергийна борса и за съседните енергийни борси като HUPX-Унгария, OPCOM-Румъния, SEEPEX-Сърбия и HENEX-Гърция. Изследването е изцяло базирано на методи за анализ на експонентата на Хърст.

4.1.2. Резултати от изследванията

Резултатите са получени чрез метода на фракталния анализ на данни от две времеви серии: данни от времеви серии на клирингови цени за пазар „Ден напред“ за всеки час за 2018 г. или 8 761 точки от данни, показващи фрактално поведение и други симулирани монофрактални данни от времеви серии. При стойности на $H < 0.5$ се очаква, че цените обикновено не се отдалечават много от средната стойност за големи периоди от време. В такива случаи се казва, че цените са непостоянни или показват тенденция за връщане към средните стойности. Обратно, при $H > 0.5$, нарастванията могат да имат същия знак за по-дълъг период от време и следователно цената може да се отдалечи от средната стойност за по-дълъг период от време. В този случай се казва, че цената е постоянна. Резултатите от изследванията чрез R/S метода, приложен към анализираниите времеви редове на спот цените на българския пазар „Ден напред“ и стойността на експонентата на Хърст, са показани на фигура 4.5. Определената стойност на експонентата на Хърст за анализираниите данни е изчислена на $H = 0.7494$. Полученият резултат показва, че анализираниите времеви редове са корелирани, тоест представляват фрактални времеви редове. Това разкрива, че ценовото формиране е свързано или че е по-вероятно увеличение на цената в миналото да бъде последвано от допълнително увеличение, отколкото от намаление на цената. Следователно ежедневните движения на цените са постоянни и подчинени на тенденции.



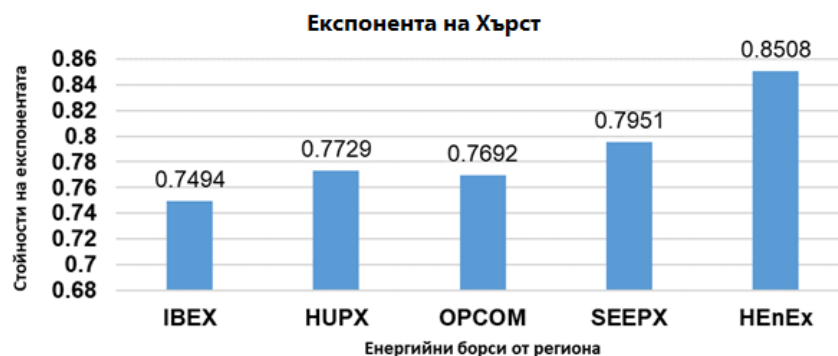
Фигура 4.5: Резултати от метода R/S, приложен към изследваните сигнали за определяне на стойността на експонентата на Хърст с данни от БНЕБ за 2018г.

Разработеният метод и софтуер за моделиране и анализ на фрактални данни се основават на използването на MATLAB и имат универсален характер по отношение на използването на данни от времеви редове. В Таблица 4.1 са обобщени получените резултати за изчислените стойности на изследваните параметри за цените за „Ден напред“ на съседните енергийни борси IBEX - България, HUPX - Унгария, OPCOM - Румъния, SEEPX - Сърбия и HEnEx - Гърция. Таблица 4.1 показва типичните характеристики на индикаторите за волатилност на цените на електроенергията на пазар „Ден напред“: средна стойност, медиана, стандартно отклонение, асиметрия, ексцес, максимална и минимална стойност на наблюдаваните данни за почасовите вариации на цените на електроенергията за „Ден напред“ през 2018 г. за IBEX и съседните борси.

Таблица 4.1: Стойности на изследваните параметри за IBEX и съседни енергийни борси HUPX, OPCOM, SEEPX и HEnEx. Данните са публично достъпни и могат да бъдат изтеглени от интернет страниците на борсите за електроенергия за 2018г.

Параметри	IBEX	HUPX	OPCOM	SEEPX	HEnEx
Експонента на Хърст	0.7494	0.7729	0.7692	0.7951	0.8508
Средно [EUR]	39.77	51.10	46.58	50.22	60.38
Медиана [EUR]	38.36	50.24	44.89	50.00	61.39
Стандартно отклонение [EUR]	19.31	19.04	21.37	19.03	11.34
Изкривеност	1.76	0.18	0.19	0.67	1.20
Ексцес	4.36	0.30	0.09	6.46	4.78
Минимум [EUR]	0.02	-25.97	0.00	0.00	0.00
Максимум [EUR]	143.18	124.95	124.95	282.40	145.00

Стойности на експонентата на Хърст по-големи от 0.5, съответстват на характеристика на дългосрочна памет, за разлика от по-ниските стойности на експонентата на Хърст. Както се вижда от таблица 4.1, наблюдаваните стойности на експонентите на Хърст варират приблизително между 0.7 и 1 и това показва дългосрочна памет на процеса. От резултатите, представени в таблица 4.1 и фигура 4.6, се вижда, че стойностите на експонентата на Хърст са: $0.5 < H < 1$, което показва устойчиво поведение или положителна автокорелационна тенденция. Стойността H , която е по-близка до единица, показва висок риск от големи и резки промени и следователно може да се заключи, че за HEnEx стойността на H е най-високата сред сравняваните стойности за съседните пазари, а за IBEX стойността на H е най-ниската. Извършеното моделиране и анализ показват повишен риск от нестабилност на цените при прогнозиране на HEnEx. Получените стойности на експонентата на Хърст показват, че времевите редове имат дългосрочна положителна автокорелация и че високата стойност в серията с голяма вероятност ще бъде последвана от друга висока стойност и че бъдещите дългосрочни стойности също ще бъдат високи.



Фигура 4.6: Експонентата на Хърст на изследваните времеви редове за цените на пазар „Ден напред“ на БНЕБ и съседните ѝ енергийни борси HUPX, OPCOM, SEEPX и HenEx за 2018г.

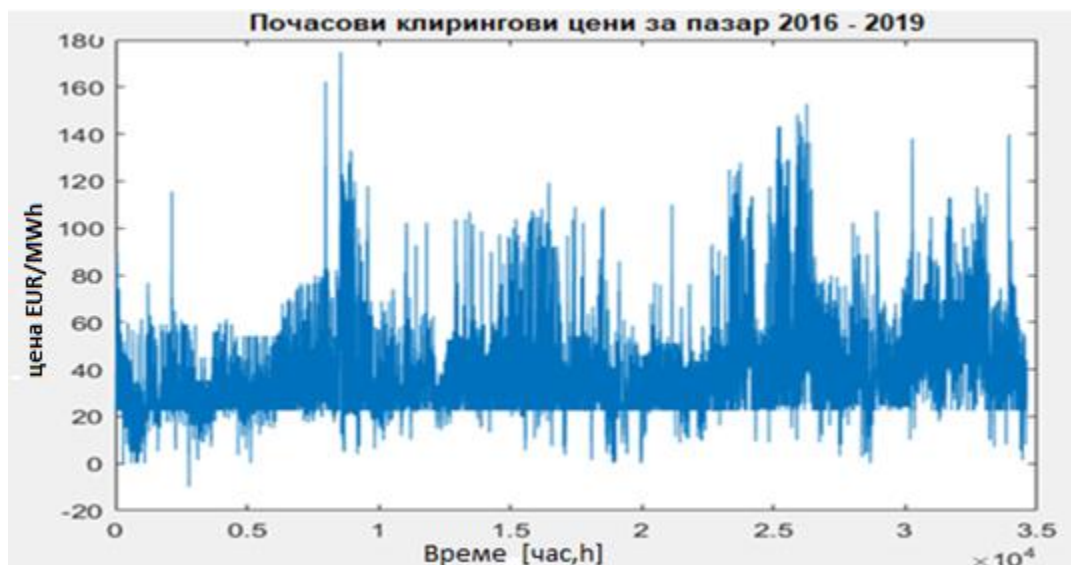
4.1.3. Изводи от направените изследвани яа пресмятане на експонента на Хърст чрез R/S метода

Получените резултати от този метод показват, че изследваните данни проявяват фрактални характеристики и свойства. Стойността, измерена за експонентата на Хърст на българския пазар „Ден напред“ $H=0.75$ показва постоянни времеви серии с дългосрочна памет. Това означава, че случващото се днес, влияе върху бъдещето. Тази дълга памет се появява независимо от времевата скала. Ежедневните промени са свързани с всички бъдещи ежедневни промени и седмичните промени ще корелират с всички бъдещи седмични промени. Изводът, който може да се направи за времевите редове анализирани за период от време е, че във всички разглеждани случаи експонентите на Хърст са с по-голяма стойност от 0.5, което е индикатор за фрактална характеристика на наблюдаваните пазари на електроенергия.

4.2. Емпиричен анализ на времевите серии на цените на електроенергия чрез DFA
DFA се използва за анализиране на времеви редове с такава сложна структура като спот цените на електроенергията за да се определи фракталната характеристика на нестационарни и стационарни времеви редове. Подробен анализ на пазарните цени на електроенергията за „Ден напред“ и тяхната динамика не е забелязано да е правен в България. За целта на изследването в тази глава е използван програмния език MATLAB.

4.2.1. Описание на входните данни и анализ

Предмет на анализ са почасовите спот цени на пазар „Ден напред“ на българския електроенергиен пазар между 20 януари 2016 г. и 31 декември 2019 г. Общия брой на наблюдения в разглежданата времева серия са 27 986 точки от данни за всеки час, което е илюстрирано на фигура 4.8. Изследваните данни са публично достъпни и могат да бъдат изтеглени от страницата на Българската енергийна борса



Фигура 4.8.: Почасови цени за пазар „Ден напред“ между 20.01.2016 г. и 31.12. 2019 г.

Таблица 4.2 показва типичните характеристики на индикаторите за променливост на цените на електроенергията за „Ден напред“: средна стойност, медиана, стандартно отклонение, асиметрия, ексцес, максимална и минимална стойност на наблюдаваните данни на почасовите вариации на цените на електроенергията за пазар „Ден напред“.

Таблица 4.2. Обобщение на резултати от анализ на почасовите цени на пазар „Ден напред“ на БНЕБ

Почасови цени на пазар „Ден напред“ на БНЕБ	Цени [EUR/MWh]
Брой елементи в серията	27 986
Средна стойност	39.77648288
Медиана	38.36
Стандартно отклонение	19.4253
Асиметрия	1.7609
Ексцес	4.3622
Минимална стойност	(10.00)
Максимална стойност	174.83

4.2.2. Резултати от изследванията

Анализът чрез DFA е метод, който има за цел да определи параметъра α , използващ се за измерване на поведението на фрактално мащабиране и също за оценка как амплитудата на флукуациите в сигнала зависи от мащаба, в който се измерват. Чрез DFA и оценката на мащабиращия показател α по години, количествено се определя фракталната структура на българските почасови спот цени на електроенергия. Линеината връзка върху двойната логаритмична функция на $\bar{F}(n)$ спрямо размера на интервала n показва наличието на степенно мащабиране на формата $F(n) \sim cn^H$, където H е фрактална DFA експонента. Ако $0.5 < \alpha < 1$, тогава степенната на мащабиране $\alpha \approx H$. Ако наклонът е по-голям от 0.5, то процесът е постоянен, докато за анти-постоянен процес наклонът е по-малък от 0.5. По този начин стойност на $\alpha = 0.5$ показва, че няма корелации (или има само на къси разстояния). Ако $\alpha > 0.5$ за всички скали, данните са дългосрочно корелирани. Колкото по-висока е стойността на α , толкова по-силни са корелациите в сигнала. $\alpha > 1$ показва нестационарна локална средна стойност на данните. В този случай флукуационният анализ е неуспешен и дава само $\alpha = 1$. Но ако този параметър е равен на 1.5, тогава сигналът се определя като Браунов шум. Случаят $\alpha < 0.5$ съответства на дългосрочни антикорелации, което означава, че големите стойности е най-вероятно да бъдат последвани от малки стойности и обратно. Резултатите от направения анализ по години са показани в таблица 4.3.

Таблица 4.3. Резултати от прилагане на метода DFA

Година	2016	2017	2018	2019
Коефициент на корелация α	0.9851	0.8924	0.9111	0.9574

Установено е, че стойността на експонентата е близо до 1 за периода от 2016 до 2019 г. на всички серии от почасови данни с малка вариация през 2017 г. Пазарното поведение показва силна дългосрочна корелация, което означава, че колкото по-висока е стойността на α , толкова по-силни са корелациите в сигнала и че корелациите, присъстващи във времевите редове, са с по-големи колебания в по-дълги времеви мащаби. Данните в таблица 4.3 за α показват съответствие на спектъра на мощността и методите на DFA.

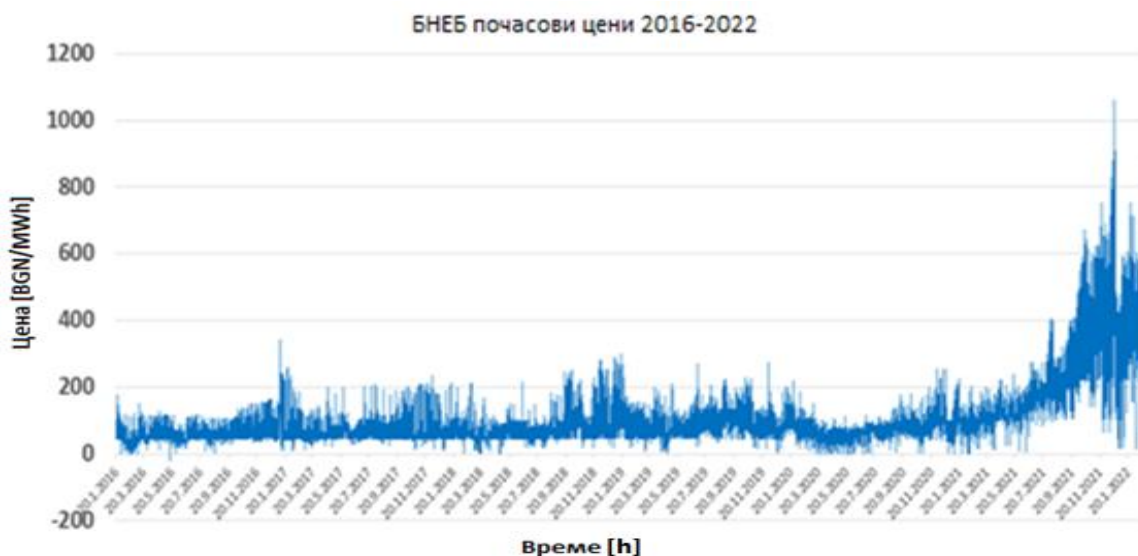
4.2.3. Изводи от направените изследвания

Резултатите от методологията на DFA показаха, че поведението на цените са в дългосрочна положителна корелация. Почасовите цени на електроенергията са нестабилни и значителна част от волатилността се дължи на почасовата, дневната, седмичната и сезонна периодичност. Следователно сигналите могат да се анализират с помощта на DFA метод за анализ. Получените резултати от всички тези методи показват, че изследваният сигнал има мултифрактални свойства. $F(\alpha)$ показва колко често могат да се появят събития с α мащабиращ показател. Ширината на фракталния спектър показва разликата между максималната вероятност и минималната вероятност. В нашия случай широчината на спектъра разкрива, че има силна мултифракталност при възвръщаемостта на почасовите цени на пазара „Ден напред“. Показано е, че данните за цените във времеви редове притежават устойчиво поведение с дългосрочна зависимост. Това постоянство показва на практика същия знак за следващия покриващ се времеви сегмент в реда, което означава, че увеличаването на процеса е най-вероятно да бъде последвано от увеличение в следващия времеви сегмент и обратното, намаляване на процеса е най-вероятно да бъде последвано от намаление в следващия времеви сегмент.

4.3. Емпиричен анализ и резултати от моделите за прогнозиране на времевите серии от дневни и месечни цени на електроенергия чрез ARIMA

4.3.1. Описание на входните данни и анализ

Диапазонът на прогнозиране е с продължителност от 24 часа с почасова честота. Базовият модел използва исторически цени за прогнозиране на бъдещи стойности. Данните съдържат реални почасови данни за цените за периода от 20.02.2016 г.-05.03.2022 г. за пазара „Ден напред“ - България, което е показано на фигура 4.17.



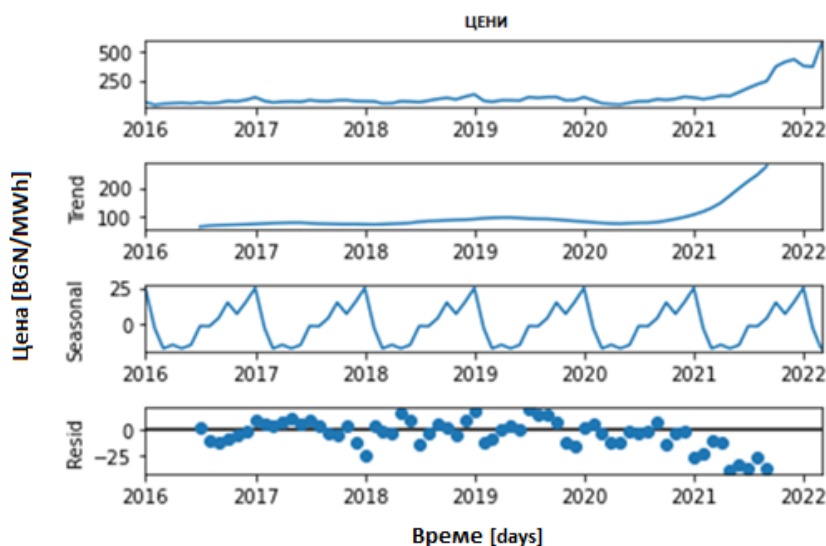
Фигура 4.17. Почасови данни за цените на пазар „Ден напред“- България.

Данните са групирани на почасови, дневни и месечни цени и след това обобщени, тъй като цените на електроенергията следват модели за дневни, седмични и сезонни времеви рамки. Основната цел е да се прогнозира среднодневната цена на електроенергията на БНЕБ ЕАД за дни и месеци напред. След успешно прилагане на методите на пречистване, подобряване

и трансформиране, данните преминават в етап зареждане в Pandas Library, част от програмният език Python.

4.3.2. Анализ на данни от дневни цени

Времевите редове са стационарни, ако те остават постоянни във времето. Времевият ред от теста на Дики-Фулър се счита за стационарен, ако p -стойността е по-ниска от 0.05. Стационарни данни са данни, които нямат тенденция по отношение на времето. Декомпозицията на данни от времеви редове е метод за разделяне на серии от данни на компоненти като тенденция, сезонен и неравномерен шум. Можем да разложим данните от времевите серии с функцията `seasonal_decompose`, както е показано на фигура 4.20 по-долу.



Фигура 4.20. Декомпозиция на данните за периода 20.02.2016 г. до 05.03.2022 г. за Българския електроенергиен пазар „Ден напред“

След като вече са избрани променливите на изследването и са създадени времевите редове, те се подлагат на проверка за стационарност. Това става чрез прилагането на разширения тест на Дики – Фулър (ADF). Проверяваната нулева хипотеза е за наличието на единичен корен, за да се изследва дали серията е стационарна или не след разликите. Резултатите от тестването за наличие на единичен корен на динамичните времеви редове за цените на електроенергията са представени в таблица 4.4. От данните може да се счита, че има слаби доказателства срещу нулевата хипотеза, времевият ред има единичен корен, което показва, че е нестационарен.

Таблица 4.4. Резултати от разширен тест на Дики-Фулър

Статистика на теста на ADF:	1.850218578480598
p -стойност:	0.9984435851817096
#Използвани закъснения:	27
Брой използвани наблюдения:	2209

Тестването на променливите във времеви редове за цените на електроенергията показва, че те са нестационарни. Стойността на p е по-голяма от 0.05 и това означава, че времевият ред е нестационарен. С други думи, той има някаква зависима от времето структура и няма постоянна вариация във времето. В случая, когато временният ред е нестационарен, стойностите на автокорелационните коефициенти намаляват много бавно и не достигат

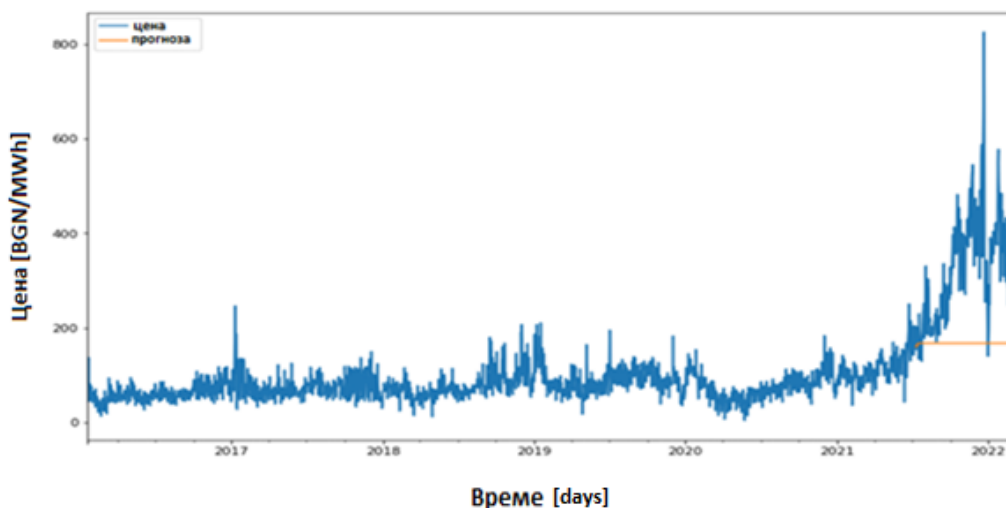
нулата дори и за отдалечените лагове. Проверката за статистическа значимост на автокорелационните коефициенти се реализира посредством приложението на теста на Бартлет, съгласно който при стационарен динамичен ред автокорелационните коефициенти следват приблизително нормално разпределение със средна стойност 0 и стандартно отклонение $1/T$, където T е броят на членовете на временния ред. Установяването на статистическата значимост се извършва посредством построяването на доверителен интервал, който ограничава статистически незначимите коефициенти. Математическият анализ на нестационарни времеви серии води до ненадеждни резултати. Поради тази причина е необходимо данните от времевите серии да бъдат преобразувани, така че да станат стационарни. Ако нестационарният процес следва произволни движения, то той се преобразува в стационарен процес чрез разграничаване. От друга страна, ако анализираните данни от времеви серии показват детерминистична тенденция, може да се предотврати генерирането на неточни резултати чрез коригиране. Резултатите от втория разширен тест на Дики-Фулър, с цел да се провери дали серията остава нестационарна след прилагането на разликата, са показани в таблица 4.5.

Таблица 4.5. Резултати от втори разширен тест на Дики-Фулър след първо диференциране

Статистика на теста на ADF:	-9.491961306426585
p-стойност:	3.632537140497319e-16
#Използвани закъснения:	27
Брой използвани наблюдения:	2209

Идентифицирането на AR модела се извършва с частична автокорелационна функция (Partial Autocorrelation Function - PACF). Идентифицирането на MA модел често се извършва най-добре с ACF, а не с PACF. Ползват се параметрите P, D, Q, които се използват за идентификация и параметризация на ARIMA модела, където P е лаг на AR модела, D е диференциацията и Q показва изоставането на MA. За несезонни данни P=1 (спецификация на AR), D=1 (ред на интегриране), Q=0 или 1 (спецификация на MA/полином).

На фигура 4.23 са показани прогнозните стойности получени чрез модела на ARIMA спрямо действителните данни за целия период на прогнозиране. От фигурата може да се види, че прогнозирането не показва добри резултати с помощта на модела ARIMA, тъй като времевите редове все още показват сезонност. След като с помощта на модела ARIMA прогнозните дневните данни не дават добри резултат, ще бъде приложен модела SARIMA.



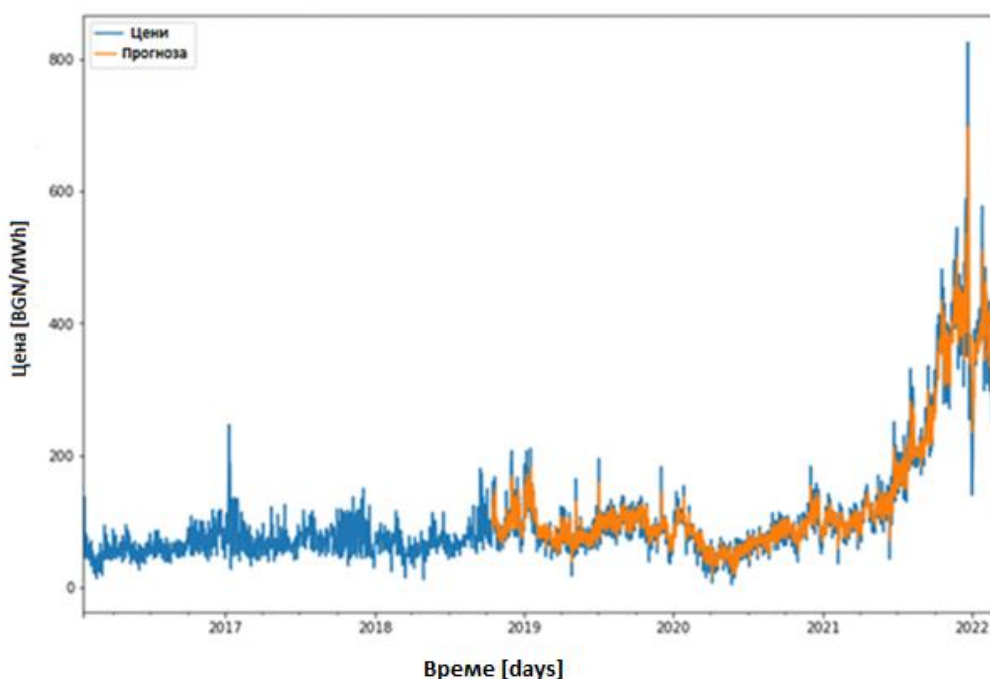
Фигура 4.23. Приложен ARIMA модел, генериращ дневни прогнози, показващ реални цени в син цвят и прогнозни стойности в оранжев цвят

Моделът SARIMA се използва, когато времевият ред показва сезонност. Този модел е подобен на моделите ARIMA, като към него трябва да се добавят няколко параметъра, за да бъдат отчетени сезоните. Сезонното диференциране отчита сезоните и разликите между текущата стойност и нейната стойност през предходния сезон. Остатъкците са бял шум и те също са нормално разпределени. Данните от модела са показани в таблица 4.6. (резултат от изпълнението на създадената софтуерна програма), като от него може да се направи констатацията, че моделът отговаря на условието за независимост в остатъците (без корелация), тъй като p -стойността на теста на Ljung-Box ($\text{Prob}(Q)$) е по-голяма от 0.05, така че не можем да отхвърлим нулевата хипотеза за независимост, но и не можем да кажем, че остатъчното разпределение е хомоскедастично (постоянна вариация), тъй като p -стойността на теста за хетероскедастичност ($\text{Prob}(H)$) е по-малка от 0.05.

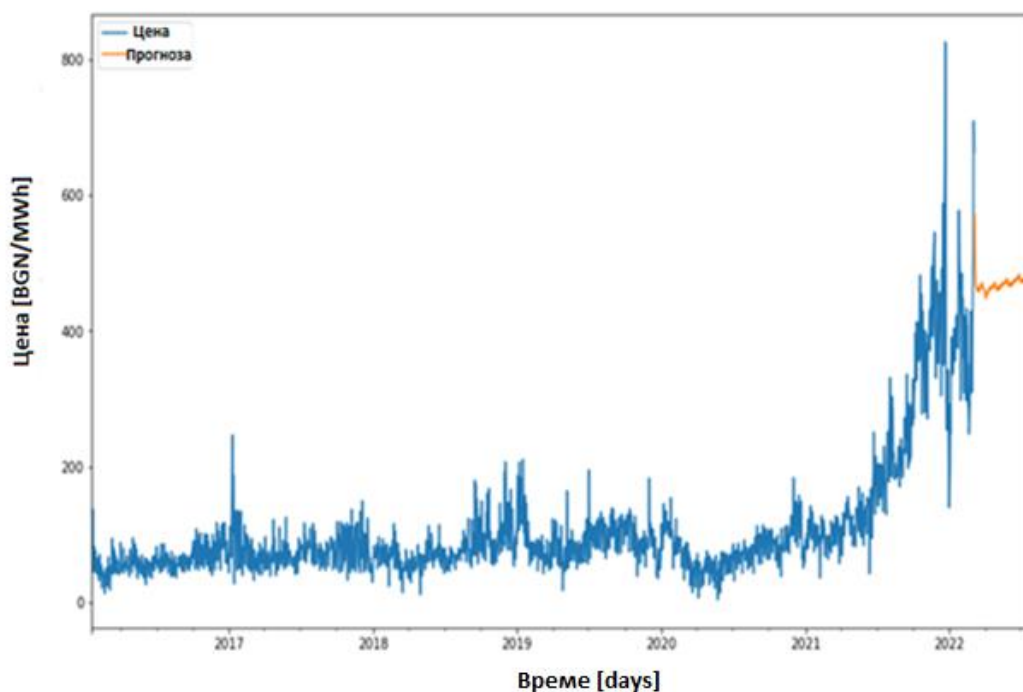
Таблица 4.6. Модел SARIMA

SARIMAX Results						
Dep. Variable:		Price	No. Observations:		2237	
Model:		ARIMA(1, 1, 1)	Log Likelihood		-10614.854	
Date:		Mon, 07 Mar 2022	AIC		21235.707	
Time:		21:34:53	BIC		21252.845	
Sample:		01-20-2016	HQIC		21241.964	
- 03-05-2022						
Covariance Type:		opg				
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.5672	0.011	52.780	0.000	0.546	0.588
ma.L1	-0.9238	0.006	-167.412	0.000	-0.935	-0.913
sigma2	777.6259	9.005	86.359	0.000	759.977	795.275
Ljung-Box (L1) (Q):		0.20	Jarque-Bera (JB):		15330.11	
Prob(Q):		0.66	Prob(JB):		0.00	
Heteroskedasticity (H):		3.39	Skew:		0.49	
Prob(H) (two-sided):		0.00	Kurtosis:		15.79	

От фигура 4.24 е видимо, че прогнозната стойност на времевите серии от цени следва плътно действителната стойност на дневните цени, тъй като от общо усреднени данни за 2237 дни се започва от стойността след 1 000 и се завършва на стойност 2 300, което добавя общо 63 точки във времеви ред или 63 дни прогнозни стойности. Фигура 4.24 сравнява тези стойности със средната стойност на тестовата серия, за да провери дали величината на грешката е приемлива. Прогнозирането с помощта на SARIMA дава добри резултати, тъй като данните показват сезонност. Резултатите от данните от прогнозните дневни цени получени със SARIMA за 150 дни напред и данни от цените за изминал период от 2016-2022 са показани на фигура 4.25.



Фигура 4.24. Приложен модел SARIMA за дневни прогнози, проверка на грешки между данни от реални цени и данни от прогнозиран цени



Фигура 4.25. Данни от резултатите от прогнозни дневни ценни със SARIMA за 150 дни напред и исторически данни за цените

4.3.3. Анализ на данни от месечни цени

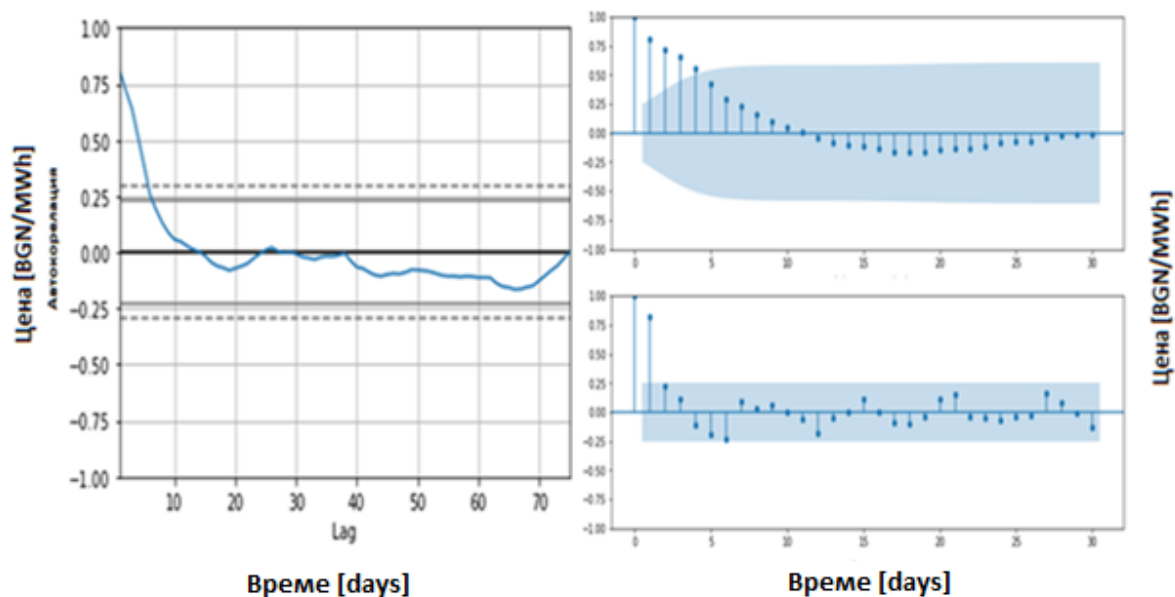
Анализирани са времевите редове на средно месечните цени на електроенергията чрез създадения алгоритъм и програмния език Python. С цел стабилизиране на реда и трансформиране на процеса в слабо стационарен процес са приложени първа и втора

сезонни диференциации. За да се тества дали серията е стационарна или не след следващото диференциране са направени няколко разширени тестове на Дики-Фулър. Втората тестова статистика на ADF показва слаби доказателства срещу нулевата хипотеза, времевият ред има единичен корен, което показва, че е нестационарен и следва да се направи второ разграничение. Резултатите показват, че данните нямат единичен корен и са неподвижни. Направено е второ сезонно диференциране.

Таблица 4.8. Резултати от трети разширен тест на Дики-Фулър, второ диференциране

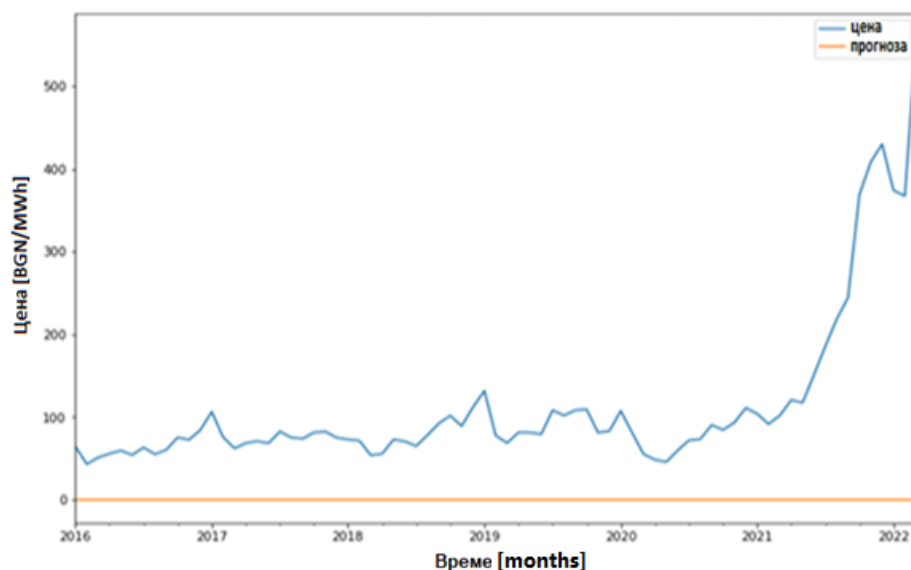
Статистика на теста на ADF:	-7.3183330481968865
p -стойност:	1.2126045470088354e-10
#Използвани закъснения:	5
Брой използвани наблюдения:	55

Идентифицирането на AR модел се извършва с техниката PACF. За несезонни данни $P=1$ (AR спецификация), $D=1$ (ред на интегриране), $Q=0$ или 1 (спецификация на MA/полином). В графиката по-долу на фигура 4.28 всеки пик т.е. закъснение над пунктираната област се счита за статистически значим.



Фигура 4.28. Диференциране и автокорелация на месечни данни

Фигура 4.29. показва прогнозите на ARIMA спрямо действителните данни за целия период на прогнозиране. Може да се види, че прогнозните данни не дават добри резултати с помощта на модела ARIMA, тъй като времевите редове показват сезонност. Затова ще бъде приложен модела Seasonal-ARIMA.



Фигура 4.29 Прогнозни месечни данни, които не дават добри резултат с помощта на модела ARIMA

Графиката с резултатите от създадената софтуерна програма, след прилагането на модела е показана в таблица 4.9. За прогнозиране на стойности за изследваните данни и след това проверка за точност с реалните данни е използван модел със следните параметри: order=(1, 1, 1), seasonal_order=(1,1,1, 12)) за ARIMA(p,d,q)x(P, D, Q)m стойности.

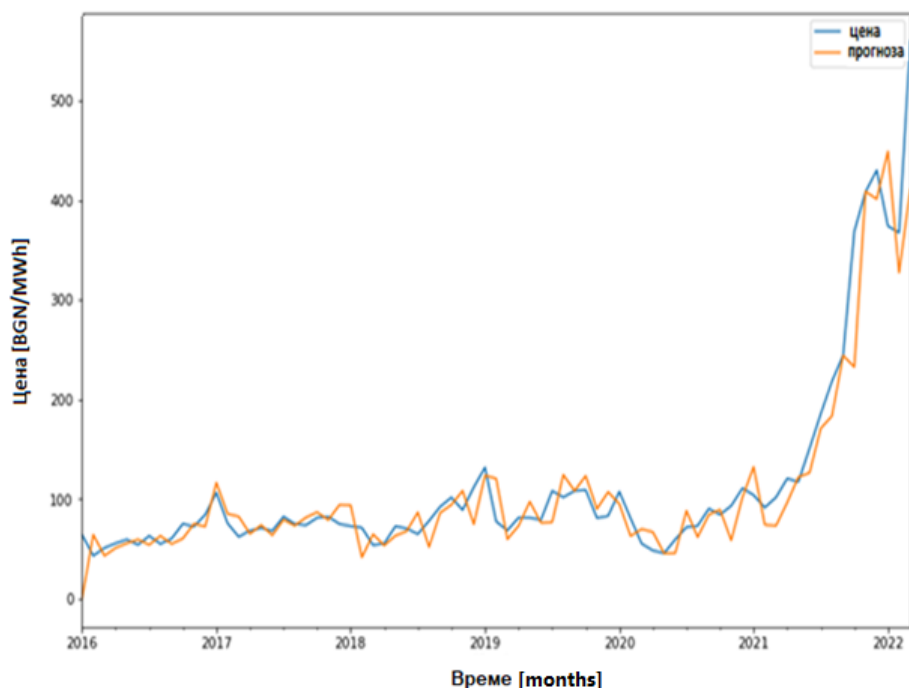
Таблица 4.9. Модел SARIMA

SARIMAX Results

Dep. Variable:	Price	No. Observations:	75			
Model:	ARIMA(2, 1, 1)	Log Likelihood	-357.649			
Date:	Mon, 22 Aug 2022	AIC	723.297			
Time:	18:52:00	BIC	732.513			
Sample:	01-01-2016	HQIC	726.974			
	- 03-01-2022					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.4910	0.569	-0.863	0.388	-1.606	0.624
ar.L2	-0.0782	0.203	-0.385	0.700	-0.476	0.319
ma.L1	0.7843	0.349	2.248	0.025	0.100	1.468
sigma2	918.9099	76.325	12.039	0.000	769.316	1068.504
Ljung-Box (L1) (Q):	0.07	Jarque-Bera (JB):	508.90			
Prob(Q):	0.80	Prob(JB):	0.00			
Heteroskedasticity (H):	15.52	Skew:	2.37			
Prob(H) (two-sided):	0.00	Kurtosis:	14.94			

Фигура 4.30 показва месечните резултати относно данните за прогнозираните цени на електроенергията със SARIMA въз основа на историческите цени и от там се вижда, че

прогнозираната цена следва плътно реалната стойност. На фигура 4.30 е направено сравнение на тези стойности със средната стойност на тестовите серии, за да се провери дали големината на грешката е приемлива. Може да се види, че прогнозирането с помощта на SARIMA дава добри резултати, тъй като данните показват сезонност.



Фигура 4.30. Приложен модел SARIMA за месечни прогнози, проверка на резултатите между данни от реални цени и данни от прогнозиран цени

4.3.4. Изводи от направените изследвания

Прогнозните резултати за цените на електроенергията с помощта на ARIMA не се характеризират с нужното качество, тъй като времевият ред показва сезонност. Получените прогнозни данни за цените на електроенергията показват, че прогнозирането с помощта на SARIMA дава добри резултати. Идентифицираният и внедрен модел SARIMA е подходящ метод за прогнозиране за променливия характер на цените на електроенергията. Методът за оценка на грешката се определя в зависимост от свойствата на данните и следователно се сравняват отделните методи за прогнозиране.

4.4. Емпиричен анализ на времевите серии на цените на електроенергия чрез рекурентна невронна мрежа с дългосрочна-краткосрочна памет - LSTM

4.4.1. Описание на входните данни и анализ на почасови и дневни цени

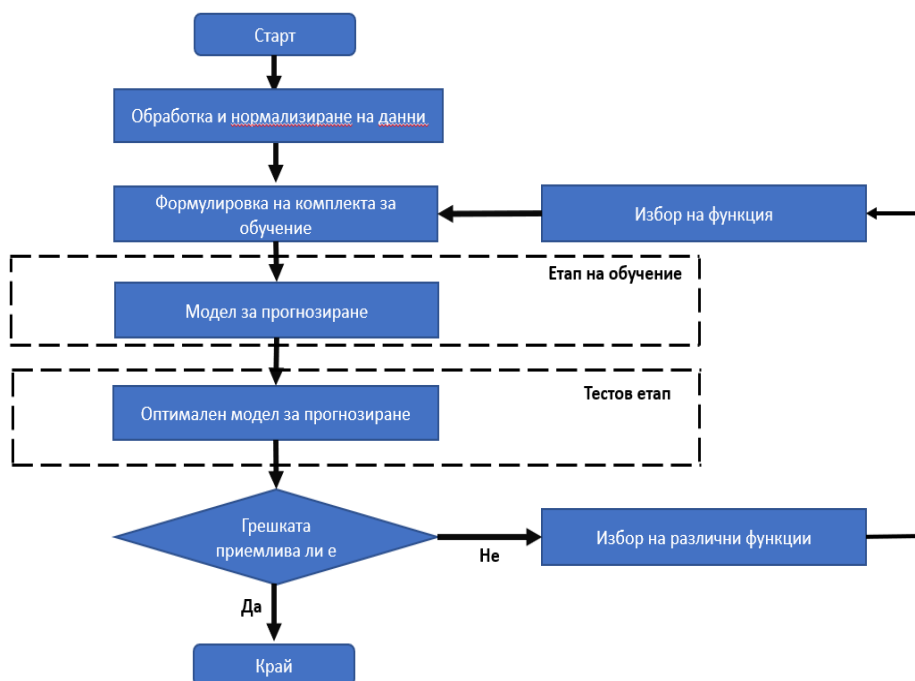
Историческите данни с които е направен анализ представляват почасов набор от данни под формата на времеви серии в размер на прозорец от общо 41 690 часа за периода от 20.02.2016 г. до 05.03.2022 г. Цените на електроенергията следват модели за дневни, седмични и сезонни времеви рамки, а графикът на прогнозните серии е на почасова и дневна база. Целта е да се предвидят бъдещите цени на електроенергията по часове и усреднени по дни, тъй като базовият модел използва исторически цени за прогнозиране на бъдещи стойности. Обучените модели ще бъдат различни от другите в смисъл, че ще бъдат обучени чрез различни хиперпараметри. След получаването на входната информация, данните са предварително обработени и структурирани и след което е създаден моделът LSTM.

Извлечените данни се съхраняват в 2D масив – низове за по-нататъшна обработка. Мащабирането на характеристиките е най-важната част от предварителната обработка на данните и резултатите от мащабирането на почасова база са показани на фигура 4.31.



Фигура 4.31. Мащабиране на почасови данни

Използвана е функцията MinMaxScaler(), за да се мащабират всички стойности в интервала между 0 и 1. Моделът се създава и обучава и се правят прогнози. Линеините графики на всички модели се начертават и се следи тяхната точност. Данните за структуриране и процесите по прогнозиране на цените са показани на блок схемата на фигура 4.32.



Фигура 4.32. Процес на прогнозиране на цените

LSTM приема серията от данни на цени на електроенергията като вход. Продължителността на серията е еквивалентна на броя на часовете. X_{train} е входът за набора за обучение, а y_{train} е изходът за набора за обучение. Подобна структура от данни се създава за изготвяне на набор от данни за прогнозиране. Наборите от данни ще бъдат структурирани за целите на обучението на почасова и дневна база. Моделът използва предишните n набора от данни като решаващ фактор и предвижда цената за стойността $n+1$. За създаване на LSTM моделът са импортирани някои модули като последователен модул за инициализиране на невронната мрежа, плътен модул за добавяне на плътно свързан слой невронна мрежа, слой

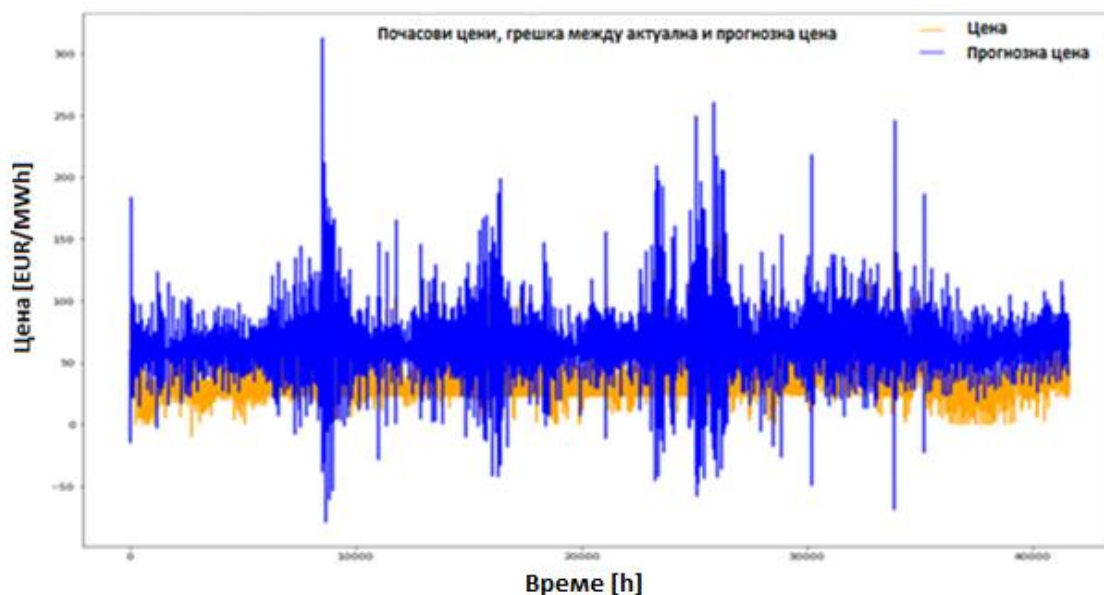
Long Short-Term Memory и модул за отпадане или за добавяне на отпадащи слоеве, които предотвратяват пренастройването. Най-важните хиперпараметри на LSTM са следните:

- Единици=96: Това означава, че се създава слой с десет неврона в него. Всеки от тези неврони получава стойности на входовете.
- Input_shape: Входът, очакван от LSTM, е в 3D формат. Данните за обучение имат формата на (1689, 50, 1) за дневни стойности и (41640, 50, 1) за почасови стойности, това е под формата на (брой проби, времеви стъпки, брой функции). Това означава, че се вземат 1689 примера, които ще бъдат научени в данните за обучение, всеки пример гледа назад 50 стъпки във времето, например каква е била цената вчера, завчера и така нататък до последните 50 дни. Те са известни като времеви стъпки. Последното число „1“ представлява броя на функциите. Тук се използва само една колона „Цена“, следователно тя е равна на „1“
- Kernel_initializer: Когато невроните започнат своето изчисление, алгоритъмът определя стойността за всяко тегло.
- Return_sequences = True: LSTM се разпространяват обратно във времето, следователно те връщат стойностите на изхода от всяка времева стъпка към следващия скрит слой. Това запазва очаквания вход на следващия скрит слой в 3D формат. Този параметър е False за последния скрит слой, защото не трябва да връща 3D изход към крайния плътен слой.
- Оптимизатор: Този параметър помага да се намерят оптималните стойности на всяко тегло в невронната мрежа.
- Batch_size = 32. Това определя колко реда ще бъдат предадени на мрежата наведнъж, след което изчислението ще започне и невронната мрежа ще започне да коригира своите тегла въз основа на грешките.
- Когато всички редове се предават в партии от по 50 реда всеки, както е посочено в този параметър, тогава наричаме този етап 1-епоха или един пълен цикъл на данни. Това също е известно като мини-партидно градиентно спускане. Малка стойност на batch_size ще накара LSTM да разглежда данните бавно, като 2 реда наведнъж или 4 реда наведнъж, което може да доведе до пренастройване, в сравнение с голяма стойност като 20 или 50 реда наведнъж, което ще накара LSTM да разглежда данните бързо, което може да доведе до недостатъчно оборудване. Следователно правилната стойност трябва да бъде избрана чрез настройка на хиперпараметър.
- Епохи=50. Същата дейност по коригиране на теглата продължава 50 пъти, както е посочено от този параметър. Казано по-просто, LSTM преглежда пълните данни за обучението 50 пъти и коригира теглата.

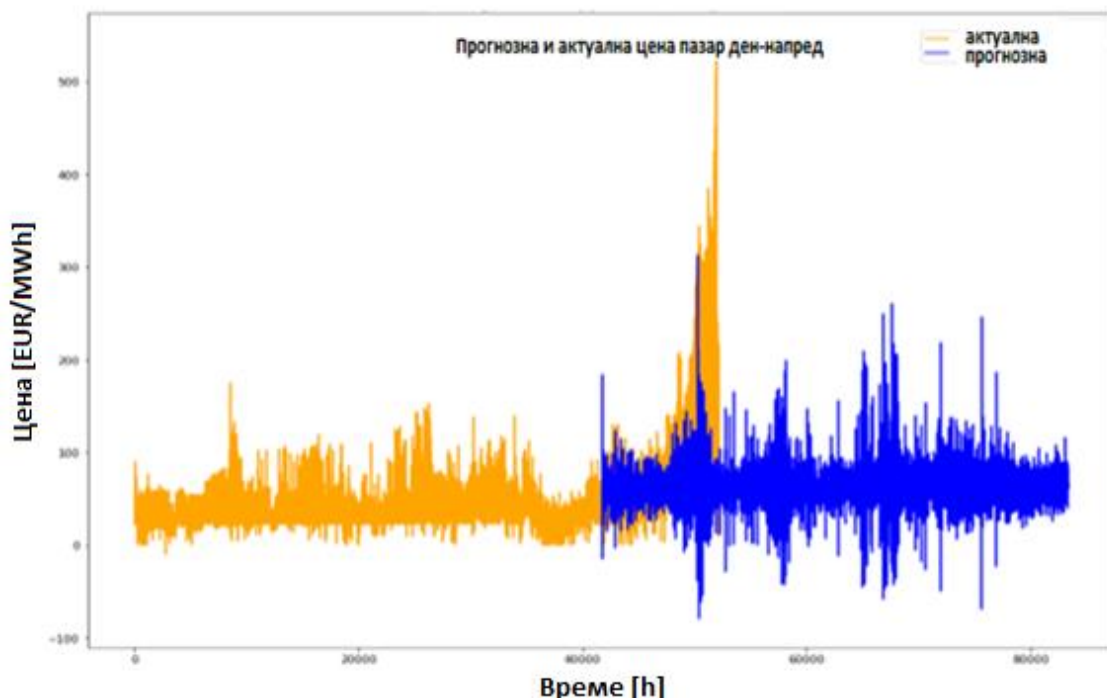
4.4.2. Резултати от изследванията

След обучение на модела се прави почасова и дневна прогноза. За тази цел се създава набор от тестови данни, който съдържа цени за минали часове и дни, така че измерването на точността на модела на почасови и дневни грешки в данните може да бъде проверено. Наборът от данни е структуриран по същия начин, както в стъпката за структуриране на данни. Тези стойности се подават една по една към входния слой LSTM и изходът се събира и съхранява в 2D масив за начертаване на диаграма и анализ на точността. Фигури 4.33. и 4.34. показват, че моделът като цяло може да предвиди правилната посока точно поради постоянната сезонност. Обучението по диференциране, а не по големина води до по-висока

точност и по-добра способност за генериране на потенциално подходящ модел. Използвана е средноквадратичната грешка MSE за да се провери колко близки са оценките или прогнозите до действителните стойности. Колкото по-ниска е MSE, толкова по-близо е прогнозата до действителната. За данните на фигура 4.33 определената MSE е 6.4.



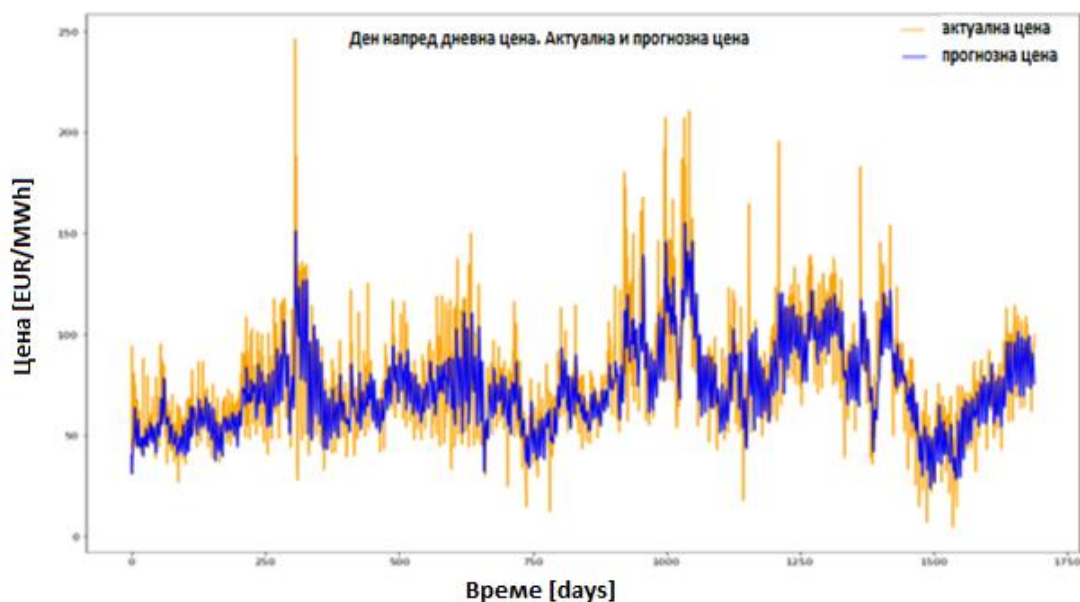
Фигура 4.33. Измерване на точността на модела – почасова проверка на грешките в данните



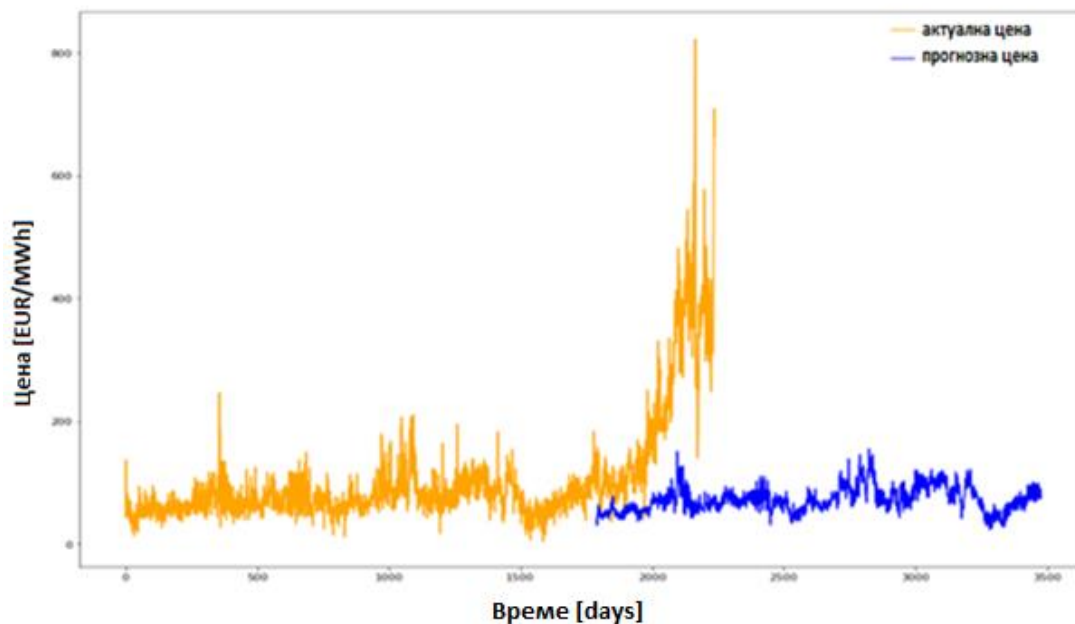
Фигура 4.34 Експериментално прогнозиране на модела върху почасови данни

От фигура 4.35 и фигура 4.36. можем да видим експерименталната прогноза на модела на почасови и дневни данни, съответно прогнозирани с LSTM. По-добри стойности могат да се видят при модела на часовите данни. От друга страна, от направените изследвания е видимо, че LSTM не е в състояние да предвиди правилни цени и да даде добри резултати с моделирани данни на дневна база. На фигура 4.36 има някои пикове на нарастващи и спадащи цени в действителната цена на реалните цени, които не са съпоставени с модела с

прогнозираната крива на данните. Изглежда, че това се дължи на неточно прогнозирани данни. Няколко некоректни части данни в набора за обучение или тестовия набор не могат да повлияят до голяма степен на модела за прогнозиране или да имат само ефект, който може да бъде игнориран. След извършване на различни симулации с различен брой параметри и епохи, се наблюдава, че с обучение се постигат малко по-добри резултати. За данните на фигура 4.35 определената MSE е 17.1. Проведените изследвания спрямо средно квадратичната грешка показват, по-добра точност на прогнозиране с LSTM модела при измерване върху почасови данни (срещу дневни данни).



Фигура 4.35. Изследване дали моделът е подходящ, чрез сравняване на база дневни данни



Фигура 4.36. Експериментално прогнозиране на модела върху дневни данни

4.4.3. Изводи от направените изследвания

Резултатите показват, че прогнозирането с помощта на метода LSTM може да бъде точно само за краткосрочна употреба, т.е. за прогнозиране на ценните за кратки интервали или почасови данни. Резултатите от работата с данните от почасовите клирингови ценни дават добри резултати. От друга страна, за прогнозирането с дневните данни, като 30 дни или 50 дни напред, методът показва, че е неуспешен. Незадоволителните резултати не са защото моделът LSTM е не е добър, а защото пазарите на електроенергия са силно нестабилни.

4.5. Получени резултати чрез различни модели

В Дисертацията се изследват 4 метода, които подпомагат процеса на прогнозиране чрез моделиране и анализ на времевите редове от почасови цени на електрическата енергия. Тук попадат методите за изчисляване на експонентата на Хърст чрез R/S анализ, DFA метод, ARIMA метод и метод на рекурентна невронна мрежа с дългосрочна-краткосрочна памет.

С помощта на програмната среда MATLAB, включвайки почасовите данни от борсовите цени през 2018 г., за България е направено изчисление на стойността на експонента на Хърст **чрез R/S анализ**, която е 0.7494 и се намира в интервала между 0.5 и 1, което означава, че времевите редове имат устойчиво поведение и запазват съществуващата тенденция. Същото може да се каже и за съседните на България пазари, Сърбия, Унгария, Румъния и Гърция, за които е направено същото изследване

DFA представлява разширение на обикновения флукуационен анализ, включвайки нестационарни процеси. С помощта на програмната среда MATLAB, е направено изчисление на стойността на експонентата на Хърст чрез метод на флукуационен анализ с редуциране на тренда. Резултатите от мащабиращата степен на корелация α количествено определя корелационните свойства на ценовите сигнали. Установено е, че стойността на експонента е близо до 1 за периода от 2016 -2019 г. на всички серии от почасови данни с малка вариация през 2017 г. Пазарното поведение показва силна дългосрочна корелация, което означава, че колкото по-висока е стойността на α , толкова по-силни са корелациите в сигнала и че корелациите, присъстващи във времевите редове, са по-големи колебания в по-дълги времеви мащаби. Данните в таблица 4.3 за α показват съответствие на спектъра на мощността и методите на DFA.

Моделирането на времевите редове с **ARIMA метод** е различно от по-традиционните подходи за класификация и регресионно прогнозно моделиране. Времеите редове са анализирани за 3 компонента: тенденция, сезонност и шум. Тъй като прогнозите с помощта на модела ARIMA спрямо действителните данни за целия период на прогнозиране не дават добри резултати, бива внедрен модела Seasonal-ARIMA. Този модел е използван, тъй като времевият ред показва сезонност. Добавени са няколко нови параметри, за да бъдат отчетени сезоните. За програмиране на създадените алгоритми за определяне на прогнозните стойности е използван програмния език Python. Прогнозирането с помощта на SARIMA дава добри резултати, тъй като данните показват сезонност.

По отношение на получените резултати с **модела LSTM**, след обучение на модела се прави почасова и дневна прогноза. За тази цел се създава набор от тестови данни, който съдържа цени за минали часове и дни, така че измерването на точността на модела на почасови и дневни грешки в данните може да бъде проверено. Наборът от данни е структуриран по същия начин, както в стъпката за структуриране на данни. Тези стойности се подават една

по една към входния слой LSTM и изходът се събира и съхранява в 2D масив за начертаване на диаграма и анализ на точността. Моделът с почасови данни може да предвиди правилната посока точно поради постоянната сезонност. Той демонстрира, че обучението по диференциране, а не по големина води до по-висока точност и по-добра способност за генериране на потенциално водещ модел.

4.6. Изводи от направените изследвания

Прогнозата за пазарната стойност е от голямо значение за максимизиране на печалбите от покупките и продажбите на електроенергия, като същевременно поддържа нисък риск. На база на получените резултати е направена оценка на грешките в зависимост от свойствата на данните и времевите хоризонти за прогнозиране чрез сравняване на отделните методи за прогнозиране, както е показано на фигура 4.37.



Фигура 4.37. Сравнителен анализ и резултати от приложените методи с препоръки за ползване на всеки един при различните времеви хоризонти на прогнозиране.

Въз основа на проведените изследвания са направени следните изводи:

1. Точните прогнози са от съществено значение за конкурентното предимство и вземането на инвестиционни решения в енергийния сектор.
2. Методите на Хърст, включително експонентата на Хърст, предоставят важна информация за тенденциите и шума във времевите редове, като се използват успешно при анализа на цените на електроенергията.
3. DFA се оказва по-подходящ метод от R/S анализа за прогнозиране на цените на електроенергията, особено за данни с краткосрочни зависимости и флуктуации.
4. ARIMA моделът среща затруднения при сезонни времеви редове и не достига до добри резултати за дългосрочни прогнози.
5. LSTM моделът предоставя по-точни прогнози, особено за краткосрочни времеви редове и почасови данни, но не постига добри резултати при по-дългосрочни прогнози.
6. SARIMA моделът се предпочита за дългосрочни и сезонни времеви редове.
7. Изборът на подходящ модел зависи от характеристиките и времевите периоди на данните, и добре проведенният анализ и моделиране предоставят ценна информация за динамиката и свойствата на цените на електроенергията, които могат да бъдат използвани за прогнозиране и вземане на решения в енергийния сектор.

Заклучение

В дисертацията са разработени ефективни модели за прогнозиране на цените на електроенергията, предлагайки нови стратегии за използване на методите за прогнозиране в зависимост от желания времеви хоризонт за изготвяне на прогнози с различни времеви рамки. Изследването подчертава важността на математическите методи в анализа на енергийните пазари, показвайки, че те могат да осигурят ценна информация и помощ при вземането на информирани решения.

Основни предизвикателства и цели на методите за анализ на цените включват идентификацията на подходящи променливи, които да подпомогнат различни модели за обясняване на промените на пазара, въз основа на ограничен брой входни данни. Комбинирането на множество фактори може да увеличи нестабилността на европейските пазари за електроенергия.

Изследването е уникално, като се фокусира върху времеви редове на почасовите цени на електроенергията и представя методи за класификация на времевите редове като стационарни и нестационарни серии. Допълнително се предоставят нови възможности за измерване на дългосрочна и краткосрочна памет в цените на електроенергията, използвайки методи като DFA и други.

Дисертацията показва, че различни математически модели, включително оптимизационни, стимулационни и статистически модели, могат да бъдат използвани за анализ на поведението на енергийните пазари и за прогнозиране на цените, търсенето и предлагането на енергия. Също така, се подчертава полезността на методи за машинно обучение при идентификацията на модели и тенденции в енергийните данни.

Изследването предлага важни теоретични и практически приноси за областта на енергийната икономика. Резултатите подобряват разбирането на поведението на енергийните пазари и подчертават ролята на математическите методи в анализа на сложни системи. Практически, те могат да бъдат от полза за политици, енергийни компании и инвеститори, които трябва да вземат решения на основата на точни и надеждни прогнози.

Въпреки това, работата има свои ограничения, включително допусканията в моделите и ограниченията на данните. Бъдещи изследвания могат да се занимават с разширяване на методите и моделите, като използват по-обширни данни и да продължават да подобряват методите за анализ на енергийните пазари и прогнозиране на цените на електроенергията.

НАУЧНО-ПРИЛОЖНИ ПРИНОСИ

1. Създадена е методика за изследване, анализ и прогнозиране на цените на електроенергията, базирана на ARIMA, SARIMA и LSTM методи, които дават оптимални прогнози за цените на пазара на електроенергия в зависимост от различните входни фактори.
2. Създадена е методика за изследване и анализ на дългосрочната устойчивост на времевите серии от данни на борсови цени на електроенергия базирана на метода DFA, което не е правено досега.
3. Експериментално е изследвано приложението на статистически методи за определяне на експонентата на Хърст (R/S метод) върху цените на електроенергията. Изследванията показват максимална относителна грешка 6% при изчисляване на H , което показва приложимостта на този метод за проучване на фракталността на времевите серии на цените на електроенергията.
4. Експериментално е доказано, че DFA метода е по-подходящ при прогнозиране на продължителни данни за цената на електроенергията в сравнение с R/S метода. Направеният сравнителен анализ показва максимална относителна грешка при определяне на експонентата на Хърст 1.2% при прилагане на DFA срещу 6% при реализация с R/S статистически метод.
5. При анализа на краткотрайни данни е доказано преимущество на прогнозирането с помощта на SARIMA метода (в сравнение с ARIMA метода), тъй като променливия характер на цените на електроенергията показват сезонност.

ПРИЛОЖНИ ПРИНОСИ

1. Програмно реализирани и анализирани са алгоритми за анализ и прогнозиране на дългосрочни данни за цените на електроенергията на базата на прилагане на R/S метода за определяне на експонентата на Хърст и DFA метода.
2. Програмно реализирани и анализирани са алгоритми за анализ и прогнозиране на краткосрочни данни за цените на електроенергията на базата на прилагане на методите ARIMA, SARIMA и LSTM.
3. Създадени са демонстрационни процедури за анализ и прогнозиране на цените на енергетиката в зависимост от вида на зададени входни параметри.

Благодарности

С благодарност и вълнение изразявам безмерна признателност към всички онези, които ме водиха и подкрепяха по време на пътя към завършване на докторската ми дисертация.

Първо и най-важно, искам да благодаря на научния ми ръководител, доц. д-р инж. Галя Цанева, която ме насърчаваше в тежките моменти. Тя беше моят светлинен фар през този дълъг и изискващ път. Нейните съвети и подкрепа бяха ключови за всяка стъпка на този научен път. Без нейното търпение, помощ и насоки, докторската ми дисертация нямаше да бъде възможна.

Със съжаление споменавам доц. д-р Митко Господинов, който беше моят вдъхновител и научен ментор. Неговото научно наследство и труд ще останат завинаги в сърцата на всички, които го познаваха. Той беше и винаги ще бъде неразделна част от тази дисертация.

Освен тях, искам да изразя моята благодарност към доц. д-р Евгения Господинова, която се превърна във значим ментор и насърчител на научната ми кариера. Нейната безкрайна подкрепа и съвети бяха източник на сила и увереност за мен.

Огромни благодарности към ръководството на ИР-БАН в лицето на Академик Чавдар Руменин и професор д-р Август Иванов, които ми предоставиха възможността да изследвам в тази удивителна област. Специални благодарности и към проф. д-р Анна Лекова, проф. д-р Мая Димитрова и проф. д-р Румен Захариев, които ме подкрепяха и вдъхновяваха през целия този път.

Без вашата подкрепа и насоки, тази дисертация нямаше да бъде възможна. Благодаря ви от сърце за вашето безценно влияние. Този труд и постижение носи следите на вашата мъдрост и страст към науката.

С най-дълбока благодарност и обич, искам да изразя моето признание и към семейството ми - моят съпруг и родители за тяхното неизмеримо търпение и подкрепа през целия този научен път. Те понасяха моите дни и нощи, посветени на изследвания и анализи, с търпение и усмивка. Те ме насърчаваха в тежките моменти и бяха там, за да ми предоставят обич и подкрепа.

Посвещение

С неизмерима благодарност и обич, посвеждавам тази докторска дисертация на моите обичани родители, моите най-големи вдъхновители и подкрепители. Вие бяхте моите първи учители и пристанище на всеки етап от моят път.

На теб, татко, който ми даде крила да се изправя срещу всяка трудност и предизвикателство и непримиримо да се боря. Твоят силен дух и непрекъснатата ти подкрепа ме научиха ценността на упоритостта, настойчивостта и решителността.

И на теб, майко, която не дочака да ме видиш днес. Твоят дух и влияние са с мен винаги, продължават да бъдат моят източник на сила и остават вечно с мен да ме водят в бъдещето. Със сълзи в сърцето искам да ти благодаря, майко, за всичко, което ми даде, за подкрепата и обичта, която винаги ще нося в сърцето си. Този труд е твоят труд, и аз се надявам, че където и да си, че си горда с мен и с всичко, което постигнах.

Благодаря ви, че вярвахте в мен и ме подкрепяхте през всички времена.

С обич и благодарност,

Екатерина

СПИСЪК НА ПУБЛИКАЦИИ ПО ТЕМАТА НА ДИСЕРТАЦИОННИЯ ТРУД

1. **Popovska, E.**, Georgieva-Tsaneva, G. (2022). Day-ahead Electricity Price Forecasting using Long-short Term Memory Recurrent Neural Network, Science Series "Innovative STEM Education", vol. 4, ISSN: 2683-1333, Institute of Mathematics and Informatics – Bulgarian Academy of Sciences, pp. 139-148, DOI: <https://doi.org/10.55630/STEM.2022.0417>.
2. **Popovska, E.**, Georgieva-Tsaneva, G. (2022). ARIMA Model for Day-Ahead Electricity Market Price Forecasting, Science Series "Innovative STEM Education", vol. 4, ISSN: 2683-1333, Institute of Mathematics and Informatics – Bulgarian Academy of Sciences, pp. 149-161. DOI: <https://doi.org/10.55630/STEM.2022.0418>.
3. Gospodinov, M., Gospodinova, E. and **Popovska, E.** (2020). Comparative Analysis of Statistical Methods for Estimating Hurst Exponent. CompSysTech '20: Proceedings of the 21st International Conference on Computer Systems and Technologies '20, ACM International Conference Proceeding Series, 2020, ISBN:978-1-4503-7768-3, Pages 148–155, <https://doi.org/10.1145/3407982.3408012>, **sjr 0.182 (SCOPUS)**
4. **Popovska, E.**, Gospodinov, M. (2020). Fractal Behavior in Bulgarian Day-Ahead Prices Based on Detrended Fluctuation Analysis, Science Series "Innovative STEM Education", vol. 02, ISSN: 2683-1333, Institute of Mathematics and Informatics – Bulgarian Academy of Sciences, pp. 56-64, DOI: <https://doi.org/10.55630/STEM.2020.0207>.
5. **Popovska, E.** and Gospodinov, M. (2019). Computation of Time Series Scaling Exponent for Electricity Prices Forecasting. CompSysTech '19: Proceedings of the 20th International Conference on Computer Systems and Technologies, ACM International Conference Proceeding Series, June 2019, pp. 194–199, <https://doi.org/10.1145/3345252.3345267>, **sjr 0.2, Q4 (SCOPUS)**.
6. Gospodinov, M., **Popovska, E.** (2018). A brief overview of the contemporary and hybrid applied mathematical methods for electricity prices forecasting, Proceedings of the XXIII International Conference “Robotics&Mechatronics, and Social Implementations”, Varna, Bulgaria, 2018, Complex Control Systems, ISSN 1310-8255, pp. 59-63, http://ir.bas.bg/ccs/2018/11_gospodinov.pdf.

ЗАБЕЛЯЗАНИ ЦИТИРАНИЯ

Mitko Gospodinov, Evgeniya Gospodinova, and **Ekaterina Popovska**. 2020. Comparative Analysis of Statistical Methods for Estimating Hurst Exponent. Proceedings of the 21st International Conference on Computer Systems and Technologies '20. Association for Computing Machinery, New York, NY, USA, 148–155. <https://doi.org/10.1145/3407982.3408012>, SRJ=2.679.

Цитира се в:

1. Sweta, Kumari et al. "Performance assessment for three statistical models of landslide susceptibility zonation mapping: A case study for Dharamshala Region, Himachal Pradesh, India." *Journal of Earth System Science* 131 (2022): n. pag. [https://www.semanticscholar.org/paper/Performance-assessment-for-three-statistical-models-Sweta Goswami/ddca1821eabd023c0e36176bbf22db45c8b30fce#references](https://www.semanticscholar.org/paper/Performance-assessment-for-three-statistical-models-Sweta%20Goswami/ddca1821eabd023c0e36176bbf22db45c8b30fce#references)

Popovska, E., Gospodinov, M. (2019). Computation of time series scaling exponent for electricity prices forecasting. CompSysTech'19: Proceedings of the 20th International Conference on Computer Systems and Technologies, June 2019, pp. 194–199. <https://doi.org/10.1145/3345252.3345267>, SRJ=2.2.

Цитира се в:

1. Ferenc, Bakó (2022). Application of the visibility graph procedure to the power exchange processes of European countries for examination, Journal Article, University in Győr, Hungary, https://rgdi.sze.hu/images/RGDI/honlapelemei/fokozatszerzesi_anyagok/Doktori%20%C3%A9rtekez%C3%A9s.pdf
2. Gospodinova, E. (2022). Methods of Nonlinear Dynamics for Heart Rate Variability Analysis, Science Series "Innovative STEM Education", volume 04, ISSN: 2683-1333, Institute of Mathematics and Informatics – Bulgarian Academy of Sciences, 32-38., <http://www.math.bas.bg/vt/stemedu/book-4/STEMedu-2022-v.pdf>
3. Lebamovski, P. (2022). Analysis of Methods and Approaches for Evaluation of Heart Rate Variability, Science Series "Innovative STEM Education", volume 04, ISSN: 2683-1333, Institute of Mathematics and Informatics – Bulgarian Academy of Sciences, 39-47., <http://www.math.bas.bg/vt/stemedu/book-4/STEMedu-2022-vi.pdf>
4. Gospodinova, E. (2020). Application of Methods from Nonlinear Dynamics for Heart Rate Variability Analysis, Science Series "Innovative STEM Education", volume 02, ISSN: 2683-1333, Institute of Mathematics and Informatics – Bulgarian Academy of Sciences, 28-35., <http://www.math.bas.bg/vt/stemedu/book-2/04-STEMedu-2020.pdf>

Popovska, E., Gospodinov, M. (2020). Fractal Behavior in Bulgarian Day-Ahead Prices Based on Detrended Fluctuation Analysis. In processing: Scientific conference with international participation STEMEDU-2020, Veliko Tarnovo, pp. 56-64.

Цитира се в:

1. Gospodinova, E. (2022). Methods of Nonlinear Dynamics for Heart Rate Variability Analysis, Science Series "Innovative STEM Education", volume 04, ISSN: 2683-1333, Institute of Mathematics and Informatics – Bulgarian Academy of Sciences, 32-38., <http://www.math.bas.bg/vt/stemedu/book-4/STEMedu-2022-v.pdf>

2. Lebamovski, P. (2022). Analysis of Methods and Approaches for Evaluation of Heart Rate Variability, Science Series "Innovative STEM Education", volume 04, ISSN: 2683-1333, Institute of Mathematics and Informatics – Bulgarian Academy of Sciences, 39-47., <http://www.math.bas.bg/vt/stemedu/book-4/STEMedu-2022-vi.pdf>

Mathematical Methods for Research Modeling, Analysis, and Forecasting in the Energy Sector and Energy Markets

Summary

The doctoral dissertation presents research on methods and models for forecasting electricity prices, with a focus on the Bulgarian energy market and a comparison with other neighboring markets. It examines various aspects of time series analysis of hourly electricity price data, applying methods such as Hurst R/S exponent and DFA parameter calculation, ARIMA, and LSTM models. The work identifies key factors influencing electricity prices and demonstrates forecasting methods depending on the time horizons as hours, days, months and years.

The results emphasize the importance of accurate forecasts for competitiveness and investment decisions in the energy sector. Through comparative analyses, the dissertation delves into the strengths and limitations of each forecasting method, shedding light on their effectiveness under various conditions.

The Hurst methods are demonstrated to offer valuable insights into time series data. They assist in the discernment of trends and noise in price data, thus being invaluable for in-depth price analysis. Specifically, the Detrended Fluctuation Analysis (DFA) emerges as a more suitable approach compared to the R/S analysis when forecasting electricity prices, particularly for data exhibiting short-term dependencies and fluctuations. In contrast, the ARIMA model encounters challenges when confronted with seasonal time series data, rendering it inadequate for accurate long-term predictions. However, LSTM model enhances forecasting accuracy, particularly for short-term time series and hourly data. It exhibits certain limitations in long-term forecasting. For long-term and seasonal time series, the SARIMA model proves to be the most appropriate choice. It excels in capturing seasonal patterns in electricity price data.

Based on the findings of the research, the Hurst and DFA methods provide valuable information on trends and noise in time series data, while ARIMA and LSTM are applied for price prediction. Comparative analysis of the methods shows different advantages and limitations depending on time periods and data characteristics.

In conclusion, the research is of significance both in theoretical and practical terms, providing tools for making informed decisions in the energy sector. The findings can be useful for electricity traders, consumers, generators, investors, and policymakers involved in energy markets who need to manage risk and make strategic decisions. However, the research has limitations, and future studies can expand and improve the methods and models for electricity price forecasting.