 **„Academia de Studii Economice”**

**Facultatea de Cibernetica, Statistica si Informatica Economica**

**Mater Statistică aplicată și Data Science**

**PROIECT ANALIZA AVANSATĂ A SERIILOR DE TIMP**

**Previziuni asupra consumului de energie electrică în România**

**analizând perioada 2008-2022**

PROFESORI COORDONATORI:

**Prof. Univ. Dr. DAVIDESCU ADRIANA – ANAMARIA**

**Prof. Univ. Dr. PELE DANIEL TRAIAN**

STUDENȚI:

**CONȚOLENCU BIANCA**

**DOMBICI ANDREEA**

**MARIN RAMONA-ALEXANDRA**

Cuprins

[**Introducere** 3](#_Toc136520698)

[**Descrierea bazei de date** 4](#_Toc136520699)

[**Modele cu trend determinist sau stochastic** 4](#_Toc136520700)

[**Modele univariate** 7](#_Toc136520701)

[**Testarea staționarității seriei** 8](#_Toc136520702)

[Testul ADF – Augmented Dickey-Fuller 8](#_Toc136520703)

[Testul KPSS 9](#_Toc136520704)

[Testul Zivot and Andrews 10](#_Toc136520705)

[**Modelul ARIMA** 12](#_Toc136520706)

[**Modele ARCH - GARCH** 15](#_Toc136520707)

[Modelul ARCH 19](#_Toc136520708)

[Modelul GARCH 20](#_Toc136520709)

[Modelul E-GARCH 22](#_Toc136520710)

[Modelul AP-ARCH 22](#_Toc136520711)

[**Netezirea exponențială** 24](#_Toc136520712)

[**Modele multivariate** 28](#_Toc136520713)

[**VAR – Analiza relației dintre produsul intern brut și randamentul consumului de energie electrică** 28](#_Toc136520714)

[Analiza staționarității seriei de timp privind Produsul Intern Brut per capita 29](#_Toc136520715)

[Cointegrarea seriilor 30](#_Toc136520716)

[Analiza modelului propriu-zis 31](#_Toc136520717)

[Testarea ipotezelor pentru verificarea diagnosticului pe reziduuri 32](#_Toc136520718)

[Cauzalitatea Granger 33](#_Toc136520719)

[Funcția de răspuns la impuls 34](#_Toc136520720)

[Descompunerea varianței 35](#_Toc136520721)

[Forecast 36](#_Toc136520722)

[**Machine Learning. Modelul XGBosst** 37](#_Toc136520723)

[**Concluzii** 38](#_Toc136520724)

[**Bibliografie** 39](#_Toc136520725)

[**Anexă** 40](#_Toc136520726)

# **Introducere**

În zilele noastre, energia este unul dintre factorii importanți care afectează dezvoltarea vieții umane modern. Creșterea populației, creșterea nivelului de viață, urbanizarea, evoluțiile tehnologice și industrializarea cresc cererea de energie a țărilor. Energia poate fi produsă din diferite surse, cum ar fi gaze naturale, petrol, cărbune, soare, vânt, valuri oceanice, biocombustibili, căderea apei etc. Sursele mondiale de producție de energie constau în 36,1% petrol, 18% cărbune, 26% gaz natural, 5,8% % biocombustibili și deșeuri, 9,8% nuclear, 2,2% hidro și 2,1% altele în 2015. Din cauza încălzirii globale și a schimbărilor climatice cauzate de emisiile de gaze cu efect de seră provenite din utilizarea combustibililor fosili, sursele de energie regenerabilă precum soarele, vântul, biomasa, geotermala au ajuns în prim-plan și au crescut procentul de utilizare în ultimii ani față de altele. [1]

Prognoza consumului de energie are un rol important în planificarea strategiilor energetice atât pentru factorii de decizie, cât și pentru organizațiile conexe din orice țară. În literatura de specialitate, mulți cercetători au aplicat diferite metode precum media mobilă, regresia multiplă, netezirea exponențială, rețeaua neuronală, gri etc. pentru a prognoza consumul de energie pe baze sectoriale sau total. Modelul ARIMA a fost utilizat pe scară largă pentru modelarea și prognozarea multor medii medicale, de mediu, aplicații financiare și de inginerie. În cazul Turciei pentru perioada analizată 1970 – 2015, s-a folosit modelul ARIMA (0, 1, 2) pentru prognoza consumului total de energie. Rezultatele au indicat faptul că consumul de energie al Turciei va continua să crească până la sfârșitul anului 2040. [1]

Atenția din ce în ce mai mare acordată problemelor energetice globale și politicilor internaționale necesare pentru reducerea emisiilor de gaze cu efect de seră au dat un stimulent reînnoit interesului de cercetare în legăturile dintre sectorul energetic și performanța economică la nivel de țară. În centrul acestei probleme se pune întrebarea care variabilă are prioritate față de cealaltă - este consumul de energie un stimul pentru creșterea economică sau creșterea economică duce la consum de energie? În studiile recente, cointegrarea și cauzalitatea Granger au fost utilizate pe scară largă pentru a examina prezența echilibrul pe termen lung și direcția de cauzalitate dintre consumul de energie și creșterea economică. Literatura empirică pe legătura **consum de energie-creștere** au dat rezultate mixte și rezultate adesea contradictorii din cauza setului de date diferit, diferite caracteristici ale țărilor și diferite metodologii econometrice folosite. [2]

Obiectivul principal al lucrării este identificarea unui model de predicție pentru consumul de energie electrică, urmând să examinăm relația pe termen lung dintre consumul de energie electrică și Produsul Intern Brut real per capita, pentru România din 2008 până în 2022.

În ceea ce privește piața energiei electrice în România, se remarcă printr-o continuă dezvoltare, ținând cont de măsurile stabilite la nivel global, având ca prioritate majoră garantarea unui raport corect între consumul de energie și resursele disponibile. Pe fondul dezvoltării industriale mai accentuate din ultimii ani, s-a generat și o creștere semnificativă a consumului energetic. Astfel, una dintre preocupările actuale constă în utilizarea cât mai rațională a energiei electrice, în contextul unor resurse naturale tot mai limitate. [3]

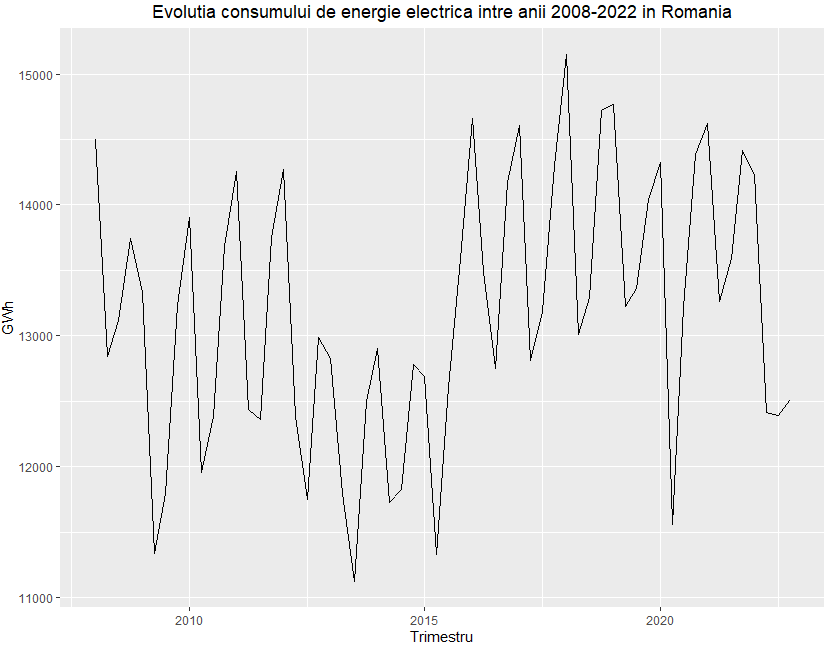
# **Descrierea bazei de date**

Analiza propusă a fost realizată cu ajutorul datelor preluate de pe site-ul Eurostat. Baza de date conține informații trimestriale despre:

* **consumul de energie electrică**, exprimat in GWh din perioada Ianuarie 2008 (Q1 2008)– Decembrie 2022 (Q4 2022). Frecvența de măsurare a datelor este trimestrială. În total au fost înregistrate 60 de observații.
* **Produsul Intern Brut la prețuri curente**, exprimat în milioane de euro din perioada Ianuarie 2008 (Q1 2008)– Decembrie 2022 (Q4 2022). Frecvența de măsurare a datelor este trimestrială. În total au fost înregistrate 60 de observații.

# **Modele cu trend determinist sau stochastic**

Înainte de implementarea propriu-zisă a modelelor pe baza cărora se va realiza predicția consumului de energie electrică, am realizat o scurtă analiză exploratorie grafică a setului de date.



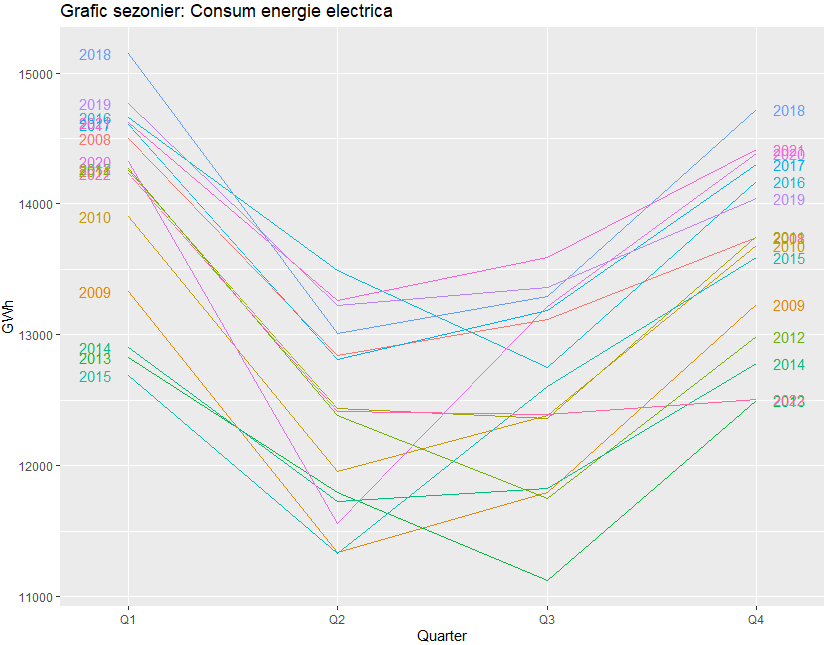
*Figura 1. Evoluția consumului de energie electrică din România în perioada 2008-2022*

Se poate observa că evoluția consumului de energie electrică în România în perioada 2008-2022 nu prezintă trend, evidențiindu-se impactul negativ al crizei economice globale dintre anii 2007-2009 prin scăderea bruscă a consumului.

În ceea ce privește sectorul energetic, de la 1 ianuarie 2015 se liberalizează piața gazelor naturale pentru IMM-uri, iar la energia electrică, preţul mediu de vânzare la tarifele reglementate creşte cu 0,26%.

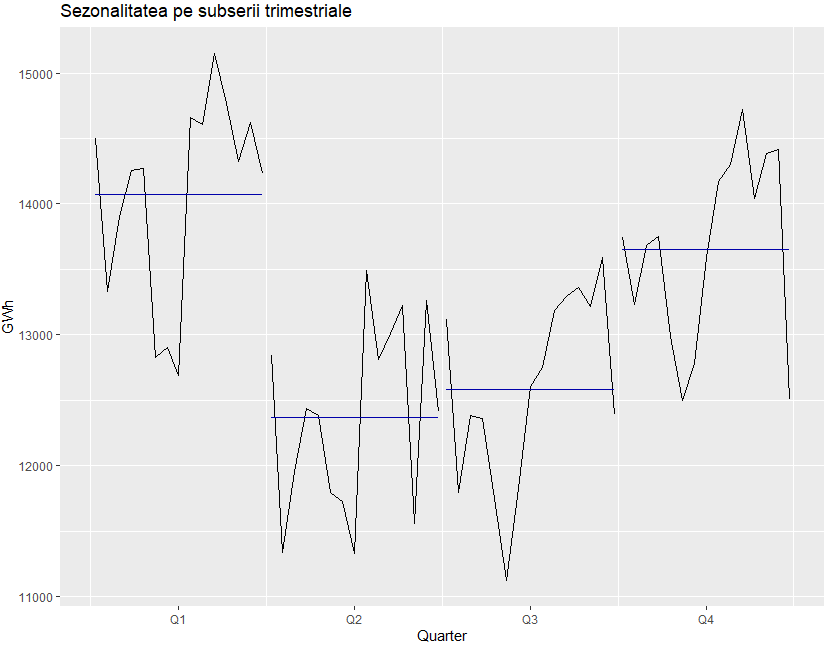
Ca în orice serie de timp care acoperă perioada pandemiei de SARS-CoV-2 este evidențiat impactul acestui virus și în domeniul energetic. Acest impact este negativ și se datorează închiderii a majorității industriilor și companiilor din România din lockdown.

În ceea ce privește sezonalitatea seriei de timp, putem remarca în graficul de mai jos (Fig. 2) consumul ridicat al energiei electrice înregistrat în primul si ultimul trimestru al anului 2018 față de restul perioadei analizate. Totodată, în toți anii se înregistrează o scădere de la Q1 la Q2 și o creștere de la Q3 la Q4.



*Figura 2. Grafic sezonier pentru consumul de energie electrică*

În plus, pe baza graficului (Fig. 3), putem afirma că există sezonalitate în setul de date deoarece media consumului de energie electrică diferă în funcție de fiecare trimestru al anului.



*Figura 3. Sezonalitatea pe subserii trimestriale*

Setul de date nu conține valori lipsă sau date care necesită formatare, deci putem continua analiza prin implementarea modelelor.

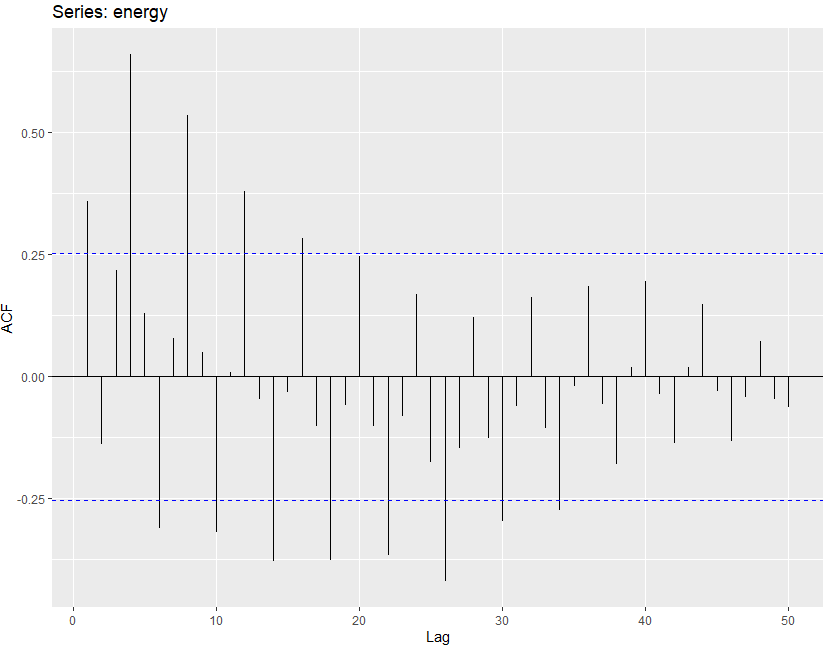
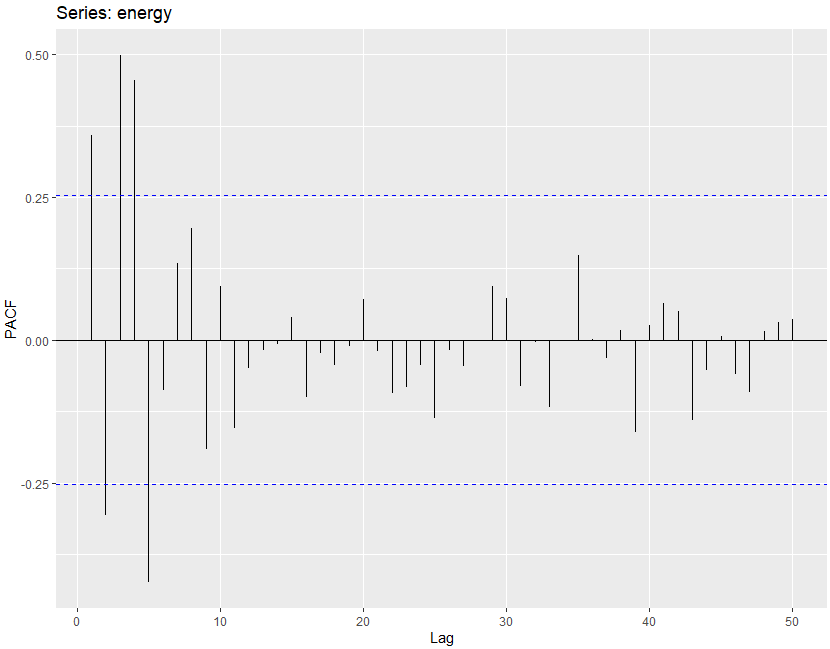
# **Modele univariate**

Înainte de a identifica modele univariate pentru seria de timp a consumului de energie electrică, avem nevoie să cunoaștem dacă seria este staționară și implicit ordinului de integrare al seriei.

Din punct de vedere economic, o serie este staționară dacă un șoc asupra seriei este temporar (se absoarbe în timp). Condițiile ce trebuie îndeplinite pentru ca o serie de timp să fie staționară sunt:

* media să fie constantă, observațiile trebuie să fluctueze în jurul medie;
* varianța să fie constantă.

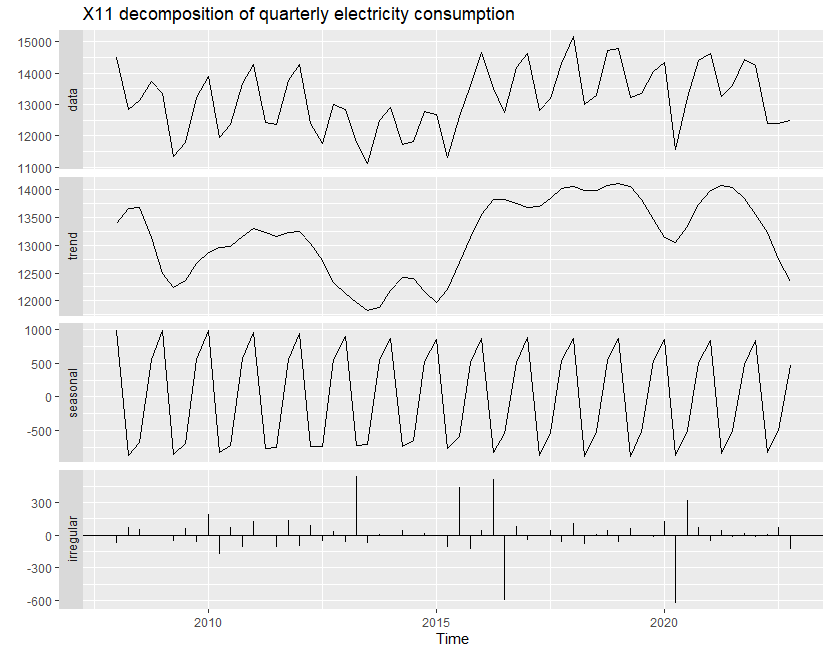
Primul pas în aplicarea metodologiei Box-Jenkins pentru a construi modele ARIMA este identificarea modelului ceea ce implică vizualizarea corelogramei (ACF) și corelogramei parțiale (PACF), alături de testarea staționarității folosind testele de rădăcină unitate: Augmented Dickey-Fuller (ADF), KPSS, Zivot- Andrews. Cel din urmă este pentru a vedea dacă există rupturi în serie și ne ajută să decidem dacă seria este staționară, în caz că celelalte teste au rezultate contradictorii.

*Figura 4. Corelograma ACF și corelograma parțială PACF*

Pe baza corelogramei ACF (Fig. 4) a seriei de date referitoare la consumul de energie electrică, se observă staționaritatea, dar și sezonalitatea trimestrială, din 4 în 4 lag-uri.

**Descompunerea seriei de timp**



*Figura 5. Descompunerea seriei de timp a consumului de energie electrică*

În urma descompunerii seriei, se observă lipsa trendului în seria de date a consumului de energie electrică în perioada 2008-2022, ceea ce înseamnă că seria este staționară. În schimb, este evidentă prezența sezonalității.

## **Testarea staționarității seriei**

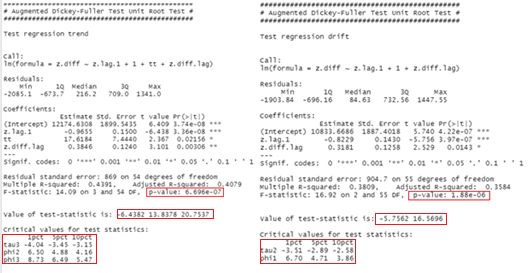
### Testul ADF – Augmented Dickey-Fuller

Testul Augmented Dickey-Fuller (ADF) este o metodă utilizată în analiza seriilor de timp pentru a evalua prezența rădăcinii unității și, implicit, staționaritatea seriei de timp. Testul ADF se concentrează pe detectarea prezenței unei tendințe sau a componentei deterministe în serie, care poate indica faptul că seria de timp nu este staționară. Tipurile testului ADF sunt cele cu trend, drift sau fără trend și fără drift.

**Ipotezele testului:**

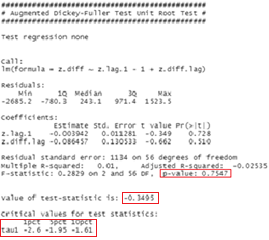
: seria de timp nu este staționară (p-value > 0,05);

: seria de timp este staționară (p-value < 0,05).



*Figura 6. Testul ADF cu trend și cu drift*

În cazul nostru, p-value este mai mic de 0,05 și statisticile calculate ale testului sunt mai mari decât valorile critice, astfel avem destule motive pentru a respinge ipoteza nulă pentru testele cu trend și cu drift, ceea ce înseamnă că pentru un nivel de semnificație de 99% seria de timp este staționară.

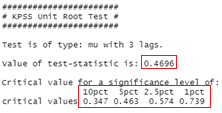


*Figura 7. Testul ADF fără trend și fără drift*

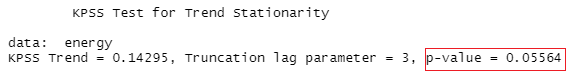
Iar în cazul testului fără trend și fără drift, acceptăm ipoteza nulă conform căreia seria nu este staționară.

### Testul KPSS

Un alt test pentru staționaritatea seriei este testul KPSS, a cărei ipoteze nulă susține faptul că seria este staționară. O diferență majoră între testele KPSS și ADF este capacitatea testului KPSS de a verifica staționaritatea în „prezența unei tendințe deterministe”. Pentru a respinge ipoteza nulă, statistica testului ar trebui să fie mai mare decât valorile critice furnizate sau valoarea lui p-value trebuie sa fie mai mică decât pragul de semnificație. Dacă valoarea p-value este mai mică de 0,05, statistica KPSS va fi mai mare decât valoarea critică de 5%.



*Figura 8. Testul KPSS – statistica testului*



*Figura 9. Testul KPSS – p-value*

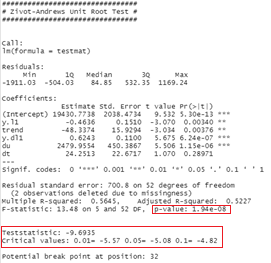
Pe baza rezultatelor din figura 8 și figura 9, valoarea lui p-value este mai mare decât 0.05, dar statisticile testului sunt mai mari decât valorile critice doar pentru un nivel de semnificație mai mic de 0.25.

### Testul Zivot and Andrews

O problemă comună cu testele convenționale de rădăcină unitară - cum ar fi testele ADF și PP, este că acestea nu permit posibilitatea existenței unei ruperi structurale. Această procedură poate identifica data rupturii structurale, facilitează analiza dacă o ruptură structurală a unei anumite variabile este asociată cu un anumit eveniment, cum ar fi o schimbare a politicii guvernamentale, o criză valutară, război și așa mai departe.

: prezența unei rădăcini unitare în seria de timp, seria de timp nu este staționară

: seria de timp este staționară



*Figura 10. Testul Zivot-Andrews*

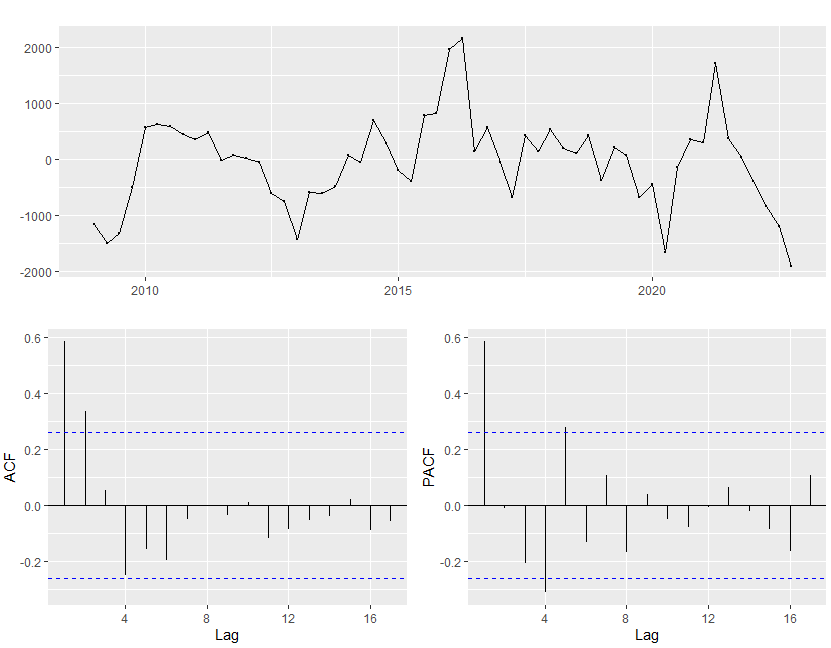
Testul Zivot-Andrews indică faptul că seria este staționară, întrucât valoarea lui p-value este mai mică decât probabilitatea de 0.05 și statistica calculată a testului este mai mare decât valorile critice, astfel încât se respinge ipoteza nulă.



*Figura 11. Grafic pentru testul Zivot-Andrews*

Prin intermediul graficului testului Zivot-Andrews este evidențiată ruptura din serie în trimestrul al 32-lea, mai exact în trimestrul al patrulea al anului 2015.

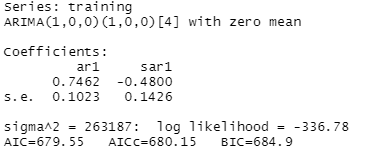
În continuare, am ajustat seria pentru a elimina componenta sezonieră utilizând diferențierea sezonieră de ordinul întâi (seasonal difference of order 1), care implică calculul diferenței între punctele de date consecutive din aceeași perioadă sezonieră. Putem observa în figura 12 că seria rămâne staționară după ajustarea sezonieră conform corelogramei ACF, astfel încât sezonalitatea nu mai este evidentă.



*Figura 12. Seria fără componenta sezonieră*

## **Modelul ARIMA**

Inițial, pentru identificarea celui mai bun model ARIMA, am aplicat in R, funcția „auto.arima()” fără diferențiere, deoarece seria noastră este staționară, obținând modelul ARIMA(1,0,0)(1,0,0)[4].



*Figura 13. Modelul SARIMA obținut cu funcția auto.arima*

În plus, pe baza graficele funcțiilor ACF și PACF prezentate mai sus în figura 13, am ales să elaborăm diferite modele ARIMA și SARIMA cu ajutorul funcției „arima()”.

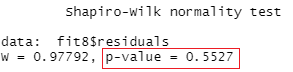
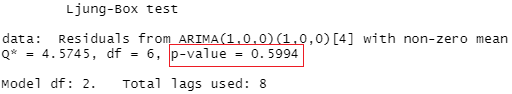
Pentru a selecta cel mai bun model potrivit pentru previziune, vom alege modelul cu cele mai mici valori ale criteriilor AIC (Akaike Information Criterion), BIC(Schwarz criterion), log likelihood și RMSE (Root Mean Square Error) obținute cu funcția „summary()”.

Tabelul nr.1. Testarea modelelor ARIMA și SARIMA

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **log** | **AICc** | **BIC** | **RMSE** |
| ARIMA (1,0,1) | -341.38 | 691.78 | 697.89 | 562.9446 |
| ARIMA (1,02) | -338.93 | 689.44 | 696.78 | 495.8917 |
| ARIMA (1,0,0) | -341.38 | 689.37 | 694.12 | 562.9946 |
| ARIMA (0,0,1) | -343.89 | 694.38 | 699.13 | 595.9024 |
| ARIMA (0,0,2) | -341.16 | 691.46 | 697.46 | 525.0264 |
| **SARIMA (1,0,0) (1,0,0)** | **-336.78** | **682.58** | **688.69** | **501.2201** |
| SARIMA (1,0,1) (1,0,0) | -336.76 | 685.1 | 692.44 | 501.0676 |
| SARIMA (0,0,1) (1,0,0) | -342.24 | 693.5 | 699.61 | 571.5735 |

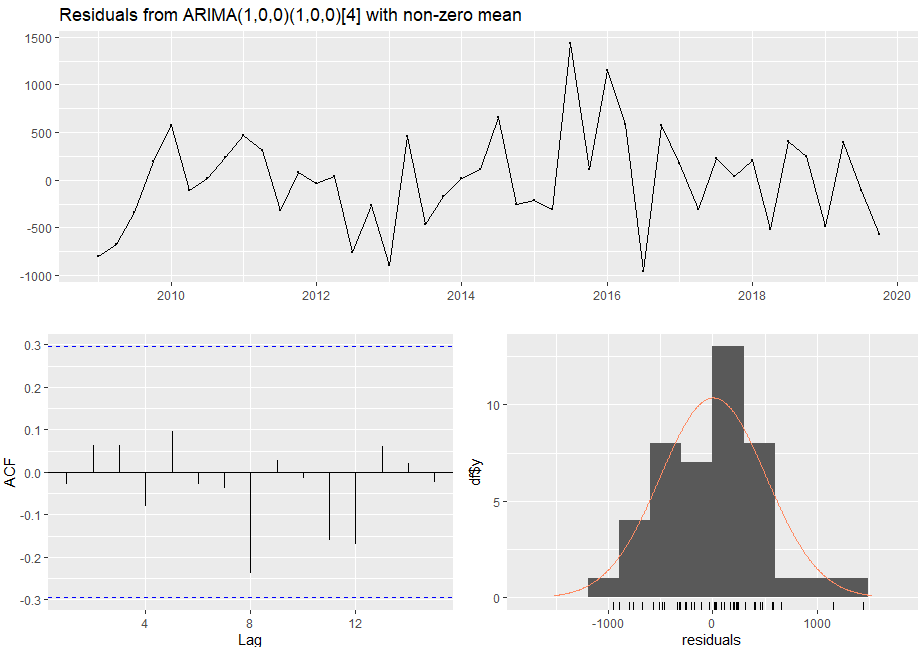
În tabelul nr. 1, observăm că cele mai potrivite valori ale criteriilor corespund modelului SARIMA (1,0,0) (1,0,0) identificat ca fim optim și de funcția „auto.arima()”.

Mai departe, înainte de a realiza predicția consumului de energie electrică, vom analiza grafic reziduurile modelului ales și vom aplica testul Ljung-Box.



*Figura 14. Testul Ljung-Box și testul Shapiro-Wilk*

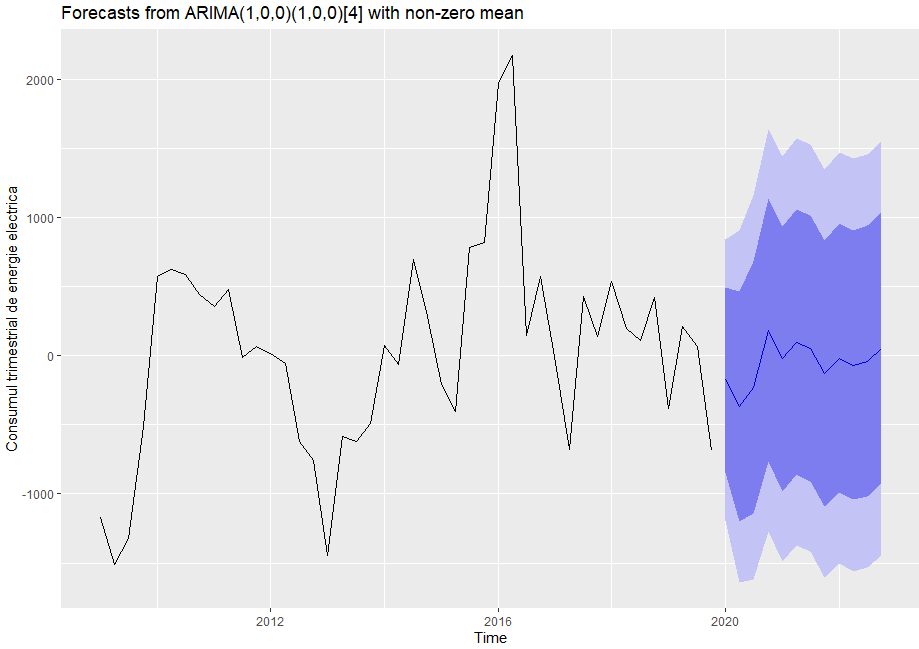
Conform testului Ljung-Box, erorile nu sunt autocorelate pentru un nivel de încredere de 95% cu p-value de 0.5994 mai mare decât 0.05. Pe baza testului Shapiro-Wilk, concluzionăm că erorile sunt normal distribuite deoarece valoarea lui p-value este de 0.5527, mai mare decât 0.05.



*Figura 15. Reprezentarea grafică a reziduurilor*

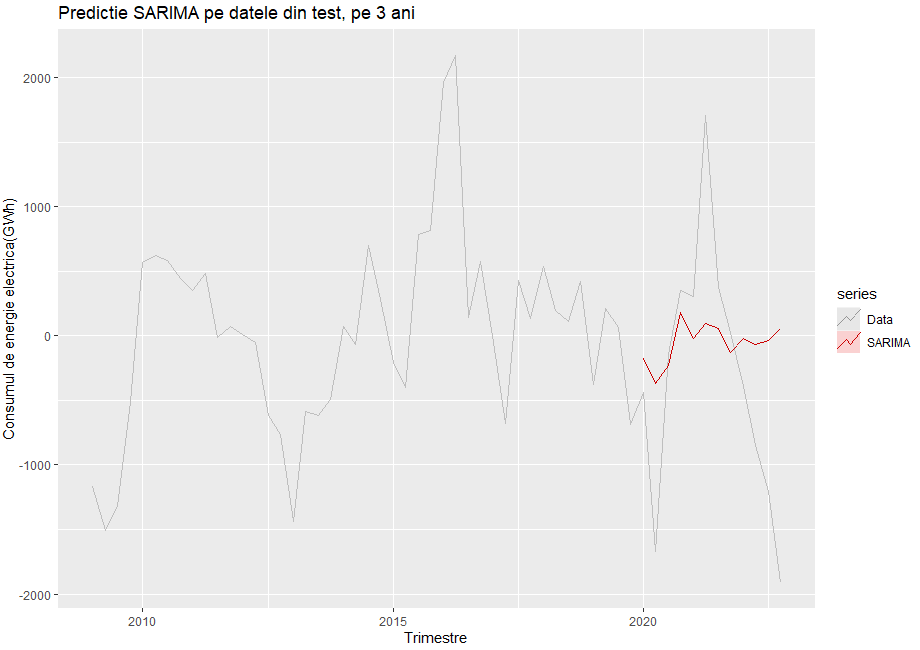
Observăm în figura 15 faptul că toate autocorelațiile sunt în limitele de prag, indicând că reziduurile se comportă ca un zgomot alb, deci nu sunt corelate. În plus, reziduurile urmează cu aproximare curba lui Gauss.

Folosind funcția „forecast()”, am realizat o predicție pentru următorii 3 ani, din anul 2020.



*Figura 16. Predicția pentru modelul SARIMA*

Următorul pas a fost să suprapunem predicția obținută pe datele de test din seria de timp pentru a analiza performanța prognozei în grafic.



*Figura 17. Suprapunerea predicției cu datele reale*

 În figura 17, remarcăm o diferență semnificativă între prognoza realizată și datele reale. Din păcate, efectul pe care pandemia l-a avut asupra consumului de energie electrică nu a putut fi prevăzut. Însă fluctuațiile seriei de date este urmărită cu ajutorul prognozei, dar nu într-o mare măsură.

*Figură 18. Acuratețea modelului*

Una dintre cele mai frecvente valori folosite pentru a măsura acuratețea prognozei a unui model este eroarea procentuală medie absolute (MAPE). În cazul nostru, diferența medie dintre valoarea prognozată și valoarea reală este de 150,68%, o valoare foarte ridicată.

## **Modele ARCH - GARCH**

Procesul ARCH introdus de Engle (1982) recunoaște în mod explicit diferența dintre variația necondiționată și cea condiționată, permițând acesteia din urmă să se schimbe în timp în funcție de erorile trecute. Acesta este compus din trei elemente:

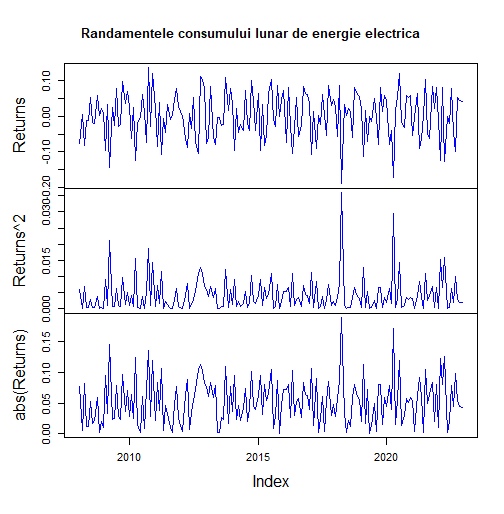
* **Autoregresiv**: Valoarea curentă poate fi exprimată în funcție de valorile anterioare, adică sunt corelate;
* **Condițional:** Informează că varianța se bazează pe erori din trecut;
* **Heteroschedasticitate**: Arată că seria prezintă o variație neobișnuită (varianță variabilă).

Modelele autoregresive pot fi dezvoltate pentru date univariate din serii de timp care sunt staționare (AR), au o tendință (ARIMA) și au o componentă sezonieră (SARIMA). De asemenea, este important ca termenii eroare din procesele stochastice care generează seria temporală să fie homoscedastici, adică să aibă o varianță constantă.

În practică, modelele ARCH sunt utilizate pentru că modelele şi seriile liniare de timp sunt incapabile să explice tendinţa de apariţie în masă a volatilităţii pe pieţele financiare, spre exemplu. Astfel rentabilitățile mari (de orice semn) au tendinţa de a apărea în urma altor rentabilităţi mari, iar rentabilitățile mici (de orice semn) au tendinţa de a apărea în urma rentabilităţilor mici. Una dintre explicațiile pentru acest fenomen care pare a caracteriza seriile de rentabilităţi financiare ar fi faptul că informațiile care pot influenţa schimbările de preţ vin la rândul lor în masă.

Pentru acest model am folosit randamentul logaritmic al consumului de energie electrică lunar din 2008 până la finalul lui 2022. De asemenea, aplicând testul ADF am putut demonstra că seria este staționară.

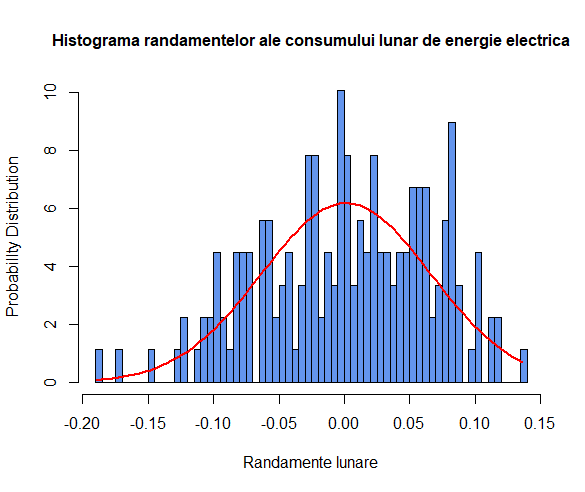
În continuare, vom verifica prezența volatilității utilizând graficul serii cronologice pentru randamente, randamente ridicate la pătrat și randamente absolute.



*Figura 19. Randamentele consumului lunar de energie electrică*

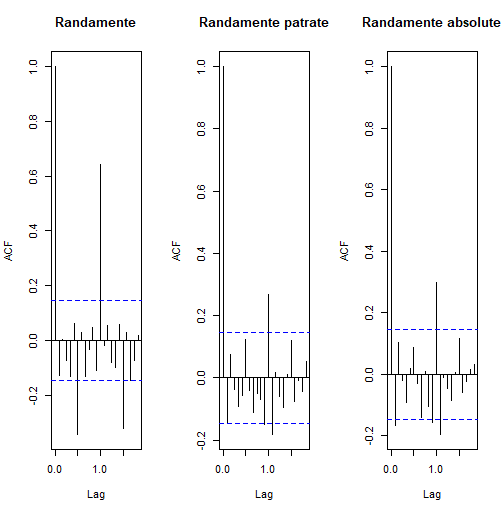
Se poate observa, în figura 18, cum varianța se modifică constant în timp cu precădere la începutul anului 2018 și la începutul pandemiei de Covid-19. Randamentele ilustrate în graficul de mai sus surprind și clusterizarea volatilității, după cum se pot observa, șocurile sunt grupate în jurul punctelor extreme.

O imagine care conține text, Font, captură de ecran, linie

Descriere generată automat****

*Figura 20. Histograma randamentelor și testul Jarque-Bera*

Conform testului Jarque-Bera (figura 19), ipoteza normalității reziduurilor este acceptată pentru un nivel de încredere de 95%.



*Figura 21. Corelograma ACF pentru cele trei randamente*

În ceea ce privește autocorelația seriilor, putem observa în corelogramele ACF din figura 20 că nu există o autocorelație puternică în niciunul din cazuri.

Vom efectua o a doua verificare a prezenței auto-corelației în randamentele pătrate prin aplicarea testului Ljung-Box pe randamentele ridicate la pătrat și absolute.



*Figura 22. Testul Box-Ljung pentru randamente*

Cum p-value > 0,05, datele sunt independente. Acceptăm ipoteza nulă, înseamnă că nu este prezentă autocorelația atunci când ne referim la randamentele ridicate la pătrat sau randamentele absolute.

În continuare, am testat efectul ARCH folosind ARCH-LM test și am obţinut următorul output:

O imagine care conține text, Font, captură de ecran, linie

Descriere generată automat

*Figura 23. ARCH-LM test pentru seria inițială*

La o semnificație de 5% se respinge ipoteza nulă, seria randamentelor simple având un efect ARCH (1). Astfel că, vom estima modelele ARCH-GARCH.

De asemenea, în încercarea găsirii celui mai potrivit model pentru estimarea seriei randamentelor consumului de energie electrică am eliminat componenta sezonieră, însă rezultatul (figura 24) nu a fost cel așteptat astfel încât prin intermediul testului ARCH-LM putem susține că seria randamentelor fără prezența componentei sezoniere nu are efecte ARCH.

O imagine care conține text, Font, captură de ecran, linie

Descriere generată automat

*Figura 24. ARCH-LM test pentru seria fără componenta sezonieră*

### Modelul ARCH

Modelul ARCH (autoregressive conditional heteroskedasticity) a fost propus pentru modelarea corelației seriale din reziduurile pătratice, sau heteroschedasticitate. Modelul lui Engle avea așadar următoarea formă:

, unde reprezinta valoarea așteptată condiționată de informația disponibilă la momentul t-1, iar este o secvență de variabile aleatoare independente și identic distribuite (iid) cu media zero si varianța unitară.

Abordarea de tip ARCH ia în considerare, în elaborarea previziunilor, şi informaţiile conţinute în varianţa condiţionată a procesului (momentul condiţionat de ordinul doi). De asemenea, poate fi folosită atât pentru serii staționare în medie cât si nestaționare în medie. Însă, modelul de tip ARCH este specific seriilor cu varianță variabilă în timp (nestaţionare relativ la varianţă).

O imagine care conține text, captură de ecran, Font, număr

Descriere generată automat

*Figura 25. Semnificația coeficienților modelului GARCH(1,0)*

Conform rezultatului din R (Anexa 1), realizat cu ajutorul pachetului rugarch(), a fost expus parametrul estimat al modelelor ARIMA(0,0,0) și GARCH(1,0). Conform acestora, modelul are coeficienții semnificativi la 1% (Figura 25), observându-se și valoarea varianței așteptate la momentul t pentru valorile de varianță reziduale zero și întârziate.

Apoi, sunt date câteva criterii de informare pentru a fi utilizate pentru selectarea celor mai bune modele GARCH pentru seria randamentelor consumului de energie electric. Modelul care are cele mai mici criterii de informare este mai bun decât celelalte.

Rezultatele testului Ljung-Box arată că reziduurile nu au autocorelație, rezultând același lucru și în cazul reziduurilor ridicate la pătrat. În plus, rezultatele testul ARCH LM arată că procesul GARCH este adecvat deoarece valorile lui p-value sunt semnificativ mai mari decât probabilitatea de 5%.

Totodată, Sign Bias Test este utilizat pentru a testa efectul de levier în reziduurile standardizate. (Ipoteza nulă: nu există șocuri de reacție pozitive și negative semnificative (dacă există vom estima folosind si modele de tip AP-ARCH)).

### Modelul GARCH

Heteroskedasticitatea condiționată autoregresivă generalizată, sau GARCH, este o extensie a modelului ARCH care încorporează o componentă medie mobilă împreună cu componenta autoregresivă. Bollerslev (1986, Journal of Econometrics) a generalizat modelul ARCH al lui Engle și a introdus modelul GARCH. Introducerea componentei medii mobile permite modelul:

* Pentru a modela modificarea condiționată a varianței în timp;
* Modificări ale variației dependente de timp.

Pentru ca parametrii GARCH (βj) să poată fi estimați, este necesar ca cel puțin unul dintre parametrii ARCH (αi) să fie diferiți de zero.

Când q = 0, modelul GARCH se reduce la modelul ARCH. Termenii GARCH (β) reprezintă persistența șocurilor trecute asupra volatilității sau persistența impactului veștilor trecute asupra volatilității. Studiile empirice au arătat că cel mai reprezentativ pentru modelarea seriilor financiare este un model simplu GARCH (1,1).

Modelul standard GARCH (1,1) propune așadar următoarea ecuație pentru varianța condiționată:

O imagine care conține text, captură de ecran, Font, număr

Descriere generată automat

*Figura 26. Semnificația coeficienților modelului GARCH(1,1)*

În acest caz, au fost estimați parametrii modelelor ARIMA (0,0,0) și GARCH (1,1). Singurul parametru semnificativ al acestui model este beta (Figura 26). Conform criteriilor informaționale Akaike = -2.6, Bayes = -2.53 și Hanna-Quinn = -2.58 rezultatele detaliate regăsindu-se în anexă (Anexa 2). Altfel spus, acest model este mai performant decât primul.

Rezultatele testelor Ljung Box arată că reziduurile nu au autocorelație, rezultând același lucru și în cazul reziduurilor ridicate la pătrat pentru o probabilitate de 10%, iar rezultatele testului ARCH LM arată că procesul GARCH este adecvat deoarece valorile lui p-value sunt semnificativ mai mari decât probabilitatea de 5%.

Pe baza testului Sign Bias, concluzionăm că există un efect de pârghie, așadar se poate potrivi și un tip de model mai specific, și anume AP-ARCH. Pentru alegerea unui model mai performant, vom estima si parametrii **modelelor** **E-GARCH și AP-ARCH.**

### Modelul E-GARCH

**Modelul E-GARCH** specifică varianta condiționată sub forma logaritmică, ceea ce înseamnă că nu mai este nevoie sa fie impuse constrângeri asupra estimărilor pentru a fi evitată varianta negativă.

O imagine care conține text, captură de ecran, Font, document

Descriere generată automat

*Figura 27. Semnificația coeficienților modelului E-GARCH(1,1)*

Modelul folosit este EGARCH (1,1) și prezintă următoarele caracteristici ale modelului, modelul este valid (Figura 27), valoarea lui p value fiind > 0.05 pentru testul Ljung-Box și astfel respingând ipoteza nulă a testului pentru reziduuri. Iar de asemenea, pentru testul ARCH LM acceptăm ipoteza nulă și putem afirma că există efectul ARCH este prezent în reziduuri, iar modelul are un coeficient AIC de -2.66 (Anexa 3).

### Modelul AP-ARCH

În cazul primului model estimat, și anume ARCH (1,0) și ARIMA (0,0,0) a rezultat că există impact asupra volatilității din partea șocurile pozitive și negative, astfel că vom estima și un model de tip AP-ARCH (Asymmetric Power ARCH).

Deși în cadrul estimării modelului GARCH (1,0) a rezultat posibilitatea estimării pe baza modelului AP-ARCH, când am aplicat AP-ARCH (1,0) modelul nu a fost semnificativ, astfel că vom schimb ordinele cu AP-ARCH (1,1).

*O imagine care conține text, captură de ecran, Font, document

Descriere generată automat*

*Figura 28. Semnificația coeficienților modelului AP-ARCH(1,1)*

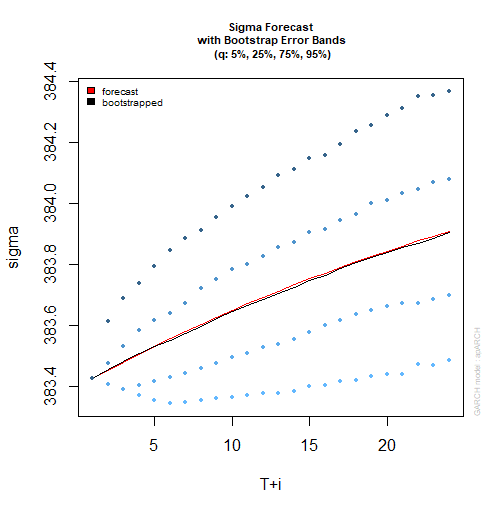
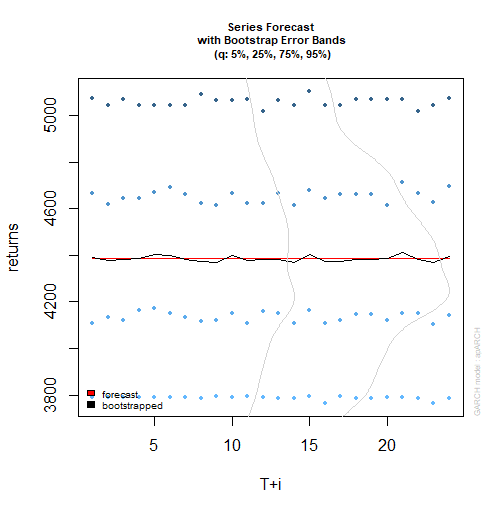
Modelul folosit este AP-ARCH(1,1) și prezintă următoarele caracteristici ale modelului, termenii delta și beta sunt semnificativi, valoarea lui p value fiind >0.05 pentru testul Ljung-Box și astfel respingând ipoteza nulă a testului pentru reziduuri. Iar de asemenea, pentru testul ARCH LM respingem ipoteza nulă și putem afirma că există efectul ARCH este prezent în reziduuri, iar modelul are un coeficient AIC de -2.58. Output-ul este anexat la finalul documentului (Anexa 4).

Cu scopul comparării tuturor modelelor din cadrul modelelor ARCH-GARCH vom analiza valorile criteriilor de informare (de exemplu, AIC sau BIC), concluzionând că modelul **EGARCH(1,1)** este mai performant decât *GARCH (1,0), GARCH (1,1) și AP-ARCH (1,1),* atunci când ne dorim analizarea consumului de energie electrică în România în perioada 2008-2022.

Tabelul nr.2. Testarea modelelor

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **ARCH(1,0)** | **GARCH (1,1)** | **EGARCH (1,1).** | **AP-ARCH (1,1)** |
| **Akaike** | 74,066 | -2,654 | **-2,654** | -2,581 |
| **Bayes** | 74,119 | -2,529 | **-2,565** | -2,474 |
| **Shibata** | 74,065 | -2,601 | **-2,655** | -2,583 |
| **Hannan - Quin** | 74,088 | -2,571 | **-2,618** | -2,537 |

Există două opțiuni de prognoză în acest pachet. Metoda Rolling și Bootstrap. Metodele bootstrap se bazează pe reeșantionarea reziduurilor standardizate din distribuția empirică a modelului adaptat pentru a genera realizări viitoare ale seriei și sigma. Cu alte cuvinte, cu prognoza bootstrap, putem prognoza atât serie, cât și variații condiționate.



*Figura 29. Diagrame de eroare standard de serie și standard Sigma*

Pe baza graficelor de predicții din figura 29 se poate observa că modelele pot surprinde în mare măsură volatilitatea seriei. Deși prezice într-o oarecare măsura tendinței, acesta nu poate prezice cu exactitate șocurile ce vor apărea, modelul considerând șocurile aproape inexistente.

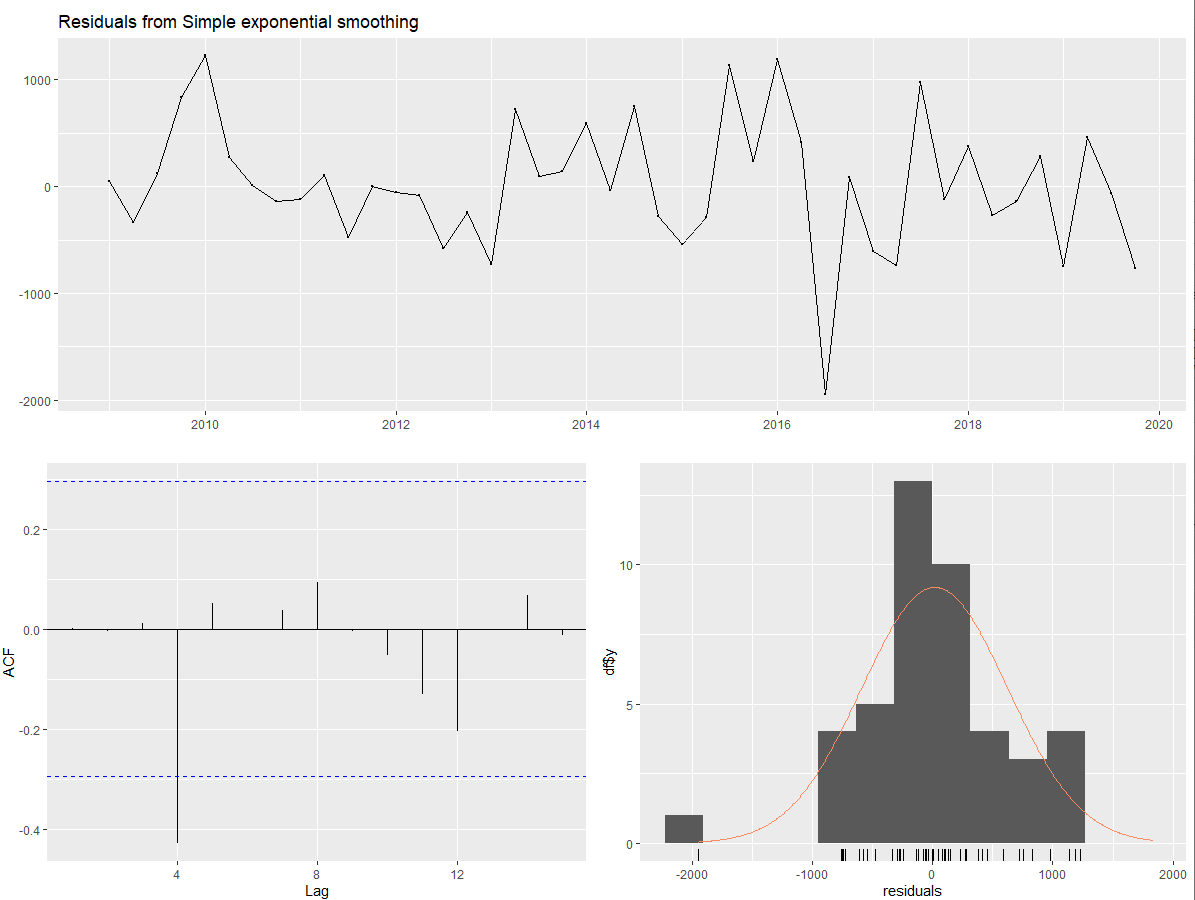
## **Netezirea exponențială**

Metoda de netezire exponențială este utilizată pentru a prognoza serii de timp univariate, care poate cuprinde și date ce prezintă trend sau sezonalitate. Exponential Smoothing (ETS) este o tehnică de predicție propusă la sfârșitul anilor 1950, care stă la baza unor metode de predicție utilizate pe scară largă. Acest cadru permite generarea rapidă a predicțiilor pentru o selecție mare de serii de timp ce au fost aplicate cu succes în industrie.

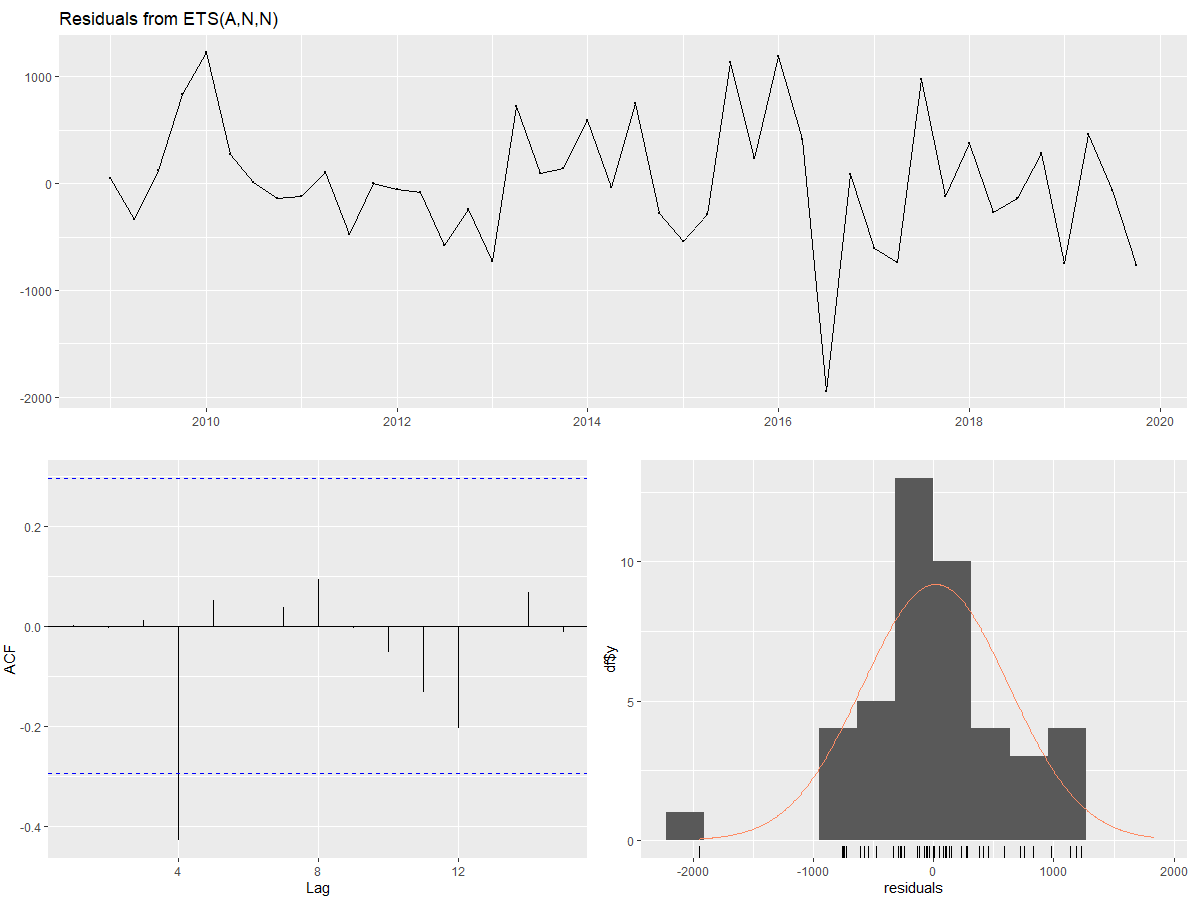
Predicțiile generate de acest model sunt, în esență, medii ponderate, unde observațiile recente primesc o pondere mai mare decât cele din trecut. Această ponderare inegală a observațiilor este posibilă datorită utilizării constantelor de netezire. Tehnicile de netezire sunt utilizate pentru a genera valori netezite (eliminând fluctuațiile aleatoare ε din date) și pentru a obține predicții. Valoarea netezită corespunzătoare unei observații este notată . Principalele avantaje ale acestei metode sunt reprezentate de simplitatea procedurii de calcul și capacitatea de a lua în considerare ponderea informațiilor inițiale.

Netezirea exponențială simplă (SES) reprezintă cea mai simplă abordare utilizată ce necesită un singur parametru denumit alfa (α), cunoscut și sub numele de *factor de netezire sau coeficient de netezire*. Acest parametru controlează rata la care influența observațiilor din pașii de timp anterioare scade exponențial. De obicei, α este setat la o valoare cuprinsă între 0 și 1. Valorile mari indică faptul că modelul acordă o atenție predominantă celor mai recente observații, în timp ce valorile mai mici înseamnă că se ia în considerare o parte mai mare din istoric atunci când se fac predicții.

Modelul Holt-Winters este alcătuit din ecuația de predicție și trei ecuații de netezire – una pentru nivelul , una pentru trendul și una pentru componena sezonieră , cu parametrii de netezire corespunzători. Pentru a utiliza acest model este necesar ca seria să aibă trend și sezonalitate. Seria din modelul nostru nu prezintă trend, din acest motiv nu putem aplica acest testul.

**

*Figura 30. Reziduurile pentru Simple Exponential Smoothing*



*Figura 31. Reziduurile pentru modelul ETS*

Prin efectuarea testului Ljung-Box, pentru modelele SES și ETS, am obținut un p-value egal cu 0,26 mai mare decât 0,05, ceea ce înseamnă că acceptăm ipoteza nulă care indică prezența autocorelației în reziduri.

După ce am verificat distribuția și autocorelația reziduurilor pentru ambele modele, am estimat acuratețea lor, astfel:

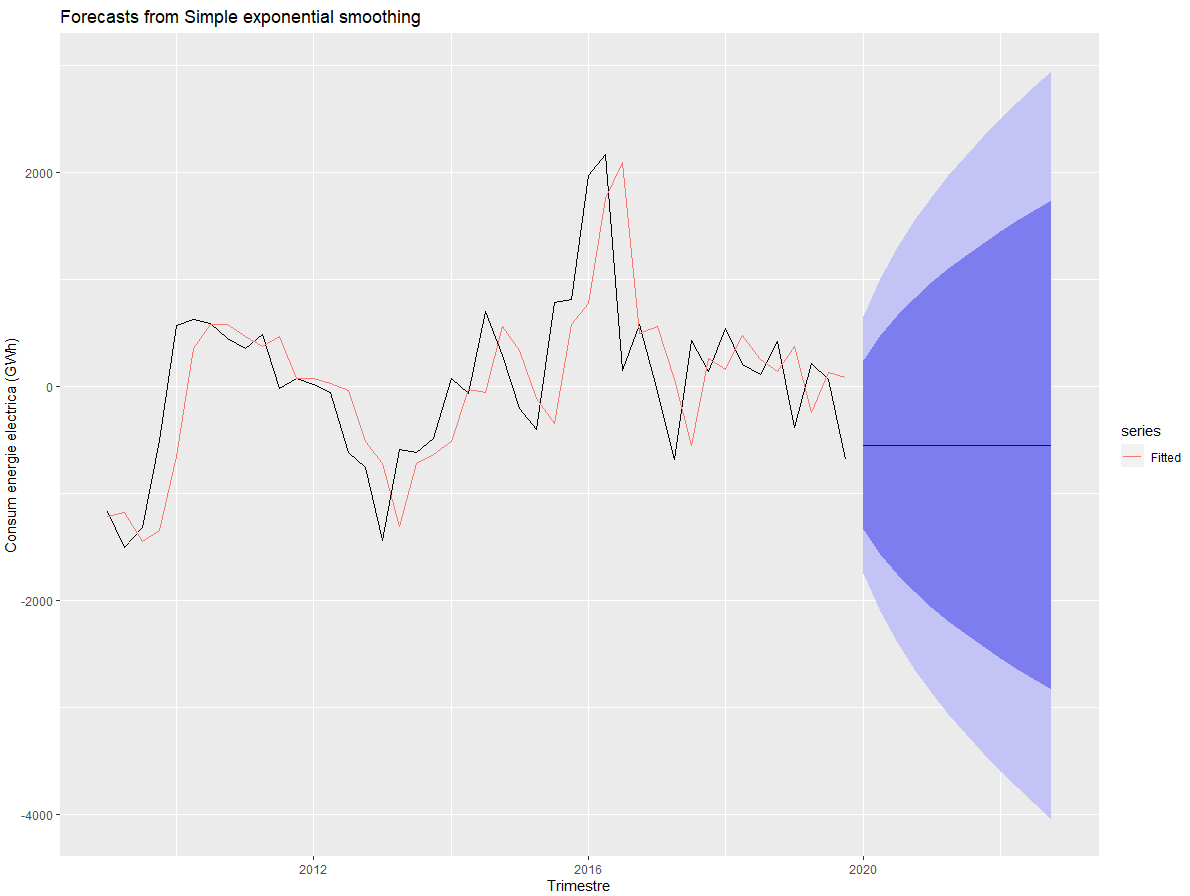


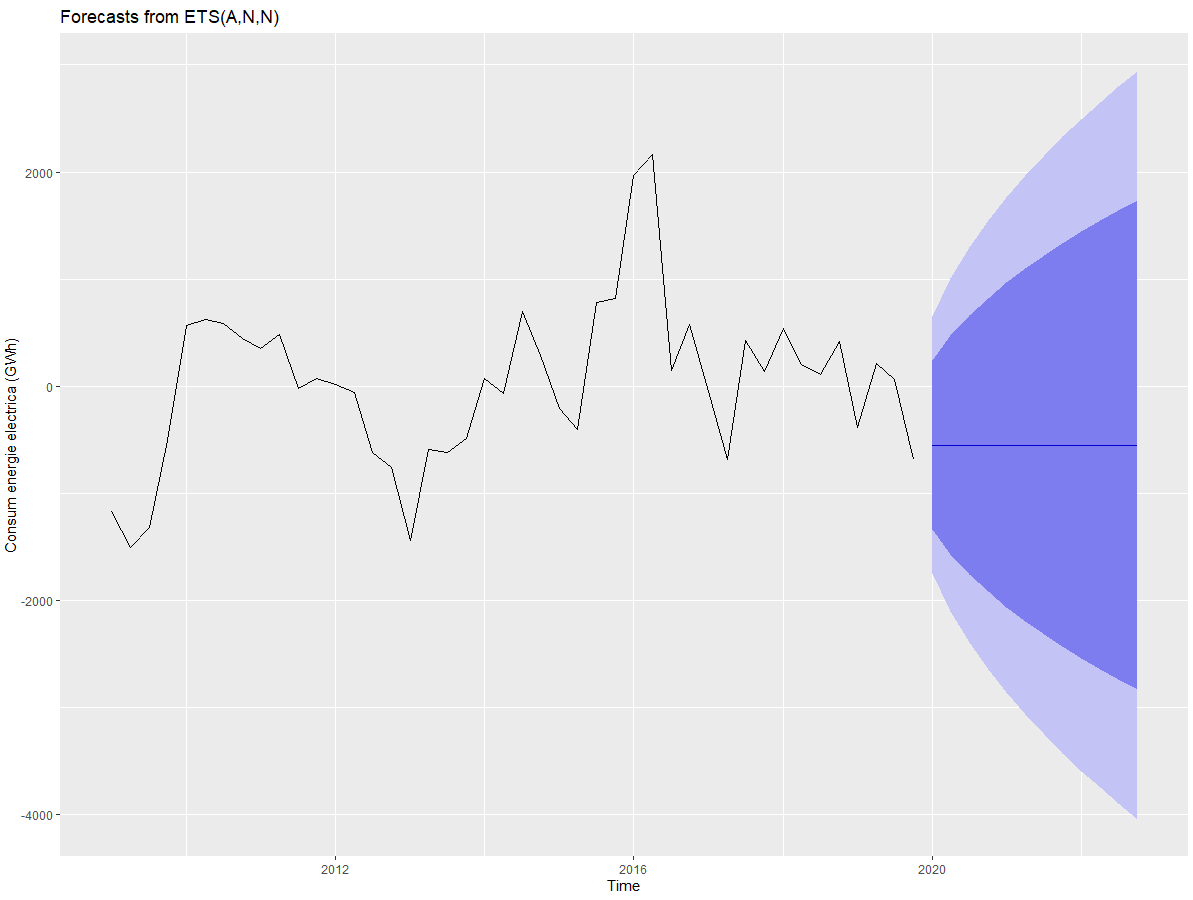
*Figura 32. Acuratețea modelului SES*



*Figura 33. Acuratețea modelului ETS*

Datorită distribuției reziduurilor ilustrate mai sus putem spune că cele două modele sunt potrivite pentru a fi luate în considerare având în vedere predicția modelului. Deoarece valoarea RMSE este egală în cele două modele le vom considera pe amândouă potrivite pentru seria noastă.

  
*Figura 34. Previziunea modelului SES*



*Figura 35. Previziunea modelului ETS*

În schimb, analizând pe baza graficelor previziunile (Figura 34, Figura 35) atât tendințele, fluctuațiile cât și variațiile prezise de model nu urmează comportamentul real al seriei de timp, întrucât rezultatul tehnicilor de netezire prezintă o linie dreaptă.

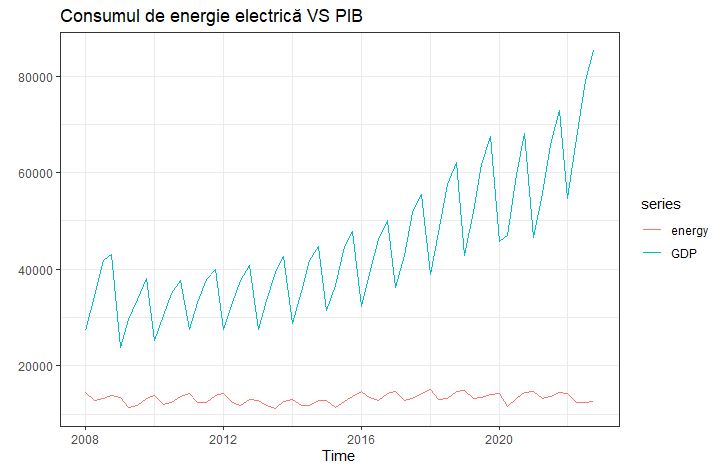
# **Modele multivariate**

## **VAR – Analiza relației dintre produsul intern brut și randamentul consumului de energie electrică**

În literatura de specialitate, creșterea economică s-a dovedit fiind drept principalul motor al cererii de energie și numai economiile avansate cu un grad ridicat de capacitate de inovare pot reduce consumul de energie fără a reduce creșterea economică. În plus, există multe studii care investighează puterea legăturii structurale dintre energie și creșterea economică, folosind analiza seriilor cronologice pentru anumite țări și seturi de date de tip panel. [2]

Pentru analiză, am ales ca factor economic Produsul Intern Brut per capita. Seria de timp este trimestrială, datele având același interval ca și seria inițială, perioada cuprinsă între anul 2018 și anul 2022.

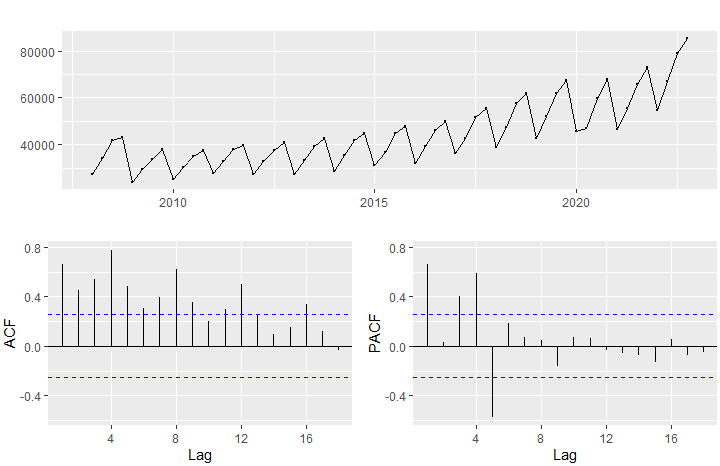
Modelul **VAR** (autoregresia vectorială) este un model statistic utilizat în domeniul economic și financiare pentru a surprinde relația dintre două sau mai multe serii de timp. Pentru un model VAR(p), primele p decalaje ale fiecărei variabile din sistem vor fi folosite ca predictori de regresie pentru fiecare variabilă.



*Figura 36. Graficul seriilor*

Din reprezentarea grafică a celor două serii (figura 30), observăm că seria reprezentată de PIB per capita prezintă fluctuații mai puternice decât cele ale seriei consumului de energie electric.

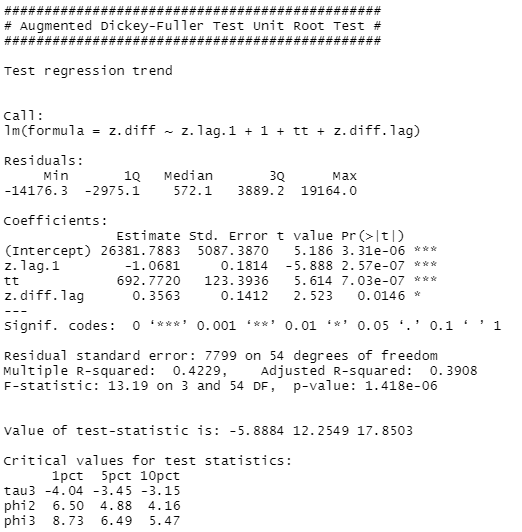
### Analiza staționarității seriei de timp privind Produsul Intern Brut per capita



*Figura 37 .Corelograma ACF și corelograma PACF pentru variabila PIB*

În figura 37, observăm în graficul de tip linie prezența unui trend crescător, iar în corelograma ACF pare că seria GDP este nestaționară deoarece valorile autocorelațiilor nu scad brusc către valoarea zero.

Mai departe, pentru a testa staționaritatea seriei GDP, am aplicat testul Augmented Dickey-Fuller (figura 38) și am obținut un p-value mai mic decât 0,05, ceea ce înseamnă că acceptăm ipoteza alternativă conform căreia seria noastră este staționară. În plus, pentru fiecare prag de semnificație, valorile statisticii calculate ale testului sunt mai mari decât valorile critice.



*Figura 38. Testarea staționarității pentru variabila PIB*

### Cointegrarea seriilor

Cointegrarea este un concept de analiza a seriilor de timp care se referă la relația de lungă durată între doua sau mai multe variabile. În esență, cointegrarea indică faptul că doua variabile se mișcă împreună în timp, în ciuda fluctuațiilor lor pe termen scurt. Acest lucru este util în previziunea și modelarea datelor economice și financiare.

Anterior, am concluzionat faptul că atât seria trimestrială consumului de energie electrică, cât și seria trimestrială Produsului Intern Brut per capita în perioada 2018-2022, prezintă staționaritate, adică sunt integrate de ordinul 0.

Vom estima un model VAR în nivel, pentru care lagul optim se va determina pe baza criteriilor informaționale (AIC, SC, HQ, FPE), dar astfel încât testul F să fie semnificativ din punct de vedere statistic. În R Studio, cu ajutorul funcției VARselect, am obținut numărul de 5 lag-uri.

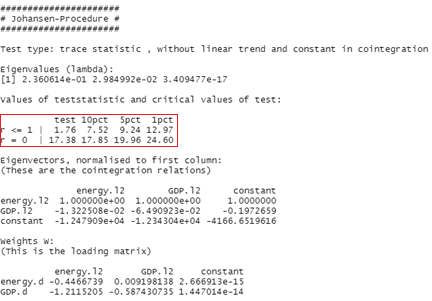
Deoarece variabilele sunt integrate de același ordin, putem folosi procedura Johansen pentru verificarea cointegrării datelor. Abordarea multivariată Johansen cuprinde două metode: Trace si Eigenvalue, unde r reprezintă numărul de cointegrări.

**Testul Johansen**

*Metoda trace* are următoare ipoteze:

: vectori de cointegrare ≤ r;

: vectori de cointegrare ≥ r + 1.



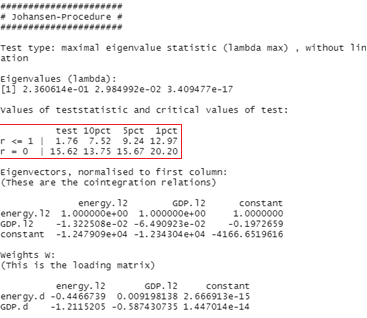
*Figura 39. Testul Johansen – metoda trace*

Observăm în figura 39, atât în prima ipoteză r = 0, cât și în a doua ipoteză r <=1, faptul că statistica testului 17,38, respectiv 1,76, nu depășește valorile critice pentru cele trei nivele de semnificație, concluzionând că cele două ipoteze nu pot fi respinse, deci nu există nicio cointegrare. Dar, pentru prima ipoteză, statistica testului este foarte aproape de valoarea critica pentru nivelul de încredere de 90%.

*Metoda Eigenvalue* are următoare ipoteze:

: Există r vectori de cointegrare;

: Există r + 1 vectori de cointegrare.



*Figura 40. Testul Johansen – metoda* Eigenvalue

Conform metodei Eigenvalue (figura 40, putem susține că există cel puțin o cointegrare deoarece pentru un nivel de încredere de 90%, statistica calculată a testului este 15,62 mai mare decât valoarea critică. Chiar și pentru nivelul de semnificație de 5%, valoarea statisticii este foarte apropiată de valoarea critică.

### Analiza modelului VAR propriu-zis

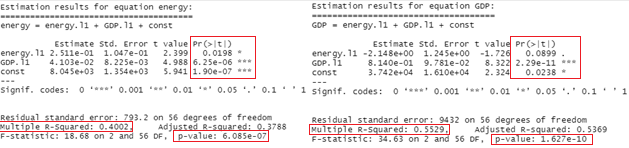
Analiza modelului VAR este structurată în două ecuații principale, unde fiecare indicator – randamentul consumului de energie și Produsul Intern Brut per capita –, va fi, pe rând, variabilă dependentă.

Modelul VAR are forma generală:



, unde:

* *Xt, Xt-1, …, Xt-p*sunt vectorii  de dimensiune () ai valorilor curente și decalate ale celor *n* variabile dependente care sunt I(1) în cadrul modelului;
* *A1,….,Ap*sunt matricele coeficienților de dimensiuni*();* *μ* este vectorul termenilor liberi de dimensiune;
* p reprezintă numărul optim de laguri;
*  este vectorul erorilor independente și identic distribuite de medie zero și matrice de varianță .



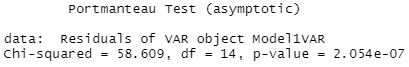
*Figura 41. Output model VAR*

Pentru prima ecuație din figura 41, în care variabila dependentă este consumul de energie electrică, toți parametrii sunt semnificativi statistic, coeficientul de determinare are o valoarea mică egală cu 40%. În ceea ce privește a doua ecuație cu PIB ca variabila dependentă, coeficienții sunt, de asemenea, semnificativi statistic, coeficientul de determinare are o valoarea mai mare egală cu 55,29%. Ambele modele poate fi considerate valide deoarece valoarea lui p-value este mai mică decât 0,01. În plus, nu putem susține prezența unor regresii false, R-Squared având valori medii.

### Testarea ipotezelor pentru verificarea diagnosticului pe reziduuri

Pentru a realiza prognoza pe baza modelului VAR, este necesară testarea și validarea cu success a ipotezelor pe reziduuri.

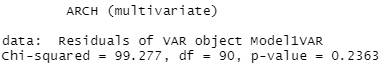
**Autocorelarea erorilor**



*Figura 42. Portmanteau Test*

La un nivel de semnificație de 1%, se respinge ipoteza nulă, concluzionând că există autocorelare în erori.

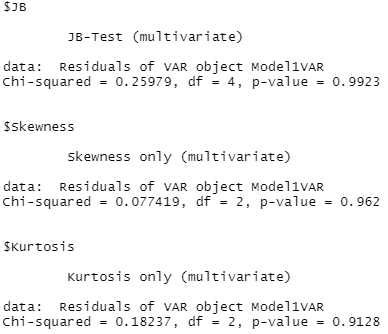
**Homoscedaticitatea eorilor**



*Figura 43. testului ARCH LM*

Valoarea p-value asociată testului ARCH Lagrange-Multiplier este 0,2, ceea ce înseamnă că pentru pragul de semnificație de 10%, erorile sunt homoscedastice.

**Normalitatea erorilor**



*Figura 44. Normalitatea erorilor*

Valoarea p-value este superioară oricărui prag de semnificație, erorile sunt normal distribuite.

### Cauzalitatea Granger

Testul de cauzalitate Granger este folosit pentru a determina dacă o serie de timp este un factor ce poate oferi informații utile în prognoza unei alte serii de timp. Astfel, o variabilă x este considerată cauza unei variabile y, dacă, ţinând cont de valorile trecute ale variabilei x, se poate realiza o predicţie bună a lui y. Testarea cauzalității implică asigurarea inexistenței autocorelării erorilor pentru seriile de date, care trebuie să fie staționare.

Tabel nr.3. Cauzalitatea Granger

|  |  |  |
| --- | --- | --- |
| **Ipoteza nulă** | **p-value** | **Decizie** |
| PIB-ul nu cauzează consumul de energie electrică | 2.248e-06 | Respingem ipoteza nulă, există cauzalitate Granger \*\* |
| Consumul de energie electrică nu cauzează PIB-ul | 0.08718 | Respingem ipoteza nulă, există cauzalitate Granger \* |
| PIB-ul nu cauzează instantaneu consumul de energie electrică | 0.007902 | Respingem ipoteza nulă, există cauzalitate Granger instantanee \*\* |
| Consumul de energie electrică nu cauzează instantaneu PIB-ul | 0.007902 | Respingem ipoteza nulă, există cauzalitate Granger instantanee \*\* |

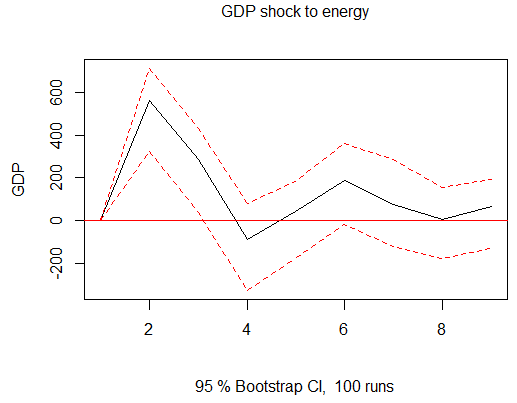
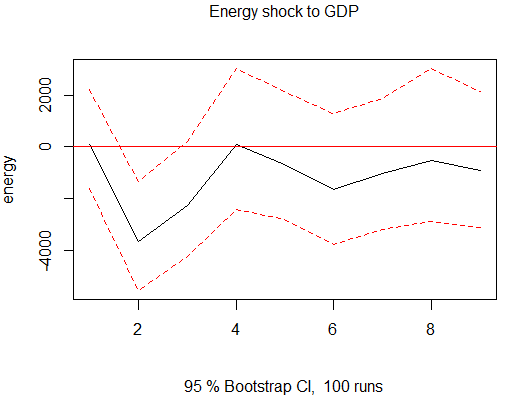
\* - nivel de semnificație de 0,1

\*\* - nivel de semnificație de 0,01

Conform datelor din tabelul nr.3, atât consumul de energie elctrică cauzează PIB-ul în mod Granger, dar și invers. În plus, există o cauzalitatea instantanee între cele două serii de timp.

### Funcția de răspuns la impuls

Funcția de răspuns la impuls este utilizată cu scopul de a descrie evoluția variabilelor unui model ca reacție la un șoc în una sau mai multe variabile. În acest sens este posibilă urmărirea transmiterii unui singur șoc în cadrul unui sistem de ecuații. În cazul modelului implementat, se consideră că, pe un orizont de 8 trimestre (2 ani), vor fi urmărite răspunsurile PIB-ului la șocurile cosnumului de energie electrică și invers.

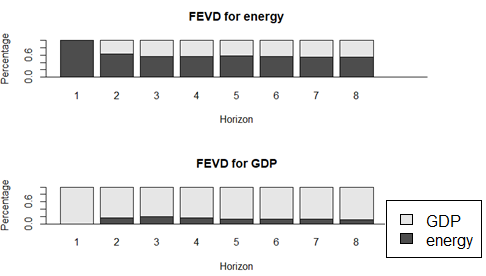
 

*Figura 45. Funcții de răspuns la impuls*

În ambele grafice (figgura 45), vizualizăm șocuri sunt semnificative și repetitive din două în două trimestre. În cazul outputului din stânga, un șoc pozitiv în cadrul consumului de energie electrică produce o creștere asupra Produsului Intern Brut per capita. Pe de altă parte, un șoc negativ în cadrul Produsului Intern Brut per capita produce un răspuns negativ a consumului de energie electrică. De asemenea, nu există saturare a șocului de-a lungul orizontului de 8 trimestre.

### Descompunerea varianței

Analiza descompunerii varianței are rolul de a ilustra ponderea șocului pe care variabila target o primește din partea celeilalte.

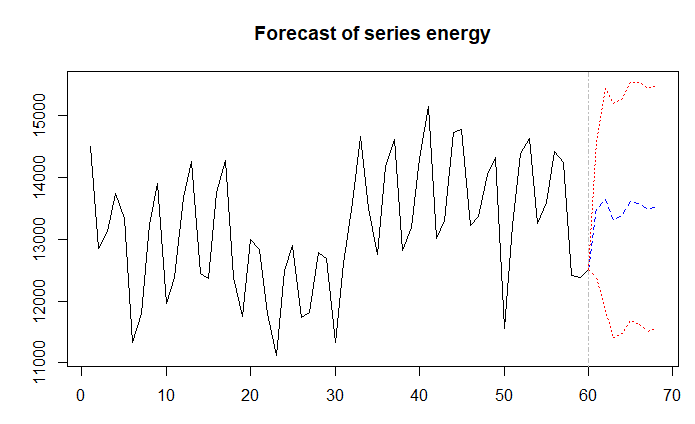


*Figura 46. Analiza descompunerii varianței*

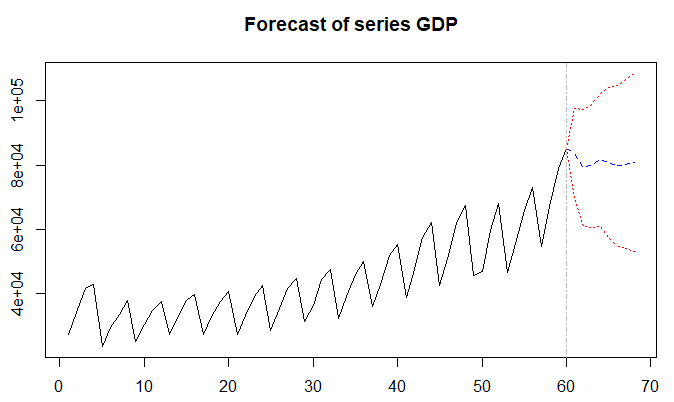
În output-ul din figura 46, observăm că șocul provine din partea variabilei în sine într-o proporție mult mai mare decât din partea indicatorului pereche. Astfel, șocurile de pe piața energetică, se datorează fluctuațiilor consumului de energie electrică. În schimb, ponderea șocului PIB-ului asupra consumului de energie electrică este de 40%, cu excepția primului trimestru. În sens invers, impactul consumului de energie electrică nu este la fel de semnificativ asupra PIB-ului.

### Forecast

Pentru prognoză, s-a ales un orizont de 8 trimestre, pentru a putea observa în grafic dinamica pentru ambele variabile considerate.



*Figura 47. Previziunea seriei consumului de energie electrică (model VAR)*



*Figura 48. Previziunea seriei PIB per capita (model VAR)*

# **Machine Learning. Modelul XGBosst**

XGBoost (Extreme Gradient Boosting) este o bibliotecă de Machine Learning (ML) scalabilă, distribuită cu un arbore decizional (GBDT). Aceasta oferă îmbunătățirea paralelă a arborilor și este biblioteca principală de ML pentru probleme de regresie, clasificare, probleme de ierarhizare sau predicții.

GBDT ( Gradient Boosting Decision Trees) este un algoritm de învățare a ansamblurilor de arbori de decizie similar cu random forecast, utilizat pentru clasificare și regresie. Algoritmii de învățare a ansamblurilor combină mai mulți algoritmi de ML pentru a obține un model mai bun.

Termenul “gradient boosting“ provine de la idea de “boosting“ sau îmbunătățirea unui singur model slab prin combinarea sa cu mai multe modele slabe pentru a genera un model colectiv puternic. Gradient boosting este o extensie a algoritmului în care procedeul de generare aditivă a modelelor slabe este organizat ca un algoritm de coborâre a gradientului asupra unei funcții obiective. Gradient boosting stabilește rezultate țintă pentru următorul model în încercarea de a minimiza erorile. Rezultatele țintă pentru fiecare caz se bazează pe gradientul erorii (de unde și numele gradient boosting) în raport cu predicția.

În literatura de specialitate din ultimii ani au fost demonstrate avantajele Deep Learning și îmbunătățirile pe care le aduc în zona de previziune a seriilor de timp, în detrimentul modelelor de ML. Însă principala condiție care se impune pentru a genera previziuni cât mai corecte o reprezintă existența unui set foarte mare de observații. Așa că, în cazul lucrării de față, principala problemă a fost dată de dimensiunea redusă a seriei de timp observate, motiv pentru care s-a optat și pentru folosirea unui model de ML, care ar putea avea rezultate mai bune în situația dată.

Pe scurt, dorim să obținem o imagine a previziunilor în raport cu valorile corecte, de asemenea, indicatorul care va fi folosit drept măsură a calității previziunii modelului va fi RMSE.

Setul de date de antrenare va cuprinde aproximtiv 80% dintre observațiile inițiale, iar setul de date de testare va cuprinde 20%. Parametrii cei mai performanți au fost:

* learning rate = 0,05
* max\_depth = 2 (controloează adâncimea arborilor)
* nrounds = 86 (numărul de iterații, epoci)

În urma efectuării previziunii pe baza modelului s-au obținut următoarele rezultate:

O imagine care conține text, diagramă, linie, Interval

Descriere generată automat

*Figura 49. Previziunea XGBoost*

Previziunea, deci nu are valori deloc satisfăcătoare, iar RMSE-ul calculat are o valoare de aproximativ 947.678. Această valoare fiind mult mai mare decât cea rezultată din modelele mai sus prezentate.Erorile au o valoare mult prea mare pentru a obține rezultate satisfăcătoare, astfel că vom admite că modelul XGBoost nu produce previziuni favorabile pe seria noastră de date.



*Figura 50. Acuratețea modelului XGBoost*

# **Concluzii**

Nu multe studii empirice au analizat acest fenomen luând în considerare diferite sectoare economice: o mare parte a literaturii de specialitate a investigat eficiența energetică doar la nivel general. Prin analiza realizată în acest proiect dintre energia electrică și Produsul Intern Brut (PIB), am constatat că energia electrică joacă un rol crucial în dezvoltarea economică și progresul unei țări. Energia electrică este esențială pentru susținerea activităților industriale, comerciale și rezidențiale, permițând creșterea economică și îmbunătățirea calității vieții.

Inițial, am dorit să observăm comportamentul consumului de energie electrică prin metode de netezire exponențială, modele ARIMA și tehnici de Machine Learning. Dintre predicțiile obținute pe baza modelelor de tip ARIMA, cu ajutorul criteriilor informaționale, am constatat faptul că cel mai optim a fost SARIMA (1,0,0) (1,0,0). Pentru a surprinde volatilitatea și șocurilor existente pe piață energetică, pe randamentele lunare ale seriei staționare, am aplicat următoarele modele de tip ARCH-GARCH: ARCH(1,0), GARCH(1,1), EGARCH(1,1) și AP-ARCH(1,1). Modelul EGARCH(1,1) a reușit să surpindă cel mai bine șocurile și a avut perfomanța cea mai bună.

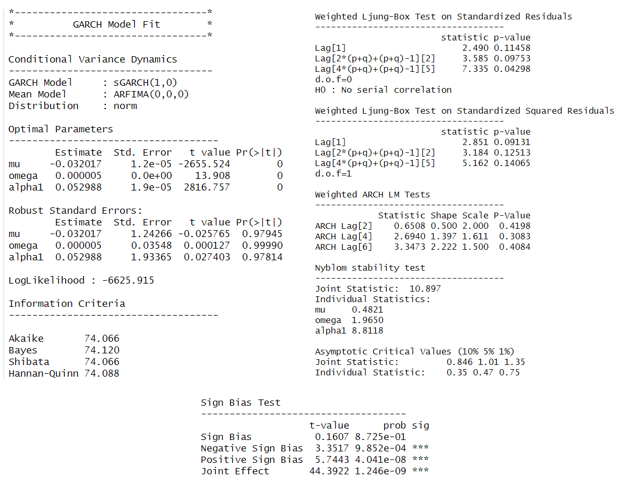
Pentru analiza comportamentelor seriilor din perspectivă multivariată am realizat un model VAR, unde sunt valide toate ipotezele reziduurilor, cu excepția autocorelării erorilor la un prag de semnificație de 10%. Modelul a reușit să surprindă fluctuațiile consumului de energie electrică, dar nu a fost anticipată apariția pandemiei. Cele două variabile – consumul de energie electrică și Produsul Intern Brut per capita – se cauzează Granger reciproc. În schimb, ponderea șocului PIB-ului asupra consumului de energie electrică este de 40%, cu excepția primului trimestru. În sens invers, impactul consumului de energie electrică nu este la fel de semnificativ asupra PIB-ului.

În ultima parte a proiectului, a fost realizat modelul din sfera ML – XGBoost. Pentru XGBoost, previziunea, deși nu are valori deloc satisfăcătoare, RMSE-ul calculat aici are o valoare de aproximativ 947,678. Astfel, nu putem considera modelul XGBoost potrivit pentru analiza noastră.

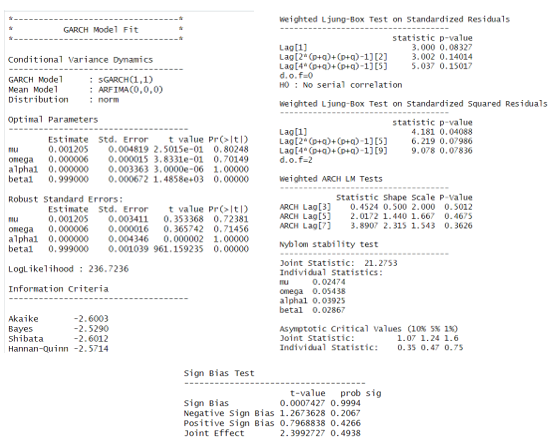
# **Bibliografie**

* Ozturk, S. and Ozturk, F. (2018). Forecasting Energy Consumption of Turkey by Arima Model. Journal of Asian Scientific Research. doi: <https://doi.org/10.18488/journal.2.2018.82.52.60>
* Costantini, V. and Martini, C. (2010). The causality between energy consumption and economic growth: A multi-sectoral analysis using non-stationary cointegrated panel data. Energy Economics. doi: <https://doi.org/10.1016/j.eneco.2009.09.013>
* Hwang, D.B.K. and Gum, B. (1991). The Causal Relationship Between Energy and Gnp: The Case of Taiwan. The Journal of Energy and Development. Available at: <https://www.jstor.org/stable/24807960>
* CEZ Vânzare. (2021). Producția și consumul de energie electrica în România. Available at: <https://www.cezinfo.ro/portal-eficienta/stiri-si-articole/productia-si-consumul-de-energie-electrica-in-romania/>
* What is XGBoost? An Introduction to XGBoost Algorithm in Machine Learning: Simplilearn (2023). Available at: <https://www.simplilearn.com/what-is-xgboost-algorithm-in-machine-learning-article>
* XGBoost Algorithm (2019). Available at: <https://towardsdatascience.com/https-medium-com-vishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d>

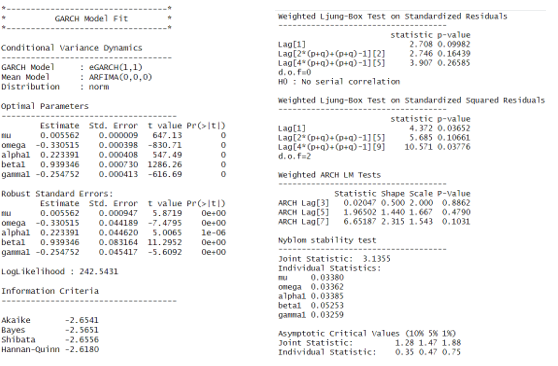
# **Anexă**



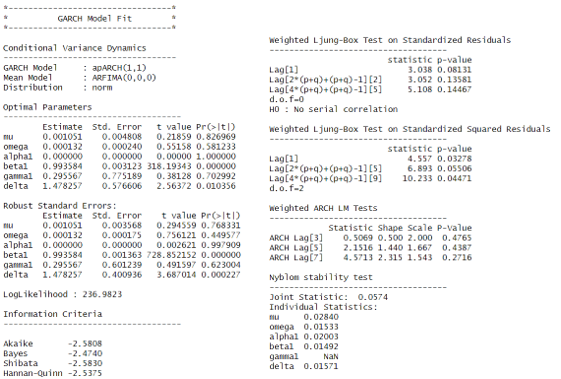
Anexa 1. Output ARCH (1,0)



Anexa 2. Output GARCH (1,1)



Anexa 3. Output E-Garch (1,1)



Anexa 4. Output AP -Arch (1,1)