**Academia de Studii Economice din București**

**Facultatea de Cibernetică, Statistică și Informatică Economică**

**Specializarea: Cibernetică Economică**

****

**Proiect de cercetare**

**SERII DE TIMP**

**Analiza Seriilor de Timp în Industria Modei: Studiu de Caz asupra Performanței Financiare a Inditex**

**Cadre didactice coordonatoare:**

**Prof. univ. dr. DAVIDESCU ADRIANA-ANAMARIA**

**Cadru asociat drd. MANTA EDUARD-MIHAI**

**Student : Stafie Diana Flavia**

**Anul III, Grupa 1081, Seria B**

CUPRINS

[INTRODUCERE 3](#_Toc166459915)

[Descrierea companiei Inditex 3](#_Toc166459916)

[Motivația alegerii 4](#_Toc166459917)

[Problema concretă 4](#_Toc166459918)

[APLICAȚIA 1 5](#_Toc166459919)

[ARIMA 5](#_Toc166459920)

[Literatură de specialitate 5](#_Toc166459921)

[Metodologia cercetării 8](#_Toc166459922)

[Descrierea datelor 9](#_Toc166459923)

[Modelarea unei serii de timp utilizând metodologia Box-Jenkins. Previziuni. 10](#_Toc166459924)

[Cerință 10](#_Toc166459925)

[Încărcarea setului de date 10](#_Toc166459926)

[Graficul seriei și corelograma 11](#_Toc166459927)

[Testul Canova și Hansen 12](#_Toc166459928)

[Staționaritatea 13](#_Toc166459929)

[Akaike Information Criterion, Testul Dickey Fuller și Augmented Dickey Fuller 14](#_Toc166459930)

[Definirea setului de rentabilități 16](#_Toc166459931)

[Evaluarea staționarității seriei de rentabilități 17](#_Toc166459932)

[Testul Philips Peron și Kwiatkowski-Philips-Schmidt-Shin 19](#_Toc166459933)

[Testul HEGY 20](#_Toc166459934)

[Diferențierea a doua a seriei 21](#_Toc166459935)

[Metoda ARIMA 22](#_Toc166459936)

[Diagnostic reziduuri 23](#_Toc166459937)

[SARIMA 27](#_Toc166459938)

[Literatură de specialitate 27](#_Toc166459939)

[Metodologia cercetării 30](#_Toc166459940)

[Descrierea datelor 30](#_Toc166459941)

[Motivarea utilizării modelului SARIMA 30](#_Toc166459942)

[UTILIZAREA MODELULUI SARIMA 31](#_Toc166459943)

[Estimarea parametrilor optimi 31](#_Toc166459944)

[Ecuația modelului SARIMA 33](#_Toc166459945)

[Graficul și corelograma reziduurilor 34](#_Toc166459946)

[MODELAREA ARCH 36](#_Toc166459947)

[BIBLIOGRAFIE 36](#_Toc166459948)

# INTRODUCERE

În era actuală a globalizării și a schimbărilor continue din industria modei, analiza datelor și înțelegerea evoluției acestora devin cruciale pentru succesul și durabilitatea companiilor din domeniu. În acest context, analiza seriilor de timp devine o unealtă esențială pentru înțelegerea și anticiparea tendințelor, modelarea comportamentului pieței și luarea deciziilor strategice într-o manieră informată.

Proiectul nostru se axează pe aplicarea conceptelor și tehnicilor de analiză a seriilor de timp în industria modei, cu accent pe studiul de caz al companiei Inditex. Inditex reprezintă unul dintre cei mai importanți jucători din industria modei la nivel global, fiind cunoscut pentru brandurile sale de renume precum Zara, Massimo Dutti, Bershka sau Pull&Bear. Într-un mediu competitiv și în continuă schimbare, înțelegerea evoluției indicatorilor cheie precum vânzările, stocurile sau performanța financiară devine esențială pentru adaptarea la cerințele pieței și menținerea unui avantaj competitiv.

Prin acest proiect, ne propunem să explorăm și să aplicăm diverse concepte și tehnici de analiză a seriilor de timp, pornind de la modele cu trend determinist sau stochastic, până la modele avansate de predicție și evaluare a performanței. Înțelegerea și aplicarea acestor concepte ne va permite să investigăm și să interpretăm evoluția și relațiile dintre diferiții indicatori financiari și operaționali ai Inditex, contribuind astfel la elaborarea unei analize comprehensive a performanței și perspectivei companiei.

Prin intermediul acestui proiect, ne propunem să evidențiem importanța analizei seriilor de timp în contextul industriei modei și să oferim o abordare practică și aplicată asupra utilizării acestor concepte în cadrul unei companii reprezentative precum Inditex. Rezultatele obținute vor putea servi drept bază pentru luarea deciziilor strategice și pentru identificarea oportunităților de optimizare și creștere într-un mediu extrem de dinamic și competitiv.

## Descrierea companiei Inditex

Inditex este unul dintre cei mai importanți și influenți retaileri din industria modei la nivel mondial. Fondată în 1985 în Spania de către Amancio Ortega, compania s-a dezvoltat rapid și a devenit cunoscută pentru abordarea sa inovatoare și agilă în domeniul designului, producției și distribuției de îmbrăcăminte și accesorii. Inditex operează o serie de branduri de renume global, printre care se numără Zara, Massimo Dutti, Bershka, Pull&Bear, Stradivarius, Oysho și Uterque. Fiecare brand are propriul său concept și identitate distincte, adresându-se diferitelor segmente de piață și preferințelor consumatorilor.

De-a lungul anilor, Inditex și-a consolidat poziția ca unul dintre liderii mondiali în domeniul modei prin strategia sa de expansiune globală, inovație în design și producție, precum și prin adaptabilitatea sa la schimbările din piață și la evoluția tendințelor de consum. Compania se distinge prin modelul său de afaceri integrat vertical, care îi permite să controleze întregul lanț de aprovizionare, de la design și producție, până la distribuție și vânzare în magazinele proprii. Această integrare a lanțului de aprovizionare contribuie la flexibilitatea și eficiența operațională a companiei, permițându-i să răspundă rapid la cerințele pieței și să ofere produse la prețuri competitive.

Inditex se remarcă și prin angajamentul său față de sustenabilitate și responsabilitate socială, implementând o serie de inițiative și programe pentru reducerea impactului său asupra mediului și pentru îmbunătățirea condițiilor de muncă în întreaga sa rețea de producție. Prin portofoliul său diversificat de branduri, abordarea sa inovatoare și capacitatea sa de a se adapta la schimbările din industria modei, Inditex continuă să rămână unul dintre cei mai puternici și influenți jucători din industrie, consolidându-și poziția sa ca lider global în moda contemporană.

## Motivația alegerii

Alegerea companiei Inditex ca subiect pentru proiectul nostru se datorează importanței și relevanței sale în industria modei la nivel mondial. Inditex reprezintă nu doar unul dintre cei mai importanți retaileri din lume, ci și un pionier în inovația și adaptabilitatea într-un mediu comercial extrem de dinamic și competitiv. Portofoliul său diversificat de branduri și abordarea sa inovatoare în design, producție și distribuție au făcut din Inditex un lider în industria modei contemporane.

Prin alegerea Inditex ca subiect pentru analiza noastră a seriilor de timp, ne propunem să investigăm și să înțelegem mai profund evoluția și performanța uneia dintre cele mai influente companii din domeniul modei. Studiul nostru va oferi oportunitatea de a explora modul în care Inditex gestionează provocările și oportunitățile din industria modei prin intermediul analizei seriei de timp, oferind insight-uri valoroase pentru practicile și strategiile viitoare în cadrul companiei și, în general, în industrie.

## Problema concretă

Se dorește să se previzioneze evoluția prețurilor acțiunilor pentru compania Inditex, unul dintre principalii jucători din industria modei, cu prezență globală. Prin analiza seriilor de timp, ne propunem să investigăm modelele și tendințele care influențează prețurile acțiunilor Inditex, oferind astfel insight-uri valoroase pentru investitori și factorii de decizie din domeniu. În plus, ne propunem să analizăm relația dintre evoluția prețurilor acțiunilor Inditex și indicele bursier relevant, oferind o perspectivă amplă asupra performanței companiei în contextul pieței financiare globale.

# APLICAȚIA 1

# ARIMA

## Literatură de specialitate

Prognozarea tendințelor pieței de valori este considerată o sarcină important și primește o atenție deosebită, deoarece prognozele de success în ceea ce privește prețul acțiunilor pot aduce profituri semnificative atât indivizilor, cât și comunităților, fiind considerate un factor important legat de întreaga societate economică. Prognozarea pieței de valori reprezintă o provocare majoră din cauza datelor non-staționare, zgomotului și haosului, ceea ce face ca investitorii să aibă dificultăți în a-și investi banii pentru a obține profituri.

Prețurile acțiunilor nu sunt valori generate aleatoriu, ci pot fi tratate ca un model de serie temporală discretă și trendul lor poate fi analizat corespunzător, putând fi de asemenea prognozate. Din necesitatea identificării unui model pentru a analiza tendințele prețurilor acțiunilor cu informații relevante pentru luarea deciziilor, se recomandă transformarea seriei temporale folosind ARIMA ca fiind o abordare mai bună decât prognozarea directă, deoarece oferă rezultate mai precise. (Prapanna Mondal, Labani Shitsi Saptarsi Goswami, 2014)

Multe cercetări au folosit modelul ARIMA pentru prognoze. Autorii au combinat ARIMA cu rețele neuronale pentru a prezice cererea de mărfuri în Chile, au comparat variante ale modelului ARIMA și au implementat tehnici hibride pentru prognoza prețurilor acțiunilor. De asemenea, s-au realizat prognoze pentru indicii de piață și prețul acțiunilor folosind diverse metode, inclusiv algoritmi genetici și tehnici de optimizare. Toate aceste cercetări au ca scop îmbunătățirea preciziei prognozelor pieței de valori pentru a ajuta investitorii în deciziile lor de investiții.

În anul 2020 a fost efectuat un studiu intitulat "ARIMA Model for Accurate Time Series Stocks Forecasting" de Shakir Khan și Hela Alghulaiakh, care a fost realizat pe tema utilizării modelului ARIMA pentru prognozarea precisă a seriilor temporale de acțiuni, în special folosind datele istorice ale acțiunilor Netflix pentru ultimii cinci ani. Autorii compară rezultatele obținute cu modelul auto ARIMA și două modele personalizate ARIMA pentru a obține o mai bună predicție a valorilor acțiunilor. Procesul de prognozare se bazează pe valorile de închidere ajustate ale acțiunilor companiei Netflix, care sunt scalate pentru a obține citiri mai precise. Pentru a evalua acuratețea modelului, se calculează funcțiile de autocorelație (ACFs) și funcția de autocorelație parțială (PACF), iar rezultatele sunt comparate în mai multe experimente. Această analiză metodologică este importantă pentru a înțelege eficacitatea și precizia modelului ARIMA în prognozarea valorilor acțiunilor Netflix.

Rezultatele studiului au demonstrat că modelul ARIMA (1,1,33) a oferit cea mai mare precizie în predicția valorilor acțiunilor Netflix pe baza datelor istorice. Deși diferența în precizie între acest model și alte variante ARIMA a fost minimă, ARIMA (1,1,33) s-a evidențiat ca fiind cea mai bună opțiune, având în vedere rezultatele testului de reținere. Prognoza realizată pe baza acestui model a arătat o continuitate în valorile acțiunilor Netflix pentru următoarele trei luni. Astfel, concluzia proiectului este că modelul ARIMA poate oferi predicții precise pentru valorile acțiunilor pe baza datelor istorice, fiind util pentru investitorii în acțiuni în luarea deciziilor de investiții. Însă, studiul a analizat doar performanța modelului ARIMA în predicția valorilor acțiunilor Netflix și este menționabil că pentru o mai mare precizie în viitor, cercetătorii ar trebui să extindă analiza și să compare mai multe modele, cum ar fi GARCH, LSTM sau modele de învățare automată.

Un alt studiu realizat în 2021 intitulat “On the accuracy of ARIMA based prediction of COVID-19 spread”, având drept autori pe H. Alabdulrazzaq, Mohammed N. Alenezi, Yasmeen Rawajfihsialtii, investighează precizia modelului ARIMA în prezicerea răspândirii COVID-19 în Kuwait pe o perioadă extinsă de timp. Astfel, prin optimizarea parametrilor și folosirea datelor raportate, autorii validează precizia modelului și constată că predicțiile ARIMA sunt adecvate și satisfăcătoare. Concluzia acestui studiu este că, deși modelul ARIMA nu este considerat potrivit pentru situații complexe și dinamice, precum pandemiile, acesta poate oferi predicții rezonabile și satisfăcătoare în ceea ce privește numărul de cazuri confirmate și recuperate de COVID-19. Așa cum am menționat anterior, autorii au testat și validat modelul ARIMA folosind datele din Kuwait pe o perioadă extinsă, în timp ce țara implementa un plan preventiv gradual în cinci faze. Chiar și cu revizii constante ale planului și schimbări continue, modelul ARIMA a demonstrat o corelație foarte ridicată între predicții și valorile reale raportate, indicând o acuratețe adecvată în prezicerea răspândirii virusului.

Anul acesta, în 2024, Ziyue Yu a realizat o lucrare intitulată "Stock Price Prediction using the ARIMA Model". Această lucrare își concentrează atenția pe metodele de prognozare a prețului acțiunilor, cu accent pe utilizarea modelului ARIMA. Autorul subliniază importanța prognozării în afaceri și investiții, evidențiind rolul crucial al prognozelor de vânzări și al prognozării prețurilor acțiunilor în formularea unor planuri financiare mai bune, alocarea resurselor și luarea deciziilor. Se compară metodele statistice și tehnici de calcul, evidențiind avantajele și dezavantajele fiecăreia. Articolul se concentrează pe modelul ARIMA și își propune să demonstreze utilizarea acestuia în prognozarea prețurilor acțiunilor folosind date reale și să discute despre potențialele sale avantaje și limite.

Concluziile proiectului lui Ziyue Yu demonstrează că modelul ARIMA poate fi util în prognozarea prețurilor acțiunilor în piețele financiare. Alegerea și structurarea unui model ARIMA pot oferi o direcție generală în ceea ce privește fluctuațiile pieței de acțiuni, ceea ce poate ajuta investitorii să ia decizii mai informate în timpul achiziționării. Cu toate acestea, este important să se recunoască că modelul ARIMA nu este lipsit de erori și că piețele financiare sunt mult mai complexe decât ceea ce pot prezice modelele. În cazul specific al băncii Ping An, se observă că prețul real al acțiunilor experimentează o tendință mai turbulentă decât cea prevăzută de model. În plus, modelul ARIMA nu își dovedește exactitatea în prognozarea datelor pe termen scurt. Aceste observații sugerează că este necesară continuarea cercetărilor pentru a reduce incertitudinile asociate utilizării modelului ARIMA în prognozarea prețurilor acțiunilor, astfel încât să se asigure o decizie mai sigură pentru investitori.

O altă cercetare, “ARIMA Models in Solar Radiation Forecasting in Different Geographic Locations” realizată de Ewa Chodakowska, Joanicjusz Nazarko și alții în 2023 își propune să evalueze performanța modelelor ARIMA în prognozarea radiației solare în diferite condiții climatice, utilizând date din Iordania și Polonia. Aceste țări au un interes crescut în creșterea utilizării energiei regenerabile, iar proiectul urmărește să analizeze eficacitatea modelului ARIMA în acest context. Datele inițiale luate în considerare includ capacitatea instalată de energie fotovoltaică, datele de iradiere solară și tendințele climatice ale celor două țări.

Cercetarea raportată în acest articol indică utilitatea modelelor ARIMA pentru prognozarea radiației solare în diferite locații geografice caracterizate de condiții climatice diferite. Capitalele Iordaniei și Poloniei, pentru care nivelul mediu zilnic al GHI diferă de peste două ori, au fost selectate pentru analiză. Modele ARIMA adecvate au fost identificate pentru ambele locații, iar parametrii pentru intervalele orare și lunare au fost estimați. În cazul Ammanului, s-au obținut modele ARIMA(1,0,0)(0,1,1)24 și ARIMA(0,0,0)(0,1,1)12, respectiv, iar în cazul Varșoviei, s-au obținut modele ARIMA(1,0,0)(0,1,1)24 și ARIMA(0,0,0)(1,1,0)12. Modelele ARIMA identificate arată aceeași structură pentru datele orare în ambele locații și o structură diferită pentru datele lunare. Un astfel de rezultat indică necesitatea identificării unui model ARIMA adecvat pentru fiecare dată pentru locații cu condiții climatice diferite.

Potrivirea modelelor și capacitatea lor predictivă au fost, de asemenea, evaluate. Toate modelele au prezentat o potrivire foarte bună cu datele, măsurată prin eroarea standard a modelului și coeficientul de determinare R^2 (R^2 > 85%). Modelele arată o potrivire mai bună pentru datele orare în lunile de vară, în cazul atât al Ammanului, cât și al Varșoviei. Previziunile pe termen de un an realizate pentru 2021 pe baza datelor lunare arată o precizie ridicată, determinată de intervalul lor de încredere. În același timp, intervalul de încredere este mai îngust pentru previziunea din Amman. Diferențele în rezultatele observate indică faptul că modelele ARIMA sunt cele mai potrivite pentru prognozarea în condiții stabile (însorite sau înnorate). Precizia mai scăzută obținută în Varșovia poate fi un indiciu că, dacă se observă instabilități mai mari, astfel de modele nu reușesc să ofere previziuni satisfăcătoare. Rezultatele modelării și prognozării radiațiilor solare cu modele ARIMA prezentate în acest articol indică faptul că acestea sunt suficient de precise, în special pentru prognozele pe termen mediu și lung, și pot fi utilizate eficient pentru planificarea armonizării instalațiilor fotovoltaice cu sistemul electric în diferite locații geografice.

Importanța prognozării iradierii solare pentru exploatarea și gestionarea infrastructurii rețelelor electrice este, de asemenea, demnă de evidențiat. O previziune fiabilă asigură atât beneficii tehnice (de exemplu, îmbunătățirea dependabilității sistemului, calitatea corespunzătoare a energiei, stocarea energetică de rezervă mai mică), cât și beneficii economice (de exemplu, alinierea mai bună a ofertei și cererii, costuri mai mici de investiții și de operare, costuri mai mici de retribuție). În ciuda progreselor notabile, sunt necesare cercetări suplimentare în aceste direcții. Trebuie subliniat faptul că pentru fiecare locație geografică, este necesar să se identifice un model ARIMA potrivit, să se estimeze parametrii săi, să se evalueze potrivirea și să se verifice capacitatea de prognozare a modelului.

Astfel, cele patru studii analizate demonstrează eficacitatea și utilitatea modelului ARIMA în diferite domenii și contexte. Studiul lui Shakir Khan și Hela Alghulaiakh relevă capacitatea modelului ARIMA de a oferi prognoze precise ale valorilor acțiunilor Netflix, oferind astfel investitorilor informații utile pentru luarea deciziilor de investiții. De asemenea, lucrarea lui Haneen Alabdulrazzaq și colegii săi validează utilizarea modelului ARIMA în prezicerea răspândirii COVID-19 în Kuwait, subliniind acuratețea acestuia în ciuda condițiilor dinamice și complexe asociate pandemiei. În plus, studiul lui Ziyue Yu explorează eficiența modelului ARIMA în prognozarea prețurilor acțiunilor, evidențiind beneficiile și limitările acestuia în contextul piețelor financiare. În fine, cercetarea realizată de Ewa Chodakowska, Joanicjusz Nazarko și colaboratorii lor confirmă utilizarea eficientă a modelului ARIMA în prognozarea radiației solare în diferite condiții climatice, oferind o bază solidă pentru planificarea și gestionarea infrastructurii energetice. Concluziile acestor studii subliniază importanța și versatilitatea modelului ARIMA în analiza datelor de serie temporală, în ciuda diversității contextelor și domeniilor în care este utilizat. În concluzie, modelul ARIMA reprezintă o unealtă valoroasă în analiza datelor de serie temporală, însă este esențial să se recunoască că rezultatele acestuia pot fi influențate de multiple factori și că este necesară o evaluare atentă a contextului și a datelor disponibile pentru a obține predicții precise și relevante.

## Metodologia cercetării

Procedura Box-Jenkins este un proces sistematic utilizat pentru identificarea și selectarea celui mai potrivit model autoregresiv pentru analiza unei serii de timp. Această procedură implică mai multe etape esențiale:

* **Etapa 1 -** **Etapa de identificare a staționarității seriei de timp** - Se începe prin calcularea funcțiilor de autocorelație (ACF) și de autocorelație parțială (PACF) pentru a determina dacă seria de timp este staționară. Dacă serie este staționară, procesul continuă cu etapa a treia, altfel se aplică transformări pentru a staționariza seria.
* **Etapa 2 - Etapa de staționarizare a seriei de timp** - În această etapă, seria de date este supusă diverselor transformări pentru a îmbunătăți staționaritatea. Aceste transformări pot include logaritmarea valorilor seriei sau aplicarea unei transformări Box-Cox, urmată de diferențierea seriei de date.
* **Etapa 3 – Etapa de determinare a modelelor autoregresive de start** - Bazându-se pe caracteristicile funcțiilor ACF și PACF ale seriei staționare, se stabilesc modele autoregresive inițiale pentru analiza seriei de timp.
* **Etapa 4 - Etapa de estimare a parametrilor modelelor autoregresive** - În această etapă, parametrii modelelor autoregresive sunt estimati folosind metode statistice adecvate.
* **Etapa 5 - Etapa de testare a caracteristicilor modelelor autoregresive estimate** - Modelele autoregresive estimate sunt supuse testelor statistice pentru a evalua calitatea și adecvarea lor la datele observate.
* **Etapa 6- Etapa de selecție a modelului optimal** - Se alege cel mai potrivit model dintre cele estimate, folosind diverse criterii de performanță, precum criteriile informaționale Akaike (AIC) și Bayesian (BIC), precum și evaluarea adecvării modelului la datele observate.
* **Etapa 7 - Etapa de analiză și prognoză pe baza modelului selectat** - Folosind modelul selectat, se realizează analize și prognoze pentru a înțelege și prezice comportamentul viitor al seriei de timp.

## Descrierea datelor

Setul de date selectat cuprinde 126 de înregistrări reprezentând prețurile de închidere al acțiunilor companiei Inditex, exprimate în EURO. Aceste date au fost extrase de pe Yahoo Finance (<https://finance.yahoo.com/quote/ITX.MC/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAFVmjmiXrnguGs837oq3-jkm3acO90va3FrUitEiEpAE70MUCSBjoK3ebITbtHXXY7aFwfV-20D_Q9o8v5A6O8EGUp0-pjgFAB9vvc81ur8FghrZZnweX9WketuIXoIQLEneQhrzJdsmIA3nsrqede6VYqK7rt6TjkzupSQ9ph_N>) și cuprind informații aparținând unei perioade de aproximativ 10 ani, de la începutul anului 2014 (01.01.2014) până în prezent, la data de 1 mai 2024.

## 

## Modelarea unei serii de timp utilizând metodologia Box-Jenkins. Previziuni.

### Cerință

Se cere modelarea unei serii temporale folosind metodologia Box-Jenkins. Înainte de a face acest lucru, trebuie să efectuăm analize descriptive preliminare asupra seriei, incluzând statistici de bază și grafice. Apoi, vom face previziuni folosind modelul valid și vom analiza cât de bune sunt aceste previziuni, interpretând aceste rezultate.

### Încărcarea setului de date

|  |
| --- |
| p1.jfif |
| Fig. 1. Încărcarea datelor în R |

În primul rând, am încărcat setul de date descris anterior în R, conform figurii 1.

### Graficul seriei și corelograma

|  |
| --- |
|  |
| Fig. 2. Plotul și corelograma |

Din observațiile aduse la lumină de graficul seriei temporale și corelograma asociată, se conturează următoarele concluzii:

* Seriei temporale îi este caracterizat un trend ușor descrescător până la declanșarea pandemiei, perioadă în care se disting anomalii semnificative aproape de anii 2020 și 2022. Aceste evenimente neașteptate au generat o scădere abruptă în prețul acțiunilor, fiind urmate de o remarcabilă redresare, marcând astfel o tendință puternic crescătoare în urma perioadei pandemice. Concluzia este că seria temporală poate fi caracterizată ca nestaționară, având în vedere variația neuniformă a valorilor sale în timp.
* Analiza autocorelației (ACF) relevă modele vizuale de sezonalitate, însă acestea se dovedesc a fi instabile în lungul perioadei analizate. Această instabilitate sugerează absența unui pattern clar și constant al autocorelației în serie.
* Se pare că variația seriei temporale rămâne relativ constantă pe parcursul întregii perioade studiate, indicând astfel o consistentă uniformitate în dispersia datelor.

Aceste constatări oferă o perspectivă detaliată asupra comportamentului seriei temporale, pregătind terenul pentru investigarea și implementarea ulterioară a modelelor de analiză și prognozare.

|  |
| --- |
| p4.jfif |
| Fig. 3. Seasonal subseries plot |

În figura 3 este reprezentată analiza prin "seasonal subseries plot", care nu dezvăluie un pattern stabil de sezonalitate, deoarece mediile sunt în proximitate și diferența dintre mediile consecutive este neuniformă. Această observație va fi confirmată prin aplicarea testului Canova și Hansen pentru stabilirea unei concluzii valide în privința existenței unui pattern de sezonalitate constant.

### Testul Canova și Hansen

|  |
| --- |
| p3.jfif |
| Fig. 4. Testul Canova și Hansen |

Conform rezultatelor testului CH, nu se poate identifica niciun pattern stabil de sezonalitate în serie, deoarece toate valorile p asociate sunt mai mari de 0.05. Această observație indică absența unei structuri de sezonalitate semnificative și constante în datele analizate.

În contextul utilizării modelului ARIMA pentru analiza și previziunea seriilor de timp, setul de date este împărțit în două componente principale: setul de antrenament și setul de testare. Pentru a asigura că modelul capturează o gamă variată de tendințe și variații în datele istorice, alegem să utilizăm 75% din valorile setului complet de rentabilități pentru antrenament. Această proporție generoasă ne permite să învățăm și să ajustăm în mod adecvat parametrii modelului, având în vedere schimbările semnificative cauzate de pandemie. Totuși, este important să subliniem că, datorită evoluției continue a situației și a efectelor pandemiei asupra datelor, nu putem anticipa cu mare acuratețe posibile erori de previziune, în special în proximitatea anului 2022. Prin utilizarea unui set de testare separat, evaluăm capacitatea modelului de a generaliza și de a face predicții precise pe datele care nu au fost folosite anterior în procesul de antrenament.

### Staționaritatea

Staționaritatea este un concept fundamental în modelarea seriilor de timp, având o importanță semnificativă în analiza și prognozarea acestora. O serie temporală este considerată staționară atunci când proprietățile sale statistice, cum ar fi media și varianța, rămân constante în timp și nu prezintă tendințe sau modele ciclice evidente. Importanța staționarității derivă din faptul că aceasta permite utilizarea unor metode statistice și matematice mai robuste și eficiente în analiza și interpretarea datelor. Într-o serie staționară, relațiile dintre valorile observate sunt mai clare și mai ușor de modelat, ceea ce facilitează identificarea de modele predictibile și generalizabile. Mai mult, prognozele bazate pe modele antrenate pe serii staționare sunt mai fiabile și mai precise, oferind informații utile pentru luarea deciziilor într-o varietate de domenii, cum ar fi finanțele, economia, sănătatea publică sau meteorologia. Prin urmare, asigurarea staționarității unei serii temporale este un prim pas crucial în analiza datelor și construirea de modele care să ofere înțelegere și previziuni precise ale comportamentului temporal al fenomenului studiat.

Un "unit root" este o caracteristică a unei serii temporale care indică faptul că seria nu este staționară. Practic, acesta sugerează că există o rădăcină unitară în procesul generativ al seriei, ceea ce înseamnă că schimbările sau variațiile din timp nu sunt efemerizate și se propagă în serie pe termen nelimitat. Atunci când o serie temporală prezintă o unitate root, acest lucru indică faptul că media și varianța seriei nu sunt constante în timp și că există o tendință sau o componentă persistentă care afectează evoluția sa. Acest fenomen este asociat cu lipsa unei stabilități pe termen lung în datele temporale și poate dificulta analiza și interpretarea acestora. Prin urmare, o serie temporală este considerată nestaționară dacă prezintă o unitate root, deoarece variațiile sale nu sunt stabilizate și persistă în timp. Această nestabilitate face ca seria să nu îndeplinească condițiile de staționaritate, cum ar fi media și varianța constante, și poate afecta validitatea analizelor statistice și a prognozelor viitoare.

### Akaike Information Criterion, Testul Dickey Fuller și Augmented Dickey Fuller

Vom verifica daca seria originala este stationara, verificand daca exista radacini unitate in procese autoregresive, la lag-urile specificate prin criteriul de informatie AIC.

AIC (Akaike Information Criterion) este o măsură utilizată în analiza statistică și în modelarea datelor pentru a evalua calitatea relativă a unui model statistic. Scopul său principal este de a selecta cel mai bun model dintre mai multe alternative, ținând cont atât de ajustarea la date, cât și de complexitatea modelului. Scopul AIC este de a găsi acel model care minimizează această măsură, indicând un echilibru între precizia ajustării și complexitatea modelului. Cu alte cuvinte, AIC penalizează modelele cu un număr mare de parametri, ceea ce încurajează selecția unui model mai simplu și mai interpretabil. Prin urmare, un model cu un AIC mai mic este preferat, deoarece acesta indică o bună ajustare la date, luând în considerare și numărul optim de parametri pentru a evita supraadaptarea.

Testul Dickey-Fuller (ADF) și testul Augmented Dickey-Fuller (ADF) sunt metode statistice utilizate pentru a determina dacă o serie temporală este staționară sau nu. Aceste teste sunt folosite în analiza datelor economice, financiare și de altă natură pentru a evalua proprietățile seriei temporale și pentru a asigura validitatea modelelor utilizate în analiză și prognozare. Testul Dickey-Fuller este bazat pe ipoteza nulă că seria temporală are o rădăcină unitară, ceea ce indică faptul că este nestaționară. Testul este conceput pentru a investiga dacă diferențierea seriei temporale poate transforma seria într-una staționară. Dacă ipoteza nulă este respinsă, putem concluziona că seria este staționară. Testul Augmented Dickey-Fuller este o extensie a testului Dickey-Fuller, care permite includerea unor componente suplimentare în modelul de regresie pentru a captura variații suplimentare care pot exista în serie. Aceste componente suplimentare pot include o tendință liniară sau o constantă, care sunt adăugate pentru a îmbunătăți capacitatea testului de a detecta rădăcinile unitare în serie. Așadar, testele Dickey-Fuller și Augmented Dickey-Fuller sunt utilizate pentru a evalua staționaritatea unei serii temporale prin testarea ipotezei că seriei îi lipsește o componentă staționară. Aceste teste sunt esențiale în analiza datelor de serii temporale pentru a asigura validitatea modelelor și a prognozelor bazate pe aceste date.

Astfel, în cadrul proiectului, testul Dickey-Fuller Augmentat (ADF) este utilizat pentru a testa dacă o serie temporală are o rădăcină unitară, ceea ce indică prezența nestaționarității în serie. Acest test poate fi efectuat în trei variante, în funcție de specificațiile incluse în modelul de regresie:

* **ADF fără tendință ("none"):**

În această variantă, modelul de regresie este simplu și nu include niciun termen de tendință. Scopul este de a verifica dacă seria temporală are o rădăcină unitară, indicând lipsa unei componente staționare. Esențial, testul încearcă să determine dacă diferențierea seriei temporale este suficientă pentru a o face staționară.

* **ADF cu tendință ("trend"):**

În această variantă, modelul de regresie include un termen de tendință liniară. Testul ADF cu tendință verifică dacă seria temporală prezintă o tendință liniară și, în același timp, dacă există o rădăcină unitară. Prin includerea unei tendințe liniare în model, testul poate distinge între o serie de timp cu o tendință și una care este pur și simplu nestaționară.

* **ADF cu tendință și constantă ("drift"):**

În această variantă, modelul de regresie include atât un termen de tendință liniară, cât și un termen constant. Testul ADF cu tendință și constantă verifică dacă seria temporală prezintă o tendință liniară și/sau o valoare constantă și, în același timp, dacă există o rădăcină unitară. Prin includerea unei constante în model, testul poate evalua dacă există o medie stabilă în seria temporală, pe lângă o posibilă tendință.

|  |
| --- |
| p5.jfif |
| Fig. 5. Summary pentru NONE |
| p6.jfif |
| Fig. 6. Summary pentru TREND |
| p7.jfif |
| Fig. 7. Summary pentru DRIFT |

Conform output-urilor din figurile 5, 6 și 7 valorile critice sunt mai mari decât valoarea testului statistic, indicând prezența radacinilor unitate în cazul acestei serii nediferențiate. Ca urmare, putem concluziona cu încredere că seria este nestationară.

### Definirea setului de rentabilități

Diferențierea și definirea rentabilităților prețurilor acțiunii reprezintă instrumente esențiale în analiza seriei temporale a prețurilor acțiunilor. Rentabilitatea unei acțiuni se referă la câștigurile sau pierderile realizate de un investitor în urma fluctuațiilor prețului unei acțiuni într-o anumită perioadă de timp. Diferențierea prețurilor acțiunilor implică calculul variației absolute sau procentuale între prețurile succesive ale acțiunii într-o serie temporală. Această procedură este adesea utilizată pentru a transforma o serie nespațială în una staționară, facilitând analiza și modelarea ulterioară a datelor. În timp ce rentabilitatea absolută măsoară diferența de preț în termeni monetari, rentabilitatea procentuală exprimă aceeași variație ca o proporție a prețului inițial al acțiunii. Definirea rentabilităților prețurilor acțiunilor permite investitorilor și analiștilor să evalueze performanța unei acțiuni într-un anumit interval de timp, să identifice tendințe și modele în comportamentul prețurilor și să ia decizii informate în ceea ce privește tranzacțiile și strategiile de investiții. Aceste metrici sunt esențiale în evaluarea performanței și riscului asociat investițiilor în piețele financiare, oferind o perspectivă semnificativă asupra evoluției valorii acțiunilor și a potențialelor oportunități de investiții.

Vom aplica o logaritmare a seriei de timp,urmată de o diferențiere, astfel încât să definim rentabilitățile prețurilor acțiunilor utilizând următoarea formulă:



Această formulă ne permite să calculăm rentabilitatea unei acțiuni la un anumit moment de timp t, comparativ cu prețul său la momentul anterior t-1. Prin aplicarea logaritmului, obținem o metrică care exprimă variația proporțională a prețului acțiunii între două momente de timp consecutive. Astfel, rentabilitatea calculată în acest mod ne oferă o perspectivă asupra modului în care prețul unei acțiuni evoluează în timp și facilitează analiza comportamentului său în cadrul seriei temporale.

Logaritmând seria de timp, putem elimina posibilele efecte bursty ale variației, asigurând homoschedasticitatea.

|  |
| --- |
| p8.jfif |
| Fig. 8. Plotul și corelograma rentabilităților |

Din perspectivă vizuală, seria temporară pare să prezinte caracteristici de stationaritate. Observăm un coeficient de corelație semnificativ la lag-ul 12, atât în ​​coeficienții de autocorelație obișnuiți, cât și în cei parțiali. Totuși, nu putem identifica prezența unui model stabil de sezonalitate, judecând după lag-urile superioare. De asemenea, nu se observă niciun coeficient semnificativ de autocorelație, atât normală, cât și parțială, la lag-urile apropiate. Această lipsă de autocorelație la lag-uri apropiate sugerează că un model ARIMA ar putea să nu fie eficient în acest context.

### Evaluarea staționarității seriei de rentabilități

Pentru a evalua stationaritatea seriei de rentabilități, datele au fost împărțite în seturi de antrenare și testare. 75% din observații au fost alocate setului de antrenare, iar restul pentru testare (începând din iulie 2021 până la finalul seriei). Astfel, setul de antrenare acoperă intervalul de timp până în iunie 2021. În prima etapă, vom testa existența unei rădăcini unitare în setul de antrenare. Deoarece modelul este construit pe setul de antrenare, este crucială identificarea stationarității în acest set de date.

|  |
| --- |
| p9.jfif |
| Fig. 9. Testul Dickey Fuller - DRIFT |

|  |
| --- |
| p10.jfif |
| Fig. 10. Testul Dickey Fuller - TREND |

|  |
| --- |
| p11.jfif |
| Fig. 11. Testul Dickey Fuller - NONE |

Rezultatele testelor Augmented Dickey-Fuller (ADF) indică absența radacinilor unitate în cadrul proceselor autoregresive care includ lag-urile specificate de criteriul de informație Akaike (AIC).

### Testul Philips Peron și Kwiatkowski-Philips-Schmidt-Shin

Testul Phillips-Perron (PP) și testul Kwiatkowski-Phillips-Schmidt-Shin (KPSS) sunt două teste importante utilizate în analiza seriilor de timp pentru a evalua stationaritatea datelor. Testul Phillips-Perron (PP) este similar cu testul Dickey-Fuller augmentat (ADF), însă este mai robust în ceea ce privește autocorelația și heteroscedasticitatea erorilor. Acest test este utilizat pentru a verifica dacă o serie de timp este staționară sau nu prin evaluarea ipotezei nule că seriei îi lipsesc radacinile unitare. Dacă valoarea p asociată testului este mai mică decât un anumit nivel de semnificație prestabilit (de obicei 0.05 sau 0.01), ipoteza nulă este respinsă și se consideră că seria este staționară. Pe de altă parte, testul Kwiatkowski-Phillips-Schmidt-Shin (KPSS) este utilizat pentru a testa ipoteza nulă că o serie de timp este staționară în jurul unei tendințe deterministe, adică nu are o tendință sau o componentă de creștere/degenerare. Dacă valoarea p asociată testului este mai mare decât un anumit nivel de semnificație prestabilit (de obicei 0.05 sau 0.01), ipoteza nulă este respinsă și se consideră că seria nu este staționară.

În esență, testul PP se concentrează pe identificarea radacinilor unitare într-o serie de timp, în timp ce testul KPSS se concentrează pe prezența tendinței deterministe. Ambele teste sunt importante în analiza seriei de timp pentru a asigura că aceasta îndeplinește condițiile necesare pentru aplicarea anumitor modele și metode de analiză.

|  |
| --- |
| p12.jfif |
| Fig. 12. Testul Philips- Perron |

Conform output-ului din figura 12, nu există rădăcini unitate.

|  |
| --- |
| p13.jfif |
| Fig. 13. Testul Kwiatkowski-Phillips-Schmidt-Shin |

În figura 13 este reprezentat output-ul testului KPSS, conform căruia seria este staționară.

### Testul HEGY

Testul HEGY (Hylleberg, Engle, Granger, and Yoo) este o metodă utilizată pentru detectarea radacinilor unitare sezoniere într-o serie de timp. Acest test extinde testul clasic Dickey-Fuller pentru a identifica prezența radacinilor unitare într-o manieră sezonieră. În esență, testul HEGY presupune ajustarea unui model autoregresiv pentru a include termeni sezonieri, iar apoi testarea ipotezei nule că nu există radacini unitare sezoniere în seria de timp. Dacă valoarea p asociată testului este mai mică decât nivelul de semnificație ales, atunci se respinge ipoteza nulă și se consideră că există radacini unitare sezoniere în serie. Acest test este util în identificarea componentelor sezoniere ale unei serii de timp și în determinarea dacă aceste componente contribuie la nestationaritatea seriei.

|  |
| --- |
| p14.jfif |
| Fig. 14. Testul HEGY |

Conform rezultatelor testului HEGY, nu există radacini unitare sezoniere, sugerând că seria este stationară din punct de vedere sezonier.

### Diferențierea a doua a seriei

În continuare, dorim să aplicăm un model ARIMA pentru a analiza seriile de timp. Cu toate acestea, analiza autocorelației și autocorelației parțiale nu prezintă valori semnificative pentru lag-urile apropiate. Prin urmare, vom proceda la diferențierea seriei încă o dată pentru a încerca să estimam potentiale modele ARIMA mai potrivite. Graficul seriei diferențiate de două ori va oferi o perspectivă mai clară asupra naturii seriei de timp și a potențialelor sale caracteristici stationare.

|  |
| --- |
| p15.jfif |
| Fig. 15. Graficul seriei diferențiate de 2 ori |

Bazându-ne pe corelograma seriei diferențiate de două ori, observăm coeficienți semnificativi de autocorelație atât simplă, cât și parțială, la anumite lag-uri. Acest lucru indică o relație directă puternică pentru lag-urile 1, 2, 3 și 4, precum și o relație inversă puternică pentru lag-ul 1, urmată de o scădere abruptă. Având în vedere aceste caracteristici, vom proceda la modelarea seriei folosind metoda ARIMA pentru a explora și identifica potențialele sale modele și parametrii corespunzători.

### Metoda ARIMA

Pentru estimarea parametrilor optimi pentru arima, am facut un script ce verifica termenii autoregresivi de la 0 la 5, termenii de medie mobile de la 0 la 5 pentru seria rentabilitatilor si seria rentabilitatilor diferentiata o data, si extrage cele mai bune 5 modele in functie de AIC. Se iau in considerare doar modelele care au toate variabilele independente statistic semnificative (in afara intercept).

|  |
| --- |
| p16.jfif  p17.jfif |
| Fig. 16. ARIMA – Script în R |

|  |
| --- |
| p18.jfif |
| Fig.17. Estimarea parametrilor optimi pentru modelele ARIMA |

### Diagnostic reziduuri

Pentru modelele cu cel mai bun AIC, se va efectua un diagnostic al reziduurilor. Acest proces constă în analiza reziduurilor rezultate din ajustarea modelului pe datele observate. Scopul diagnosticului este de a verifica dacă reziduurile îndeplinesc anumite asumpții ale modelului statistic, cum ar fi independența, omogenitatea varianței și lipsa autocorelației. Prin această analiză, se asigură că modelul este adecvat pentru datele observate și că nu există pattern-uri sau structuri nedorite în reziduuri care ar putea indica o specificare incorectă a modelului.

În contextul analizei seriilor de timp, "white noise" reprezintă o secvență de variabile aleatoare independente și identic distribuite, caracterizate prin media zero și varianță constantă. Aceste variabile aleatoare nu prezintă niciun tip de autocorelație între ele și sunt considerate complet imprevizibile. În esență, "white noise" este un proces stochastic în care fiecare observație este complet independentă de celelalte și nu prezintă niciun pattern sau structură discernabilă în timp. În cadrul analizei seriilor de timp, prezenta unui white noise în reziduurile modelului indică faptul că acesta nu a capturat nicio altă informație relevantă din datele observate și că nu există nicio corelație remanentă care ar putea indica o specificare inadecvată a modelului.

Astfel, se doreste ca erorile acestor modele sa fie distribuite normal si sa nu prezinte corelatie, sa fie white noise, astfel incat sa putem afirma ca toate erorile din modelul nostru sunt un rezultat pur aleator.

|  |
| --- |
| p19.jfif |
| Fig. 18. Script R – ARIMA (4,0,4) |

Pentru a filtra modelele în funcție de diagnosticul pe reziduuri, se implementează un script (Figura 18) care efectuează următoarele verificări:

* Testul de autocorelație pentru reziduurile modelului, în care se examinează dacă există autocorelație semnificativă la lagurile de la 1 la 12.
* Testul Jarque-Bera pentru normalitatea distribuției reziduurilor. Acest test verifică dacă reziduurile prezintă o distribuție normală, având în vedere skewness-ul și kurtosis-ul acestora.

Prin implementarea acestui script, se selectează modelele care trec ambele teste, indicând astfel că reziduurile lor sunt independente și au o distribuție normală, ceea ce sporește credibilitatea modelului respectiv în analiza seriilor de timp.

|  |
| --- |
| p20.jfif |
| Fig. 19. Output ARIMA (4,0,4) |

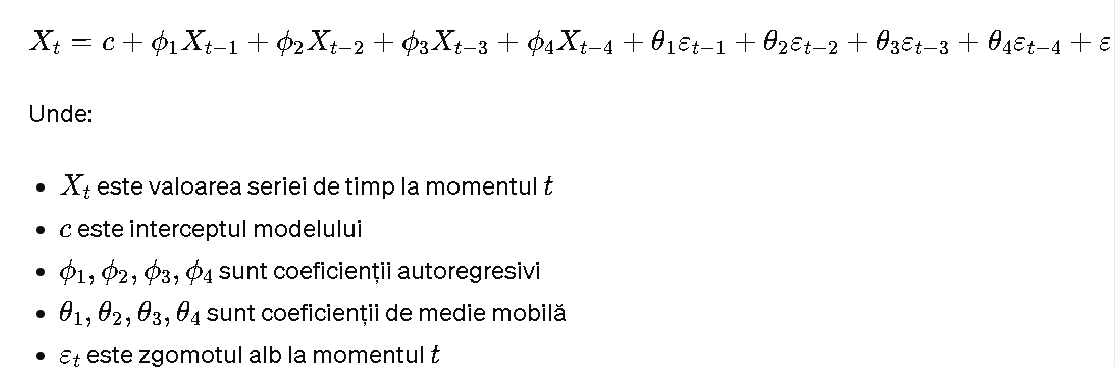
Conform rezultatelor obținute din acest script, modelul ARIMA (4,0,4) este identificat ca fiind optim pentru previzionarea valorilor acestei serii de timp. Alegerea acestui model este susținută de diagnosticul pe reziduuri, care indică faptul că reziduurile modelului respectiv sunt independente și urmează o distribuție normală. Astfel, modelul ARIMA (4,0,4) prezintă capacitatea cea mai bună de a captura structura și variabilitatea din seria de timp, oferind estimări fiabile și precise pentru valorile viitoare.

|  |
| --- |
| P21.jfif |
| Fig. 20. Graficul reziduurilor modelului ARIMA (4,0,4) |

Din analiza grafică, se poate deduce că nu există autocorelație semnificativă între reziduurile modelului la lagurile mai mici de 36. De asemenea, distribuția reziduurilor pare să se apropie de o distribuție normală, conform asumpțiilor modelului. Aceste observații indică faptul că modelul ARIMA (4,0,4) este robust și că reziduurile sale sunt în mare măsură independente și distribuite aproximativ normal, ceea ce sprijină adecvarea sa pentru analiza și previziunea seriei de timp.

În continuare, vom dezvolta modelul ARIMA (4,0,4) pentru setul de date training. Ecuația acestui model este:

Xt = c - 0,91 Xt-1 – 0.64 Xt-2 – 0.80 Xt-3 – 0.92c Xt-4 + 0.82ε t-1 + 0.50 ε t-2 + 0.81 ε t-3 + 0.99 ε t-4 , conform următoarei ecuații generale:



|  |
| --- |
| p22.jfif |
| Fig. 21. Estimarea parametrilor modelului ARIMA (4,0,4) |

Următorul pas include previzionarea următoarelor 32 de valori și graficul acestora comparativ cu datele de testare.

|  |
| --- |
| p23.jfif |
| Fig. 22. ARIMA (4,0,4) Forecast vs. Test Data |

Se constată că modelul ARIMA(4,0,4) nu reușește să estimeze datele cu o precizie semnificativă, însă reușește să surprindă câteva tendințe în perioada de testare. Acest model nu reușește să surpindă tendințele de la începutul perioadei de testare, întrucât se produc anumite anomalii cauzate de pandemie, însă reușește să captureze câteva tendințe de după finalizarea pandemiei.

|  |
| --- |
| p24.jfif |
| Fig. 23. Graficul erorilor de previzionare |

# SARIMA

## Literatură de specialitate

În analiza seriei temporale, modelele ARIMA (Autoregressive Integrated Moving Average) au fost folosite pe scară largă pentru a realiza previziuni și pentru a înțelege comportamentul datelor în timp. Cu toate acestea, atunci când se confruntă cu serii temporale cu componente sezoniere pronunțate, modelele ARIMA pot să nu fie suficient de flexibile pentru a le captura în întregime. Pentru a aborda această problemă, a fost dezvoltat modelul SARIMA (Seasonal Autoregressive Integrated Moving Average), care aduce în plus capacitatea de a modela și de a prognoza serii temporale care prezintă componente sezoniere. SARIMA extinde modelul ARIMA prin includerea unor componente sezoniere în funcția de autoregresie, integrare și medie mobilă, permițând astfel adaptarea mai precisă la tendințele sezoniere ale datelor. Prin urmare, acest model devine esențial în analiza și prognozarea datelor care prezintă variații ciclice sau sezoniere, precum vânzările lunare, producția sezonieră sau alte fenomene care manifestă modele periodice.

Un prim studiu analizat este intitulat "SARIMA: A Seasonal Autoregressive Integrated Moving Average Model for Crime Analysis in Saudi Arabia", realizat de Talal H. Noor, Abdulqader M. Almars, Majed Alwateer, Malik Almaliki, Ibrahim Gad și El-Sayed Atlam în anul 2022. Scopul acestei cercetări este de a investiga modelele de criminalitate în Arabia Saudită și de a prezice ratele de criminalitate viitoare în diferite zone ale țării. Autorii propun un model SARIMA, care se concentrează pe identificarea factorilor care afectează infracțiunile în Arabia Saudită, estimând o rată rezonabilă a criminalității și identificând probabilitatea distribuției infracțiunilor pe baza diferitelor locații. Lucrarea folosește un set de date obținut din surse guvernamentale oficiale saudite, care conține informații detaliate despre timp și loc, împreună cu statisticile privind diferite tipuri de infracțiuni. Această lucrare a adus mai multe contribuții importante în domeniul analizei și prognozării criminalității în Arabia Saudită. În primul rând, a identificat o nevoie semnificativă de a îmbunătăți modelele existente de predicție a criminalității, care pot fi mai puțin precise în anumite situații. De asemenea, a evidențiat faptul că există factori promițători care ar putea permite prognozarea ratelor de criminalitate cu o precizie mai mare. Principala contribuție a acestei lucrări este introducerea modelului SARIMA pentru analiza modelelor de criminalitate și prognozarea criminalității în mai multe orașe din Arabia Saudită. Utilizând un set de date obținut din surse guvernamentale oficiale, lucrarea a demonstrat eficacitatea modelului propus în identificarea distribuției infracțiunilor în diferite orașe, precum și în prognozarea criminalității viitoare și luarea măsurilor corespunzătoare pentru prevenirea acestora.

În anul 2023, un alt studiu intitulat " Using the Statistical Machine Learning Models ARIMA and SARIMA to Measure the Impact of Covid-19 on Official Provincial Sales of Cigarettes in Spain " a fost realizat de Andoni Andueza, Miguel Ángel Del Arco-Osuna, Bernat Fornés, Rubén González-Crespo și Juan Manuel Martín-Álvarez. Acest studiu se concentrează pe utilizarea modelelor ARIMA și SARIMA pentru a prognoza impactul pandemiei Covid-19 asupra vânzărilor de țigări în provinciile spaniole. Folosind datele oficiale privind vânzările de țigări și o serie de modele statistice și metode de evaluare, cercetătorii au analizat modul în care pandemia Covid-19 a influențat vânzările de țigări în diverse regiuni, cu accent pe provinciile de frontieră și cele cu un flux ridicat de turiști. Concluzia acestei cercetări este că modelul SARIMA (Seasonal Autoregressive Integrated Moving Average) s-a dovedit a fi o unealtă utilă în analiza și prognozarea vânzărilor de tutun în diferite provincii din Spania în contextul pandemiei COVID-19. Utilizarea acestui model a permis identificarea impactului sezonier al pandemiei asupra vânzărilor de tutun și diferențierea acestui impact în funcție de caracteristicile specifice ale fiecărei provincii. Concluziile obținute cu ajutorul modelului SARIMA au oferit informații valoroase decidenților politici și autorităților în luarea unor măsuri eficiente pentru gestionarea vânzărilor de tutun și a impactului pandemiei în diferite regiuni ale Spaniei.

O altă cercetare realizată în același an, 2023, de către Ankita Vaswani, Dr. Poorna Chandra Prasad și Dr. Prafulla Kumar Padhi are titlul “Time Series Analysis: An application of SARIMA Model in General Trade to Forecast Sales” și se concentrează pe importanța comerțului general în industria bunurilor de larg consum (FMCG) și pe rolul său în asigurarea disponibilității produselor pentru consumatori într-un mod eficient. Autorii explorează utilizarea modelului de serie temporală SARIMA pentru a prognoza vânzările de la comerțul cu ridicata către comerțul cu amănuntul. Ei analizează o cantitate mare de date pentru a prezice vânzările, folosind modele ARIMA și SARIMA și constată că modelul SARIMA este cel mai potrivit pentru a prognoza vânzările în retail atât pentru distribuitorii, cât și pentru comercianții cu amănuntul. Studiul subliniază importanța prevederii cererii viitoare și a gestionării stocurilor în mod eficient în industria FMCG. Concluzia acestui studiu evidențiază că modelul SARIMA (Seasonal Autoregressive Integrated Moving Average) este extrem de semnificativ în prognozarea vânzărilor de la comerțul cu ridicata către comerțul cu amănuntul. Rezultatele obținute indică că estimarea vânzărilor este deosebit de importantă prin reducerea erorii estimate și înțelegerea factorului de neregularitate. Obiectivul principal al studiului a fost să dezvolte și să utilizeze un model solid în comerțul cu ridicata, prin intermediul căruia comercianții cu ridicata să poată prognoza tendința vânzărilor pentru anii viitori. Utilizând abordarea seriei temporale și modelul SARIMA, studiul a demonstrat că este posibilă estimarea vânzărilor pe baza datelor istorice și înțelegerea comportamentului și modelelor expuse la neregularitate. Rezultatele indică faptul că variația sezonieră a fost bine abordată în cadrul modelului, iar adoptarea acestuia în comerțul cu ridicata se dovedește a fi mai fiabilă în comparație cu alte modele, fiind cel mai potrivit în metodele de serie temporală.

A naliza seriei temporale și utilizarea modelului SARIMA (Seasonal Autoregressive Integrated Moving Average) au demonstrat să fie instrumente valoroase în diferite domenii de cercetare și aplicații practice. Studiile examinate în acest context au ilustrat eficacitatea acestui model în analiza și prognoza diferitelor fenomene, cum ar fi criminalitatea în Arabia Saudită sau vânzările de tutun în Spania în contextul pandemiei COVID-19, precum și în prognozarea vânzărilor în comerțul cu ridicata către comerțul cu amănuntul. Utilizarea modelului SARIMA a permis cercetătorilor să identifice și să înțeleagă mai bine tendințele sezoniere și modelele de comportament ale datelor, facilitând luarea deciziilor informate și implementarea măsurilor adecvate în diferite domenii. Concluziile obținute din aceste studii evidențiază importanța și relevanța modelului SARIMA în gestionarea datelor de serie temporală și în realizarea previziunilor precise, subliniind necesitatea continuării cercetărilor în acest domeniu și explorării potențialului său extins în diferite contexte și aplicații.

## Metodologia cercetării

Modelul SARIMA (Seasonal Autoregressive Integrated Moving Average) este o metodă esențială utilizată pentru analiza și prognoza seriilor de timp care prezintă variații sezoniere. Acest model poate fi împărțit în două categorii: modelul SARIMA simplu și modelul SARIMA multiplu. În general, seria temporală inițială {Yt} este supusă unui proces de transformare într-o secvență staționară pentru a îmbunătăți precizia modelelor de prognoză. Această etapă implică eliminarea comportamentului ciclic și sezonier din seria temporală, de obicei prin diferențiere. Identificarea componentelor sezoniere și/sau ciclice se face prin analiza autocorelației (ACF), iar în cazul în care secvența este nestationară, se aplică una sau două diferențe temporale pentru a netezi seria prin procesarea diferențelor. Următoarea etapă constă în identificarea modelului potrivit. Prin analiza autocorelației (AC) și a funcției de autocorelație parțială (PAC) a seriei netezite, se determină inițial ordinea modelului. De obicei, se utilizează o metodă de încercare și eroare pentru a estima mai multe modele diferite pe baza criteriului de informare (AIC). Modelul final este ales prin comparație.

După selectarea modelului potrivit, parametrii acestuia sunt estimati folosind metoda celor mai mici pătrate neliniare. Estimările preliminare sunt obținute prin optimizarea parametrilor modelului, iar metoda celor mai mici pătrate neliniare este utilizată pentru estimarea parametrilor. În final, se testează modelul de predicție folosind metoda testului χ2 pentru a evalua secvența reziduală și a verifica dacă aceasta se comportă ca un zgomot alb. Dacă secvența reziduală nu are caracteristicile unui zgomot alb, înseamnă că modelul nu a capturat toate informațiile utile și trebuie să fie îmbunătățit în continuare.

## Descrierea datelor

Pentru realizarea aplicației 2 vom folosi același set de date ca și în aplicația 1, înregistrări reprezentând prețurile de închidere al acțiunilor companiei Inditex, exprimate în EURO ( perioada 01.01.2014 – 01.05.2024).

## Motivarea utilizării modelului SARIMA

În analiza corelogramei seriei de rentabilități, observăm prezența unei autocorelații semnificative la lag-ul 12, chiar în absența unui pattern sezonier stabil. Acest lucru indică faptul că există o corelație între valorile rentabilității la intervalul de timp actual și cele de 12 luni în urmă. Bazându-ne pe aceste constatări, putem formula un model auto-regresiv sau de medie mobilă, având în vedere lag-urile semnificative identificate. ACF (funcția de autocorelație) și PACF (funcția de autocorelație parțială) prezintă valori nesemnificative după lag-ul 12 pentru seria de rentabilități normală, seria diferențiată de ordin 1 și seria diferențiată sezonier, cu coeficienți de autocorelație mai puternici pentru lag-urile 12. În acest context, vom explora o varietate de modele SARIMA care includ componente auto-regresive și de medie mobilă cu valori între 0 și 1. Pentru identificarea parametrilor optimi ai acestor modele, vom dezvolta o funcție de grid search, similară cu cea utilizată pentru identificarea parametrilor optimi ai modelului ARIMA. Această abordare ne va permite să selectăm și să ajustăm modelul SARIMA care se potrivește cel mai bine datelor noastre și să obținem previziuni precise pentru serie.

## UTILIZAREA MODELULUI SARIMA

### Estimarea parametrilor optimi

|  |
| --- |
| p1.jfif |
| Fig. 24. Funcția de estimare a parametrilor optimi |

Această funcție reprezentată în figura 24 primește ca parametri un set de date și intervalele în care să caute fiecare parametru al modelului. De asemenea, este specificată perioada de sezon (în cazul nostru, 12). Funcția analizează aceste informații pentru a formula o listă care să conțină detaliile celor mai bune cinci modele, evaluându-le în funcție de criteriul informațional Akaike (AIC). Modelele sunt selectate pe baza coeficienților auto-regresivi (AR), ai mediei mobile (MA), ai auto-regresiei sezoniere (SAR) și ai mediei mobile sezoniere (SMA), asigurându-se că aceștia sunt semnificativi din punct de vedere statistic. Această listă furnizează informații cruciale pentru alegerea celui mai potrivit model SARIMA și pentru realizarea previziunilor precise pentru seria temporală dată.

|  |
| --- |
| p2.jfif |
| Fig. 25. Funcție grid search |

Funcția aferentă figurii 25 a fost folosită în scopul căutării modelelor optime prntru p,q aparținând mulțimii {0,1,2,3,4} și d,P,D,Q aparținaâd mulțimii {0,1}.

|  |
| --- |
| p3.jfif |
| p4.jfif |
| Fig. 26. Modelele cu parametri optimi |

Se dezvoltă o funcție pentru filtrarea testelor pe baza diagnosticului reziduurilor, care are rolul de a verifica dacă reziduurile fiecărui model prezintă autocorelație, ceea ce ar putea indica faptul că nu am capturat în totalitate toate modelele prezente în serie și am lăsat un model în reziduuri. De asemenea, se verifică dacă aceste reziduuri sunt distribuite normal, având caracteristici similare cu zgomotul alb. Această abordare asigură evaluarea adecvată a calității modelelor propuse și confirmarea faptului că acestea pot fi utilizate pentru realizarea de previziuni precise și fiabile.

|  |
| --- |
| p5.jfif |
| Fig. 27. Output-ul funcției de filtrare |

Așa cum se poate observa din output-ul reprezentat în figura 27, modelul optim SARIMA este cel cu parametrii ARMA (0,1), SAR(1), SMA(1).

### Ecuația modelului SARIMA

|  |
| --- |
| p6.jfif |
| Fig. 28. Coeficienții ecuației SARIMA |

|  |
| --- |
| p1777.jfif |
| Fig. 29. Ecuația SARIMA generalizată |

În contextul proiectului, ecuația modelului SARIMA este, deci :

**Xt = c – 0.26 εt-1 – 0.94 Xt-12 – 0.81 εt-12**

### Graficul și corelograma reziduurilor

|  |
| --- |
| p7.jfif |
| Fig. 30. Graficul și corelograma reziduurilor modelului SARIMA |

Din analiza plot-ului generat de funcția checkresiduals() se poate observa că erorile prezintă o tendință de a fi distribuite în mod normal, iar niciunul dintre coeficienții de autocorelație nu este semnificativ diferit de zero. Aceste constatări sugerează că modelul SARIMA potrivit a reușit să captureze majoritatea variabilității din seria temporală și că reziduurile modelului au proprietățile dorite ale unui zgomot alb. Acest lucru indică faptul că modelul este adecvat pentru analiză și prognoză, iar rezultatele sale pot fi considerate fiabile și robuste în contextul utilizării lor pentru previziuni viitoare.

|  |
| --- |
| fig30.jfif |
| Fig. 31. Reprezentare grafică Forecast vs. Test Data |

Pentru perioada de testare, se efectuează previziuni folosind modelul SARIMA și se compară aceste valori previzionate cu valorile din setul de test. Ulterior, rezultatele sunt afișate sub formă grafică (figura 31) pentru a evidenția discrepanțele și similaritățile între valorile observate și cele prognozate. Această comparație vizuală oferă o evaluare rapidă a performanței modelului și a gradului său de acuratețe în raport cu datele reale din setul de test.

Astfel, se observă că modelul SARIMA reușește să captureze mai bine pattern-urile și structura datelor din seria de timp, ceea ce se traduce printr-o previziune mai precisă a următoarelor rentabilități ale prețurilor. Acuratețea îmbunătățită a previziunilor este esențială pentru luarea deciziilor informate și pentru gestionarea eficientă a riscurilor și oportunităților în mediul economic.

Deși modelul prezice o scădere bruscă a rentabilității prețului în apropierea anului 2022, ceea ce corespunde perioadei pandemice, această previziune poate fi rezultatul unei coincidențe, deoarece noile componente adăugate sunt în concordanță cu valorile seriei din urmă cu 12 luni. Totuși, este important de menționat că datele previzionate pentru perioada post-pandemie sunt într-o anumită concordanță cu valorile actuale, sugerând că modelul reușește să capteze corect tendințele și schimbările din serie. Este esențial să evaluăm și să interpretăm aceste previziuni cu precauție, având în vedere complexitatea și incertitudinea dinamicii pieței și a mediului economic.

|  |
| --- |
| fig31.jfif |
| Fig. 32. Grafic, corelogramă și histogramă serie erori de previziune |

Se calculează seria erorilor de previziune ca diferența dintre valorile actuale și valorile previzionate, iar apoi se afișează grafic această serie, alături de corelograma și histograma ei. Aceste grafice sunt utile pentru a evalua dacă erorile de previziune au proprietățile dorite, cum ar fi distribuția normală și lipsa autocorelației.

Putem concluziona că erorile de previziune par să fie white noise, sugerând că acestea pot fi rezultatul unui proces aleatoriu. Lipsa autocorelației semnificative în seria de erori și distribuția normală a acestora, confirmată de testul Jarque Bera cu o statistică p de 0.7, sprijină această concluzie.

# MODELAREA ARCH

Testarea pentru efectele ARCH este o etapă crucială în evaluarea adecvării modelului pentru capturarea volatilității condiționate heteroscedastice în seriile de timp financiare. Modelul ARCH (Autoregressive Conditional Heteroscedasticity) presupune că volatilitatea prețurilor este condiționată de deviația standard a prețurilor la diferite lag-uri. Pentru a testa dacă reziduurile modelului SARIMA prezintă aceste efecte, vom dezvolta o funcție care aplică testul de heteroscedasticitate condiționată reziduurilor modelului. Această funcție va returna informații despre existența sau absența efectelor ARCH, furnizând astfel o evaluare suplimentară a adecvării modelului pentru previzionarea volatilității în serie.

|  |
| --- |
| fig4.jfif |
| Fig. 33. Funcție pentru verificarea efectelor ARCH |

|  |
| --- |
| p66.jfif |
| Fig. 34. Output corespunzător funcției de verificare a efectelor ARCH |

Observăm că nu există efecte ARCH semnificative la niciun nivel de semnificație, inclusiv la un prag p de 0,1, pentru lag-urile cuprinse între 1 și 24. Acest rezultat sugerează că nu este necesară utilizarea unui model ARCH pentru a captura volatilitatea condiționată heteroscedastică în datele noastre. Astfel, putem concluziona că modelul SARIMA a reușit să captureze eficient caracteristicile seriei de timp și să ofere previziuni relevante pentru volatilitatea prețurilor fără a necesita extinderea sa cu un model ARCH.

# SIMPLE EXPONENTIAL SMOOTHING, MODELUL HOLT ȘI HOLT WINTERS

Metoda Simple Exponential Smoothing (SES) este una dintre cele mai fundamentale și ușor de înțeles tehnici de previzionare în analiza seriei temporale. Această metodă se bazează pe principiul ponderării exponențiale a observațiilor anterioare pentru a estima valoarea viitoare a unei serii de timp. În esență, SES atribuie un nivel de greutate decrescător observațiilor mai vechi, cu o rată de descreștere controlată de un factor de estompare, ceea ce permite modelului să reacționeze rapid la schimbările recente în datele de intrare. Această abordare simplă, dar puternică, face ca SES să fie o alegere populară pentru previziunile de scurtă durată sau atunci când se dorește o metodă rapidă și eficientă de estompare a seriei temporale.

Metoda lui Holt, cunoscută și sub numele de estompare dublă exponențială, reprezintă o extensie a metodei simple de estompare exponențială (SES), care este utilizată pentru a gestiona atât tendințele, cât și sezonieritatea în datele seriei temporale. Holt a introdus această tehnică pentru a aborda limitările SES în capturarea tendințelor și schimbărilor pe termen lung din datele seriei temporale. Prin introducerea unui al doilea set de coeficienți de estompare, Holt a permis modelului să capteze atât nivelul, cât și tendința datelor seriei temporale, oferind astfel previziuni mai precise și mai flexibile. Această metodă este utilă în cazurile în care datele prezintă o tendință clară și consistentă în timp, precum și variații sezoniere care trebuie luate în considerare în procesul de previzionare.

Modelul lui Holt-Winters, denumit uneori și estompare triplă exponențială, reprezintă o metodă avansată de estompare exponențială care poate gestiona atât tendințele, cât și sezonieritatea în datele seriei temporale, precum și variații aleatorii. Acest model a fost dezvoltat ca o extensie a metodei lui Holt, adăugând o componentă suplimentară pentru a captura variațiile sezoniere din date. Prin introducerea unei componente de estompare sezonieră, modelul lui Holt-Winters poate furniza previziuni mai precise și mai robuste în cazurile în care datele prezintă modele de sezonieritate clar definite și consecvente. Principala sa utilizare constă în analiza și prognoza datelor seriei temporale care prezintă atât tendințe pe termen lung, cât și variații periodice.

Se formulează modele de netezire exponențială atât pentru seria prețurilor acțiunilor, cât și pentru seria rentabilităților.

### Pentru seria de rentabilități

Atunci când se modelează seria rentabilităților folosind netezirea exponentială simplă și netezirea exponentială dublă, este de așteptat ca previziunile acestor modele să fie foarte asemănătoare. Aceasta se datorează faptului că seria nu prezintă un trend semnificativ, iar ambele modele se bazează pe valori istorice pentru a estima valorile viitoare. Astfel, atât netezirea exponentială simplă, cât și cea dublă, vor oferi previziuni similare pentru această serie de date.

|  |
| --- |
| fiiii.jfif |
| Fig. 38. Parametrii modelului SES |

|  |
| --- |
| param.jfif |
| Fig. 39. Parametrii modelului Holt |

Similaritatea dintre modele poate fi observată încă din parametrii alpha și beta ai acestora. Deși modelul Holt estimează o valoare inițială a valorilor puțin mai mare, tendința este în mare parte aceeași, iar parametrul beta, corespunzător trendului, este foarte mic. Acest lucru este demonstrat și în graficul generat de metodele respective.

|  |
| --- |
| grpp.jfif |
| Fig. 40. Previziunile modelului SES și HOLT |

Pentru a modela eficient această serie, adăugăm factorul de sezonalitate, comparând modelele Holt-Winters multiplicative și aditivă. Deoarece avem valori apropiate de zero în seria de rentabilități, adăugăm o constantă seriei pentru a evita erorile în modelul multiplicativ, evitând împărțirea la zero. Această constantă va fi considerată a fi 5.

|  |
| --- |
| al1.jfif |
| Fig. 41. Previziunile modelului Holt Winters |

Din grafic, se poate observa că ambele modele previzionează cu o anumită acuratețe datele de test.

|  |
| --- |
| l1.jfif |
| Fig. 42. Graficul si corelograma modelului multiplicativ |

|  |
| --- |
| l2.jfif |
| Fig. 43. Graficul si corelograma reziduurilor din modelul aditiv |

### Pentru seria acțiunilor

Se formulează modelul de netezire exponențială simplă și modelul de netezire exponențială dublă, cel ce ține cont de trendul seriei de timp. Întrucât seria inițială prezintă un trend puternic pozitiv pe intervalul de test, putem afirma ca modelul Holt va surprinde această creștere bruscă și se va potrivi cu o acuratețe mai mare pe datele de test.

|  |
| --- |
| p44.jfif |
| Fig. 35. Parametrii modelului HOLT |

Atât pentru modelul Holt de netezire dublă cât și pentru cel de netezire exponențială simplă, observăm că coeficientul estimat pentru alpha este mare, acest lucru indicând faptul că în previziunea valorii la momentul t, se acordă o pondere mare a valorii anterioare, iar o pondere exponențial mai mică se acordă valorilor anterioare momentului t (t-1, t-2,..). De asemenea, coeficientul beta obținut în cadrul modelului Holt, precum și starea sa inițială indică existența unui trend stabil și pozitiv ((beta <0,0001).

|  |
| --- |
| graf.jfif |
| Fig. 36. Graficul previziunilor dupa metoda SES si metoda Holt |

Pentru seria de prețuri reale, se estimează modele de netezire exponentială triplă (Holt-Winters), care iau în considerare și sezonalitatea din serie. Observând o creștere bruscă în perioada de test, la o primă vedere, se poate spune că un model multiplicativ ar fi mai potrivit pentru previzionarea acestor valori. Acest lucru se datorează faptului că un model multiplicativ are mai puțină variație, dar un bias (eroare sistemată) destul de mare (HIGH BIAS, LOW VARIANCE).

Dacă dorim ca seria să fie previzionată cu mai mare stabilitate, dar fără a capta în aceeași măsură volatilitatea, putem aplica un model aditiv. Un model aditiv implică adunarea componentelor de bază ale seriei (nivel, trend și sezonieritate), iar orice creștere sau scădere este aplicată la aceste componente de bază. Astfel, modelul aditiv este mai stabil în previziunea datelor, având un bias mai mic, dar poate să nu captureze în mod adecvat schimbările proporționale din seria de timp.

|  |
| --- |
| holt.jfif |
| Fig. 37. Previziunile modelelor Holt Winters |

### Observăm că modelele Holt-Winters nu previzionează cu acuratețe scăderea bruscă din perioada pandemiei. Acest lucru este de așteptat, însă pentru previzionarea trendului puternic crescător de după pandemie, modelul multiplicativ capturează mai bine tendințele.

# BIBLIOGRAFIE

* Suport de curs SERII DE TIMP – Profesor universitar dr. DAVIDESCU ADRIANA-ANAMARIA
* Support de seminar SERII DE TIMP – Cadru asociat drd. MANTA EDUARD-MIHAI
* Smith Tim. Random Walk Theory. Investopedia, 2019
* Hayes Adam. Autoregressive Integrated Moving Average (ARIMA). Investopedia, 2022
* Gao Jie. Research on Stock Price Forecast Based on ARIMA-GARCH Model. Web of Conferences, 2021, 292.
* <file:///C:/Users/Lenovo/Downloads/Paper_65-ARIMA_Model_for_Accurate_Time_Series%20(2).pdf>
* <file:///C:/Users/Lenovo/Downloads/IFMPT2024-516-521%20(4).pdf>
* <file:///C:/Users/Lenovo/Downloads/1-s2.0-S2211379721006197-main%20(1).pdf>
* <https://www.inditex.com/itxcomweb/es/home>
* <https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/The_Box-Jenkins_Method.pdf>
* <https://www.forbes.com/companies/inditex/?sh=16a98ef251e9>
* <https://www.statista.com/topics/10161/inditex/#dossier-chapter4>
* <https://www.ambitionbox.com/overview/inditex-overview>
* <https://finance.yahoo.com/quote/ITX.MC/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAFVmjmiXrnguGs837oq3-jkm3acO90va3FrUitEiEpAE70MUCSBjoK3ebITbtHXXY7aFwfV-20D_Q9o8v5A6O8EGUp0-pjgFAB9vvc81ur8FghrZZnweX9WketuIXoIQLEneQhrzJdsmIA3nsrqede6VYqK7rt6TjkzupSQ9ph_N>