

Academia Oamenilor de Știință din România

Ilfov 3, 050044 București, Romania



RAPORT FINAL DE CERCETARE

Decembrie 2025

Proiect de cercetare:

” Îmbunătățirea prognozelor ratei inflației în România folosind analiza de sentiment și machine learning (învățare automată)”

Director de proiect: Prof. univ. dr. habil. CS I Mihaela Simionescu

Academia Oamenilor de Știință din România, Ilfov 3, 050044 București, Romania

Universitatea din București

E-mail: mihaela.simionescu@unibuc.ro, mihaela.simionescu@faa.unibuc.ro

Telefon: + 40 748 112 411

Cercetător: dr. CS III Alexandru-Sabin Nicula

Academia Oamenilor de Știință din România, Ilfov 3, 050044 București, Romania

Universitatea Babeș-Bolyai

E-mail: sabin.nicula@ince.ro

Telefon: + 40 749 980 159

Prezentul raport include lista publicațiilor în cadrul proiectului, autorii având afilierea AOSR, dar și un studiu corelat cu tema proiectului.

Diseminarea rezultatelor cercetării

Articole publicate în reviste WoS

1. Simionescu Mihaela, Nicula Alexandru-Sabin (2024). Sentiment Analysis as Innovation in the Inflation Forecasting in Romania. *Marketing and Management of Innovations*, 15(2), 13–25, ISSN (print) – 2218-4511, ISSN (online) – 2227-6718, WOS:001254017400002, IF =1.2, AIS=0.101, <https://doi.org/10.21272/mmi.2024.2-02>.
https://mmi.sumdu.edu.ua/wp-content/uploads/2024/06/02_A809-2024_Simionescu-et-al.pdf
2. Simionescu Mihaela, Alexandru- Sabin Nicula (2025). The role of sentiment analysis and central bank interest rate decisions in forecasting inflation: A Bayesian non-parametric approach for Czech Republic and Romania, *Romanian Statistical Review*, nr. 1 / 2025, 3-24, WOS:001495208700001, ISSN Print: 1018-046X, ISSN Online: 1844-7694, IF=0.2, AIS=0.021.
https://www.revistadestatistica.ro/wp-content/uploads/2025/05/A1_RRS_1_2025.pdf
3. Simionescu Mihaela (2025). Machine Learning vs. Econometric Models to Forecast Inflation Rate in Romania? The Role of Sentiment Analysis, *Mathematics*, 13(1), 168, 1-18, WOS:001393595400001, EISSN 2227-7390, IF=2.2, AIS=0.373, <https://doi.org/10.3390/math13010168>
<https://www.mdpi.com/2227-7390/13/1/168>
4. **Simionescu Mihaela**, Oancea Bogdan (2025). Does Income Inequality Influence Energy Consumption in the European Union?, *Energies*, 18(4), 787, 1-19, WOS:001431612400001, ISSN: 1996-1073, IF=3.2, AIS=0.470, <https://doi.org/10.3390/en18040787>
<https://www.mdpi.com/1996-1073/18/4/787>
5. **Simionescu Mihaela**, Oancea Bogdan (2025). Income/gender inequality and energy use in the European Union, *Energy Efficiency*, 18(32), 1-21, WOS:001458340700001, Electronic ISSN1570-6478, Print ISSN 1570-646X, IF=2, AIS=0.396, <https://doi.org/10.1007/s12053-025-10314-4>

https://link.springer.com/article/10.1007/s12053-025-10314-4?utm_source=rct_congratemail&utm_medium=email&utm_campaign=oa_20250402&utm_content=10.1007/s12053-025-10314-4

6. **Simionescu Mihaela**, Cifuentes-Faura Javier (2025). Empirical Insights on Inflation, Financial Development and Income Inequality in Central and Eastern European Countries, Czech Journal of Economics and Finance, 75 (1), 29-58, WOS:001459891400002, Print ISSN 0015-1920, Online ISSN 2464-7683, în Q4 după factorul de impact, IF=0.7, AIS=0.082, <https://doi.org/10.32065/CJEF.2025.01.02>
https://journal.fsv.cuni.cz/storage/1544_attachment.pdf
7. **Simionescu Mihaela**, Constantin Daniela-Luminița, Marin Erika (2025). The impact of demographic dynamics on income inequality at regional level: the intriguing case of Romania, Management Research and Practice, 17(1), 5-17, WOS:001439963800001, ISSN: 2067-2462, IF=0.6, AIS=0.064.
<https://mrp.ase.ro/v17i1/1.pdf>
8. Simionescu Mihaela (2025). The Impact of Income Inequality on Energy Poverty in the European Union, International Journal of Financial Studies, 13(2), 54, 1-29, WOS:001516019300001, EISSN 2227-7072, IF=2.2, AIS=0.319.
<https://doi.org/10.3390/ijfs13020054>
<https://www.mdpi.com/2227-7072/13/2/54>
9. Simionescu Mihaela (2025). Does finance contribute to income inequality in the European Union?, Borsa Istanbul Review, 1-12, Print ISSN: 2214-8450, Online ISSN: 2214-8469, în Q1 după factorul de impact și Q2 după AIS, IF=7.1, AIS=0.834.
<https://doi.org/10.1016/j.bir.2025.06.010>
<https://www.sciencedirect.com/science/article/pii/S2214845025000997>

Articol publicat în Journal of Knowledge Dynamics (revista AOSR)

Simionescu Mihaela (2025). The Impact of Inflation Rate on Income Inequality in Romania, Journal of Knowledge Dynamics, 2(1), 19-32, ISSN Online 3061-2640.
<https://doi.org/10.56082/jkd.2025.2.19>
<https://www.aos.ro/wp-content/anale/JKDVol2Nr1Art.2.pdf>

ÎNVĂȚARE AUTOMATĂ VS. MODELE ECONOMETRICE PENTRU A PROGNOZA RATA INFLAȚIEI ÎN ROMÂNIA? ROLUL ANALIZEI DE SENTIMENT

1. Introducere

Acuratețea previziunilor inflației este esențială pentru România, deoarece banca națională a țării folosește o strategie de țintire directă a inflației din 2005. Mai mult, această strategie devine mai dificil de aplicat din cauza inflației ridicate din ultimii ani, cauzată de criza energetică globală declanșată de conflictul din Rusia și Ucraina, de perturbările lanțului de aprovizionare determinate de criza sanitară recentă și de război, precum și de creșterea salariilor.

Previziunile precise sunt necesare în vederea ajustărilor necesare pentru a menține economia pe un trend de creștere și a păstra încrederea publicului în eficacitatea băncii. În primul rând, Banca Națională a României (BNR) stabilește ratele dobânzilor pe baza previziunilor inflației. O prognoză precisă permite băncii să ajusteze ratele în mod corespunzător pentru a menține inflația aproape de un nivel țintă. Dacă se preconizează că inflația va fi mai mare decât cea dorită, banca poate crește ratele pentru a încetini creșterea economiei și pentru a tempera inflația. În plus, dacă se prognozează că inflația va fi prea mică, ratele dobânzilor ar putea fi reduse pentru a stimula economia și a aduce inflația la un nivel mai ridicat (Woodford, 2007).

În al doilea rând, antreprenorii și consumatorii se bazează pe previziunile inflației pentru a-și fundamenta deciziile economice. Previziunile precise creează un mediu mai stabil pentru investiții, împrumuturi și stabilirea prețurilor. Companiile pot face planuri adecvate privind costurile și profiturile viitoare, iar gospodăriile pot lua decizii informate cu privire la economisire și cheltuieli în viitorul apropiat.

În al treilea rând, când banca națională își atinge în mod constant ținta de inflație, se consolidează încrederea publicului în capacitatea sa de a gestiona economia (Rossi, 2023). Această încredere este esențială pentru un sistem financiar sănătos.

Obiectivul de a furniza previziuni precise depinde de mulți factori: calitatea și disponibilitatea datelor, metoda de prognoză, includerea factorilor externi, actualizarea și rafinarea modelului, expertiza și aprecierile experților în previziune, comunicare și transparență etc. (Döhrn și Schmidt, 2011). Cu toate acestea, în acest raport se consideră selectarea metodei de prognoză adecvate pentru a obține previziuni cu cel mai mare grad de acuratețe.

Din acest punct de vedere, prognoza empirică recentă a inflației poate fi sistematizată în câteva direcții de cercetare: utilizarea modelelor de echilibru general stochastic dinamic (cunoscute sub denumirea de modele DSGE) datorită dezvoltării multor metode bayesiene care ar putea fi utilizate în estimări, utilizarea de metode alternative la modelele DSGE (mers aleatoriu, modele vectorial-autoregresive de tip Bayesian (VAR), modele mixte bazate pe DSGE și VAR, alte tipuri de modele econometrice, metode de învățare automată), utilizarea predicțiilor furnizate de experți în previziune. În acest cadru, au fost avansate două dezbateri (Del Negro și Schorfheide, 2013). În primul rând, unele studii susțin că modelele DSGE depășesc toate celelalte modele de prognoză a inflației, în timp ce altele arată superioritatea modelelor hibride (Smets și Wouters, 2007). În al doilea rând, capacitatea unei metode de prognoză de a depăși performanțele celorlalte depinde de natura datelor (utilizarea datelor în timp real care oferă rezultate mai bune sau a datelor revizuite)

și de capacitatea metodei de a discerne tendințele pe termen lung prezentate în date (Cardani et al., 2019).

Prognoza empirică recentă a inflației este influențată semnificativ de noi fundamente teoretice pe fondul progresului tehnologic rapid. Apariția resurselor puternice de calcul, în special a unităților de procesare grafică (GPU), a revoluționat învățarea automată utilizată în elaborarea de prognoze. Capacitățile de procesare paralelă ale GPU-urilor au permis antrenarea de modele complexe pe seturi de date masive, reducând semnificativ timpul de calcul și permițând dezvoltarea de algoritmi mai intensivi pentru date (Cano, 2018). Prin urmare, prognoza empirică a inflației bazată pe învățarea automată și alte metode ar trebui subordonată obiectivului de a obține predicții cu precizie maximă. În acest studiu, întrebarea de cercetare este legată de capacitatea învățării automate de a furniza previziuni mai bune ale inflației pentru România în comparație cu modelele econometrice. Mai mult, dezvoltarea resurselor de calcul ne ajută să extragem sentimente din cantități mari de texte și să includem aceste măsuri de sentiment în modelele econometrice pentru a îmbunătăți previziunile. Analiza sentimentului se realizează ușor prin utilizarea procesării limbajului natural (NLP), care combină beneficiile aduse de lingvistică, informatică și inteligență artificială. Scopul său este de a reduce decalajul dintre limbajul uman și înțelegerea mașinii.

În timp ce atât econometria, cât și învățarea automată își propun să construiască modele predictive folosind variabile explicative, abordările acestora diferă semnificativ. Econometria se concentrează pe modele statistic riguroase care explică fenomenele economice, bazându-se adesea pe teorii bine definite. Învățarea automată, pe de altă parte, prioritizează algoritmi puternici care pot învăța direct modele complexe din date, excelând în gestionarea seturilor de date mari și diverse (Charpentier et al., 2018).

Cercetări recente sugerează că modelele de învățare automată pot depăși performanța metodelor econometrice tradiționale în ceea ce privește precizia predicției (Li și Kockelman, 2022). Principala limitare este asociată cu costul generat de interpretabilitatea redusă. Econometricienii pot beneficia din înțelegerea acestor diferențe și din explorarea modului în care instrumentele de învățare automată pot fi integrate în modelele lor pentru a valorifica punctele forte ale ambelor abordări. Acest potențial de colaborare ar putea duce la o analiză economică mai cuprinzătoare și mai puternică.

Având în vedere contextul economiei românești care se confruntă cu o inflație ridicată, scopul principal al acestui studiu este de a testa empiric dacă previziunile ratei inflației bazate pe modele econometrice care încorporează analiza sentimentului depășesc tehnicile de învățare automată. Acest obiectiv este atins prin introducerea fundamentelor teoretice susținute de studii anterioare în secțiunea următoare și prin dezvoltarea cercetării proprii pe baza metodelor cantitative.

2. Date și metodologie

Ratele inflației trimestriale sunt previzionate pe orizontul 2023: T1-2023: T4 folosind modele econometrice și tehnici de învățare automată descrise în această secțiune a lucrării.

Rapoartele privind rata inflației publicate de Banca Națională a României sunt utilizate pentru a calcula indicii de sentiment pe baza rezumatului secțiunii *Perspective asupra inflației*. IntelliDockers a fost utilizat pentru a calcula valorile indicelui de sentiment în fiecare trimestru. Mai mult, rata șomajului trimestrială este, de asemenea, utilizată printre variabilele din metodele de învățare automată.

Rata inflației din România a crescut brusc la sfârșitul anului 2022, atingând un vârf de 16,37%. Acest episod hiperinflaționist se explică prin creșterea globală a prețurilor la energie și produse alimentare din cauza războiului din Ucraina. România, alături de multe alte țări, se bazează pe importuri pentru aceste bunuri esențiale, care au devenit mai scumpe pe plan intern. În plus, perturbările persistente ale lanțului de aprovizionare determinate de epidemia recentă au limitat disponibilitatea diferitelor bunuri în anul 2022. Acești factori coroborați cu oferta limitată care a răspuns cererii globale ridicate au determinat creșterea prețurilor în România. Pe de altă parte, România a înregistrat cea mai mare dezinflație în trimestrul al treilea al anului 2019, probabil din cauza scăderii cheltuielilor excesive ale consumatorilor.

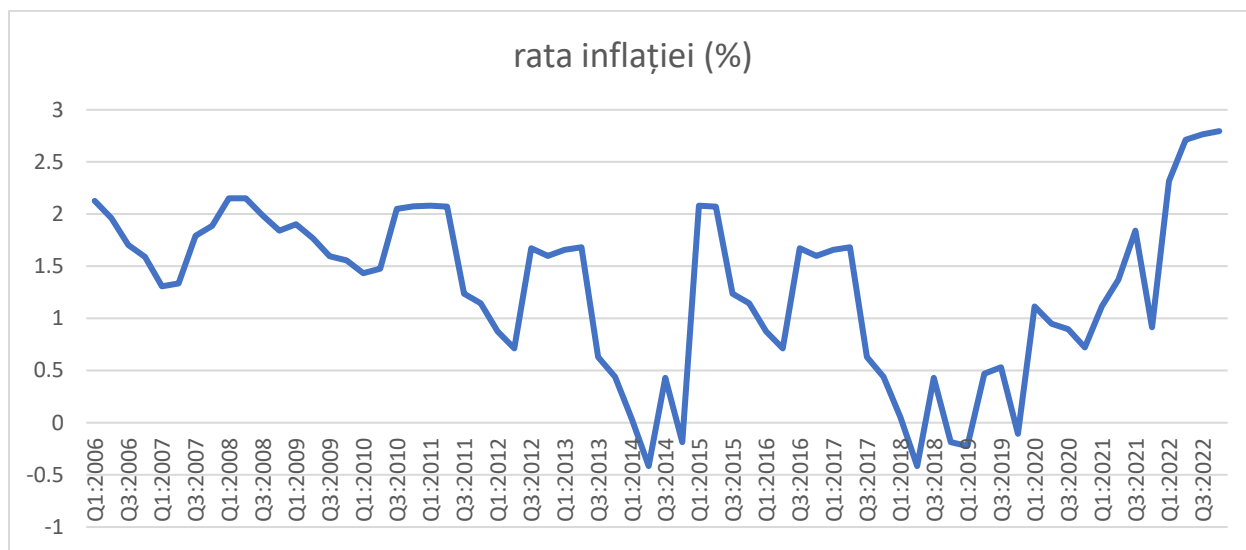


Figura 1. Evoluția ratei inflației trimestriale în România (2023: T1-2023: T4)

Modele econometrice

Modelele autoregresive cu lag distribuit (ARDL) sunt regresii liniare utilizate pentru a reprezenta relațiile dinamice din date. Aceste modele includ atât variabila dependentă, cât și variabilele exogene cu o anumită întârziere printre regresori. Modelele ARDL oferă mai multe avantaje față de tehnicile tradiționale de regresie, ceea ce le face un instrument util pentru cercetarea economică și financiară.

Principalele avantaje ale modelelor ARDL sunt:

- Evidențierea relațiilor pe termen scurt și lung: Modelele ARDL captează eficient relațiile atât pe termen scurt, cât și pe termen lung între variabile, chiar și când variabilele au serii ce prezintă ordine diferite de integrare.
- Gestionarea ordinelor diferite de integrare: Spre deosebire de tehnicile tradiționale de cointegrare care necesită ca ambele serii să fie integrate de același ordin, modelele ARDL pot folosi serii cu ordine mixte de integrare, fără a depăși ordinul 1 ($I(0)$ și $I(1)$).

- Eficiență pentru eșantioane mici: Modelele ARDL oferă estimări de încredere și eficiente chiar și pentru eșantioane de volum redus. Această robustețe este esențială când disponibilitatea datelor este limitată (Kripfganz și Schneider, 2023).

Deși modelele ARDL oferă numeroase avantaje, acestea prezintă și anumite provocări metodologice care trebuie abordate pentru a asigura validitatea rezultatelor.

- Autocorelarea: Prezența variabilei endogene printre regresori poate introduce autocorelare în termenii de eroare. Pentru a rezolva această problemă, pot fi utilizate tehnici de estimare robuste, cum ar fi regresiiile aparent necorelate (SUR).
- Multicoliniaritate: Includerea valorilor întârziate ale variabilei dependente și a variabilelor explicative poate duce la multicoliniaritate. Selecția atentă a variabilelor și metodele adecvate de transformare a datelor pot ajuta la atenuarea acestei probleme (Wang, 2022).

Pentru a utiliza adecvat modelele ARDL, trebuie îndeplinite anumite ipoteze și condiții:

- Ordinul de integrare pentru seria variabilei dependente: se presupune că variabila dependentă are o serie care este integrată de ordinul unu (I(1)).
- Ordinul maxim de integrare: seriile de timp ale variabilelor din model nu pot avea un ordin de integrare mai mare de unu.
- Ipoteza de exogenitate: variabilele explicative sunt considerate slab exogene, ceea ce înseamnă că nu sunt cauzate de variabila endogenă.

Considerând variabila dependentă Y și predictorii X_1, X_2, \dots, X_k , un model ARDL (p, q_1, q_2, \dots, q_k) este reprezentat ca:

$$Y_t = \alpha_0 + \alpha_1 t + \sum_{i=1}^p \beta_i Y_{t-i} + \sum_{j=1}^k \sum_{l_j=0}^{q_j} \gamma_{j,l_j} X_{j,t-l_j} + \varepsilon_t \quad (1)$$

α_0 -constantă;

α_1 - coeficient pentru trend;

β_i -coeficienți pentru variabila dependentă cu valori decalate;

γ_{j,l_j} - coeficienți pentru cele k variabile explicative cu valori decalate $j=1, 2, \dots, k$;

ε_t -erori.

Considerând operatorul de lag L , polinoamele în L notate cu $\beta(L)$ și $\gamma_j(L)$ pot fi scrise astfel:

$$\beta(L) = 1 - \sum_{i=1}^p \beta_i L^i$$

$$\gamma_j(L) = 1 - \sum_{l_j=1}^{q_j} \gamma_{j,l_j} L^{l_j}$$

Considerând polinoamele în L , ecuația din (1) devine:

$$\beta(L)Y_t = \alpha_0 + \alpha_1 t + \sum_{j=1}^k \gamma_j(L)X_{j,t} + \varepsilon_t \quad (2)$$

Există trei reprezentări alternative ale modelului ARDL (p, q_1, q_2, \dots, q_k). Prima reprezentare este utilizată pentru estimarea dinamicii intertemporale, a doua pentru derivarea relației de echilibru pe termen lung post-estimare, iar a treia este o reducere la reprezentarea modelului de corecție a erorilor condiționată de testul limitelor pentru cointegrare, cunoscut sub numele de testul Pesaran, Shin și Smith (2001) sau testul PSS. Toate aceste reprezentări alternative se bazează pe descompunerea Beveridge-Nelson, o procedură pentru descompunerea unei serii de timp nestaționare în componente stochastice (una permanentă care urmează un mers aleator cu drift și una tranzitorie care este staționară în jurul unei medii nule) (Granger, 2004).

Modelele autoregresive de medie mobilă integrată de tip sezonier (SARIMA) extind capabilitățile modelelor ARIMA tradiționale prin integrarea explicită a sezonității în analiză. Sezonalitatea consideră evoluții repetitive în serii care apar la intervale predictibile, notate prin "s" în acest context. Modelele SARIMA captează componenta sezonieră prin trei elemente cheie:

1. Parte sezonieră autoregresivă (ordinul P): Această parte ia în considerare impactul valorilor sezoniere din trecut (decalaje ale multiplilor de "s") asupra valorii curente. Numărul de termeni sezonieri autoregresivi este notat cu "P".
2. Parte sezonieră de medie mobilă (ordinul Q): Această componentă încorporează impactul erorilor sezoniere din trecut (reziduuri) asupra valorii curente. Numărul de termeni sezonieri de medie mobilă este notat cu "Q".
3. Diferențierea sezonieră (D): Similar cu diferențierea în modelele ARIMA, diferențierea sezonieră elimină nestaționaritatea cauzată de modelele sezoniere. Numărul de diferențe sezoniere impuse pentru a obține staționaritatea este notat cu "D".

Pentru a determina prezența sezonității și a selecta valori adecvate pentru P, Q și D, este necesar să se descrie funcția de autocorelație (ACF) și funcția de autocorelație parțială (PACF) a seriei de timp. Vârfuri semnificative în aceste funcții la decalaje care sunt multipli de "s" indică prezența componentelor sezoniere.

Un proces staționar SARIMA $(p,q) \times (P, Q)_s$ cu $D=0$ poate fi reprezentat astfel:

$$\phi(L)\varphi(L^s)(y_t - \mu) = \theta(L)\Theta(L^s)\varepsilon_t.$$

Sunt incluse componente nesezoniere de tip ARMA (p, q) și componente sezoniere $(P, Q)_s$ și numărul de diferențieri sezoniere (D).

Componentele sezoniere sunt reprezentate astfel: SAR(P): $\varphi(L^s) = 1 - \varphi_1 L^s - \varphi_2 L^{2s} - \dots - \varphi_p L^{ps}$ și SMA(Q): $\Theta(L^s) = 1 - \Theta_1 L^s - \Theta_2 L^{2s} - \dots - \Theta_q L^{qs}$

Componentele nesezoniere sunt reprezentate prin: AR(p): $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$ și MA(q): $\phi(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$

Componentele nesezoniere controlează dinamica pe termen scurt, în timp ce $p < 0.5s$ și $q < 0.5s$.

În caz de periodicitate diferită în componenta sezonieră de tip autoregresiv de medie mobilă, $(p,q) \times (P, Q)_{s,s'}$ se calculează astfel:

$$\phi(L)\varphi(L^s)(y_t - \mu) = \theta(L)\Theta(L^{s'})\varepsilon_t$$

S reprezintă componenta sezonieră de tip autoregresiv, s' reprezintă componenta sezonieră de tip medie mobilă.

Componentele sezoniere sunt: SAR(P): $\varphi(L^s) = 1 - \varphi_1 L^s - \varphi_2 L^{2s} - \dots - \varphi_p L^{ps}$ și SMA(Q): $\Theta(L^{s'}) = 1 - \Theta_1 L^{s'} - \Theta_2 L^{2s'} - \dots - \Theta_q L^{qs'}$. Componentele nesezoniere sunt: AR(p): $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$ și MA(q): $\phi(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$.

Metode de învățare automată

Random forests („pădurile aleatoare”) (RF) reprezintă un algoritm utilizat pentru clasificare și regresie. Acestea se bazează pe crearea unei colecții de arbori de decizie și apoi combinarea predicțiilor fiecărui arbore pentru a obține o predicție finală, cu precizie mai mare. Arborii de regresie utilizați în RF sunt modele neparametrice cu partiționare recursivă a spațiului covariat. Date fiind K noduri terminale, minimizarea sumei pătratelor erorilor se face pentru a selecta diviziunile:

$$\pi_{t+h} = \sum_{k=1}^K c_k I_k(x_t; \theta_k) \quad (1)$$

θ_k - parametrii

Dacă $R_k(\theta_k)$ este regiunea k, atunci:

$$I_k(x_t; \theta_k) = \begin{cases} 1, & \text{pentru } x_t \in R_k(\theta_k) \\ 0, & \text{altfel} \end{cases}$$

Bootstraparea (bootstrapping) se folosește pentru agregarea arborilor de regresie și se consideră un număr de 1 000 eșantioane de bootstrap. În cazul unui astfel de eșantion, un arbore poate include K_b regiuni. Previziunile se calculează astfel:

$$\hat{\pi}_{t+h} = \frac{1}{B} \sum_{b=1}^B \left[\sum_{k=1}^{K_b} \hat{c}_{k,b} I_{k,b}(x_t; \hat{\theta}_{k,b}) \right] \quad (2)$$

Mașinile vectoriale de suport (SVM) sunt folosite și pentru predicția valorilor viitoare pe baza datelor istorice datorită câtorva avantaje. În primul rând, datele din seriile de timp din lumea reală prezintă adesea conexiuni neliniare. SVM pot capta eficient aceste neliniarități prin utilizarea funcțiilor nucleu (kernel), care proiectează datele într-un spațiu de dimensiune mai mare unde este reprezentată o relație liniară. În al doilea rând, datele din serii de timp pot avea multe caracteristici,

care includ valorile trecute și alți factori relevanți. SVM gestionează adecvat aceste date cu dimensionalitate ridicată, reducând nevoia de selectare a caracteristicilor (ceea ce uneori poate elimina informații valoroase).

Regresia vectorială de suport (SVR) reprezintă utilizarea SVM pentru regresie și aproximarea funcțiilor. Setul de date de antrenament pentru caracteristici și ieșiri este reprezentat ca : $Z = \{X_j, Y_j | j = 1, 2, \dots, n\}$, $X_j \in \mathfrak{R}^q$ reprezintă valorile caracteristicilor pentru exemplul de antrenare j , $Y_j \in \mathfrak{R}$ arată valoarea țintă pentru exemplul de antrenare j , q reprezintă numărul de caracteristici din vectorul de intrări, n reprezintă dimensiunea setului de date de antrenare. Forma acestui model de regresie este redată mai jos:

$$Y = \omega^T \cdot \theta(X) + b \quad (3)$$

b - deplasare (bias); $\theta(X)$ - funcția de mapare neliniară; ω - vector de ponderi calculat prin minimizarea funcției regularizate din ecuația (4) folosind restricția din ecuațiile (5), (6) și (7):

$$\min \{0.5 \cdot \omega^2 + C \sum_{j=1}^N (\varepsilon_j + \varepsilon_j^{(*)})\} \quad (4)$$

$$Y_j - \left\{ \left(\omega^T \cdot \theta(X_j) \right) + b \right\} \leq \psi + \varepsilon_j, j=1, 2, \dots, N \quad (5)$$

$$\varepsilon_j, \varepsilon_j^{(*)} \geq 0, j=1, 2, \dots, N \quad (6)$$

$$f(X) = \sum_{j=1}^N (\delta_j - \delta_j^*) K(X, X_j) + b \quad (7)$$

C - coeficientul de penalizare corespunzător erorilor; $\varepsilon_j, \varepsilon_j^{(*)}$ - variabile excedent pozitive; ψ - funcția de aproximare a acurateții; δ_j, δ_j^* - multiplicatori Lagrange.

Rețelele neuronale artificiale (ANN) sunt, de asemenea, instrumente puternice pentru prognoza seriilor de timp. Acestea sunt capabile să reflecte relații complexe neliniare, permițând scenariu în care metodele statistice tradiționale ar putea întâmpina dificultăți. În primul rând, spre deosebire de unele metode statistice, ANN nu necesită presupuneri despre relațiile liniare din date. În al doilea rând, arhitectura rețelei (numărul de straturi, neuroni) poate fi adaptată la specificul seriei de timp. În al treilea rând, ANN pot identifica modele complexe și dependențe din date care nu pot fi demonstrate de modele simple.

ANN conține un eșantion interconectat de neuroni care sunt legați între ei. Pentru al j -lea neuron, se consideră vectorul de intrări n -dimensional $x = (x_1, \dots, x_n)$ care corespunde vectorului de ponderi $w = (w_{j1}, \dots, w_{jn})$ și deplasarea b_j . Neuronul de ieșire (y_j) este calculat pe baza funcției de activare f ca: $y_j = f(\sum_{t=1}^n w_{jt} x_t + b_j)$.

Prognozele inflației bazate pe SVM și ANN folosesc ca intrări inflația din perioada anterioară și rata șomajului din perioada actuală. Versiunile care includ indicele de sentiment nu au furnizat rezultate acceptabile din punctul de vedere al preciziei prognozei.

Acuratețea previziunilor

Precizia (acuratețea) prognozei este evaluată pentru a determina metoda care a generat cele mai bune predicții. Se utilizează măsurile uzuale de acuratețe: eroare medie absolută (mean absolute error- MAE), eroare medie (mean error- ME), rădăcina erorii medii pătratice (root mean square error- RMSE).

$$ME = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)$$
$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$
$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$
$$RMSE = \sqrt{MSE}$$

Cu toate acestea, aceste măsuri ale acurateții previziunilor nu au capacitatea de a detecta direcția corectă de schimbare a inflației predictată. Prin urmare, Δ_{t+h}^f și Δ_{t+h} definite mai jos sunt calculate pentru a identifica direcția prevăzută și reală a modificării inflației în trimestrul următor.

$$\Delta_{t+j}^f = \begin{cases} 1 & \text{dacă } inflație_{t+1} > inflație_t \\ -1 & \text{altfel} \end{cases}$$
$$\Delta_{t+j} = \begin{cases} 1 & \text{dacă } inflație_{t+1} > inflație_t \\ -1 & \text{altfel} \end{cases}$$

Modificarea direcțională corectă este dată de procentul de modificări anticipate corecte. Mai mult, testul Diebold-Mariano (DM) este utilizat pentru a compara prognozele din punctul de vedere al preciziei. În acest caz, comparația se face cu o prognoză naivă care consideră valoarea din perioada anterioară ca rămânând neschimbată în perioada curentă.

3. Rezultate

S-au aplicat mai multe teste pentru verificarea staționarității: Augmented Dickey-Fuller (ADF), Phillips-Perron (PP), KPSS test, Elliott, Rothenberg and Stock (ERS). Rezultatele din Tabelul 1 sugerează că toate seriile sunt staționare în primă diferență, ceea ce arată că seriile în nivel sunt integrate de ordinul întâi (I(1)).

Tabelul 1. Rezultatele testelor pentru verificarea staționarității seriilor de timp

Variabila	Testul	Seriile de timp în dublă diferență	Seriile de timp în primă diferență	Seriile de timp în nivel
Rata inflației	ADF	-7,2999 (<0,01)	-7,7822 (<0,01)	-0,5431 (0,4781)
	PP	-27,8409 (<0,01)	-7,7815 (<0,01)	-0,6161 (0,4468)
	KPSS	0,0251	0,2705	0,1955
	ERS	272,9122	0,9308	5,3131
Indicele de sentiment	ADF	-6,3919 (<0,01)	-13,5908 (<0,01)	-0,4816 (0,5035)
	PP	-62,9407 (<0,01)	-43,5127 (<0,01)	-0,4455 (0,5181)
	KPSS	0,2311	0,2292	0,1288
	ERS	5808,833	1,6429	2,3082
Rata șomajului	ADF	-8,2868 (<0,01)	-6,9159 (<0,01)	-0,4164 (0,5296)
	PP	-31,1833 (<0,01)	-6,9091 (<0,01)	-0,4497 (0,5164)
	KPSS	0,1929	0,1494	0,2338
	ERS	8,1724	0,0283	1,9278

Sursa: calcule proprii. Notă: statistici raportate și valori p între paranteze. Pentru testul KPSS, valoarea critică la un nivel de semnificație de 5% este de 0,463. Valorile critice pentru testul ERS la un nivel de semnificație de 5%: 3,02 pentru seriile în nivel, 3,01 pentru seriile în prima diferență și 3,017 pentru seriile în a doua diferență.

Cointegrarea nu se verifică între aceste serii și se construiește un model ARDL. Cel mai bun model selectat este ARDL(1,0,1), care este prezentat în Tabelul 2. Toți parametrii modelului sunt semnificativi statistic la nivelul de 10%. Modelul este corect specificat conform testului

RESET, în timp ce erorile sunt independente, homoscedastice și distribuite normal la un nivel de semnificație de 1%.

Tabelul 2. Estimările pentru modelul ARDL(1,0,1)

Variabila	Coeficientul	Valoarea p	VIF
Intercepție	5,9034	0,0106	-
Inflație(t-1)	0,8131	<0,01	2,033
Indice de sentiment(t)	-5,2753	0,0316	1,976
Indice de sentiment (t-1)	-2,8561	0,0952	1,885
R-pătrat ajustat: 0,6764	RESET stat.= 2,1844, p-value = 0,1213		-
Testul Breusch-Pagan : BP = 1,5614, p-value = 0,6682	Testul Ljung-Box : hi- pătrat calculat = 0,082879, p-value = 0,7734	Testul Shapiro-Wilk W = 0,95068, p-value = 0,163	-

Sursa: calcule proprii în R

Cel mai bun model SARIMA selectat este SARIMA (1,0,0) (2,0,0)₄ care este prezentat în Tabelul 3. Toți coeficienții sunt semnificativi statistic la nivelul de semnificație de 5%. Erorile sunt independente, homoscedastice și distribuite normal la un nivel de semnificație de 1%.

Tabelul 3. Rezultatele estimării pentru modelul SARMA (1,0,0) (2,0,0)₄ cu media diferită de zero

Variabila	Coeficientul	Valoarea p
Media	6,3129	0,0139
AR(1)	0,9044	<0,01
SAR(1)	-0,2246	0,046
SAR(2)	0,3099	0,032
σ^2	3,822	-
R-pătrat: 0.776	Stat, Durbin- Watson: 1,9556	Stat, Jarque- Bera (valoarea p în paranteze): 1,289 (0,554)

Sursa: calcule proprii în EViews 9.

Conform Tabelului 4, modelul ARDL a performat mai bine în ceea ce privește acuratețea prognozelor pe orizontul 2023: T1-2023: T4. Toate predicțiile bazate pe modelul ARDL au anticipat corect direcția de modificare a ratei inflației. Tehnicile de învățare automată au funcționat mai bine decât în cazul modelului SARIMA conform ME, MAE și RMSE. Mai mult, toate prognozele au o acuratețe mai bună decât prognozele naive, așa cum a indicat testul DM.

Tabelul 4. Evaluarea acurateții prognozelor pentru previziunile ratei inflației cu un pas înainte (orizont: 2023: T1-2023: T4)

Metoda	ME	MAE	RMSE	Stat. DM stat. (comparația cu previziunile naive)	Procentul trimestrelor cu direcția de modificare a prognozelor corect anticipată
ARDL(1,0,1)	-1,5325	1,5325	1,9238	-5,88 (<0,01)	100%
SARMA (1,0,0) (2,0,0) ₄	2,025	2,025	2,4341	-4,67 (<0,01)	75%
RF	1,987	1,987	2,237	4,07 (<0,01)	75%
SVM	-1,788	1,788	2,048	4,65 (<0,01)	75%
ANN	2,004	2,004	2,312	3,89 (<0,01)	50%

Sursa: calcule proprii.

În concluzie, modelul ARDL care a inclus indicele sentimentului printre variabilele exogene a determinat cele mai bune prognoze trimestriale pentru rata inflației pe orizontul 2023: T1-2023: T4. Acest tip de model a depășit tehnicile de învățare automată ca acuratețe a previziunilor.

4. Concluzii

Progresul tehnologic ne permite să folosim modele statistice complexe pentru a identifica tipare și conexiuni între serii economice și pentru a previziona inflația. Mai mult, tehnicile de învățare automată (ML) sunt din ce în ce mai mult folosite pentru a construi modele predictive care pot evidenția conexiuni neliniare și complexe și care se pot adapta la condițiile economice dinamice.

Atât modelele econometrice, cât și învățarea automată au puncte forte și puncte slabe în analiza economică. Alegerea abordării depinde de obiectivul cercetării. Econometria este mai potrivită pentru înțelegerea relațiilor cauzale și interpretarea fenomenelor economice când există

o bază teoretică solidă. Învățarea automată oferă capacități puternice de predicție și poate gestiona seturi de date mari și complexe, însă interpretarea acestora poate fi o provocare. O tendință în creștere este combinarea acestor abordări pentru a valorifica avantajele ambelor metode în vederea unei analize mai cuprinzătoare (Shobana și Umamaheswari, 2021).

În timp ce studiul lui Pérez-Pons et al. (2021) a arătat că în majoritatea studiilor comparative învățarea automată are performanțe mai bune decât modelele econometrice, acest studiu arată că încorporarea indicelui sentimentului în modelele econometrice a depășit performanțele prognozelor bazate pe învățarea automată.

Această lucrare prezintă câteva limitări. De exemplu, au fost folosite doar câteva tipuri de modele econometrice și metode de învățare automată. Previziunile sunt elaborate pe un termen scurt și au fost calculate doar câteva măsuri de acuratețe a prognozelor. În studii viitoare, vor fi abordate mai multe metode de prognoză și indicatori de măsurare a acurateții previziunilor. Mai mult, combinarea acestor prognoze, fiecare cu punctele sale forte și slabe, poate fi luată în considerare într-un studiu următor pentru a verifica dacă acuratețea prognozelor este îmbunătățită.

Bibliografie

Cano, A. (2018). A survey on graphic processing unit computing for large-scale data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(1), e1232.

Cardani, R., Paccagnini, A., & Villa, S. (2019). Forecasting with instabilities: An application to DSGE models with financial frictions. *Journal of Macroeconomics*, 61, 103133.

Charpentier, A., Flachaire, E., & Ly, A. (2018). Econometrics and machine learning. *Economie et Statistique*, 505(1), 147-169.

Del Negro, M., & Schorfheide, F. (2013). DSGE model-based forecasting. In *Handbook of economic forecasting* (Vol. 2, pp. 57-140). Elsevier.

Döhrn, R., & Schmidt, C. M. (2011). Information or institution? On the determinants of forecast accuracy. *Jahrbücher für Nationalökonomie und Statistik*, 231(1), 9-27.

Granger, C. W. J. (2004). Time series analysis, cointegration, and applications. *American Economic Review*, 94(3), 421-425.

Kripfganz, S., & Schneider, D. C. (2023). ardl: Estimating autoregressive distributed lag and equilibrium correction models. *The Stata Journal*, 23(4), 983-1019.

Li, W., & Kockelman, K. M. (2022). How does machine learning compare to conventional econometrics for transport data sets? A test of ML versus MLE. *Growth and Change*, 53(1), 342-376.

Pérez-Pons, M. E., Parra-Dominguez, J., Omatu, S., Herrera-Viedma, E., & Corchado, J. M. (2021). Machine learning and traditional econometric models: a systematic mapping study. *Journal of Artificial Intelligence and Soft Computing Research*, 12(2), 79-100.

Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of applied econometrics*, 16(3), 289-326.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Taieb, S. B., ... & Ziel, F. (2022). Forecasting: theory and practice. *International Journal of Forecasting*, 38(3), 705-871.

Rossi, S. (2023). The distributional impacts of inflation-targeting strategies. In *Central Banking, Monetary Policy and Income Distribution* (pp. 261-273). Edward Elgar Publishing.

Shobana, G., & Umamaheswari, K. (2021, January). Forecasting by machine learning techniques and econometrics: A review. In 2021 6th international conference on inventive computation technologies (ICICT) (pp. 1010-1016). IEEE.

Smets, F., & Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American economic review*, 97(3), 586-606.

Wang, L. (2022). Research on the dynamic relationship between China's renewable energy consumption and carbon emissions based on ARDL model. *Resources Policy*, 77, 102764.

Woodford, M. (2007). The case for forecast targeting as a monetary policy strategy. *Journal of economic perspectives*, 21(4), 3-24.