



Analiza și Prognoza Seriilor de Timp

Capitolul 10: Recapitulare Comprehensivă



Daniel Traian PELE

Academia de Studii Economice din Bucureşti

IDA Institute Digital Assets

Blockchain Research Center

AI4EFIN Artificial Intelligence for Energy Finance

Academia Română, Institutul de Prognoză Economică

MSCA Digital Finance

Obiective de învățare

La finalul acestui capitol, veți fi capabili să:

1. Aplicați fluxul complet de prognoză, de la date la evaluare
2. Selectați modelul potrivit în funcție de caracteristicile datelor
3. Evaluați acuratețea prognozelor folosind metri și validare încrucișată
4. Integrați cunoștințele din toate capitolele anterioare în practică



Cuprins

Fundamente

- Metodologia Prognozei
- Studiu de Caz 1: Volatilitatea Bitcoin (GARCH)
- Studiu de Caz 2: Ciclurile Petelor Solare (Fourier)

Aplicații

- Studiu de Caz 3: Șomajul (Prophet)
- Studiu de Caz 4: Analiză Multivariată (VAR)
- Sinteză și Ghid
- Quiz



Abordarea științifică a prognozei

Întrebarea de Cercetare

- Cum putem **evalua riguros** performanța prognozei evitând supraajustarea?

Problema Fundamentală

- Ajustarea în eșantion \neq Performanța în afara eșantionului
- Modelele pot “memora” datele de antrenament fără a învăța tipare
- Soluție:** Metodologia corectă train/validation/test

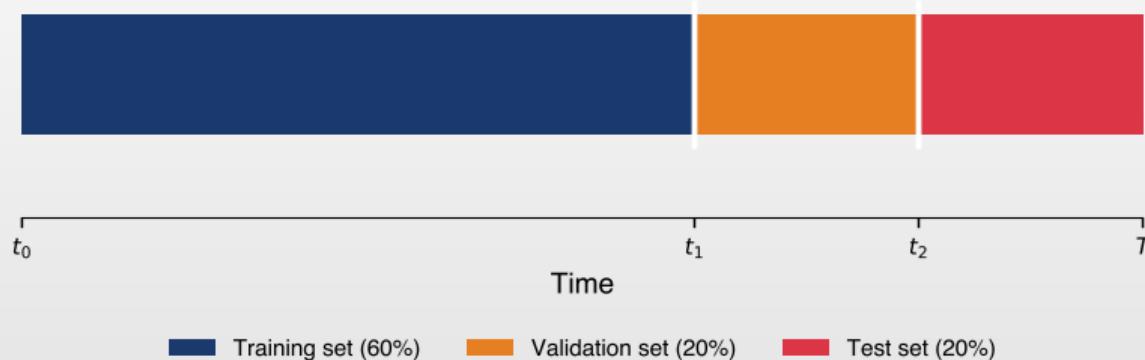
Principiu Cheie

- “Setul de test trebuie să rămână **neatins** până la evaluarea finală.”
- Practică standard în machine learning și econometrie



Cadrul Train/Validation/Test

Train / Validation / Test Split



Set Antrenament

- Estimare parametri
- Cea mai mare parte

Set Validare

- Comparare modele
- Ajustare hiperparam

Set Test

- Păstrat
- Metricti finale

Q TSA_ch10_train_val_test_split

Metrici de evaluare

Definiție 1 (Metrici ale Erorii de Prognoză)

- **Date:** Fie y_t valorile reale, \hat{y}_t progozele

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_t (y_t - \hat{y}_t)^2}, \quad \text{MAE} = \frac{1}{n} \sum_t |y_t - \hat{y}_t|, \quad \text{MAPE} = \frac{100\%}{n} \sum_t \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

Când să folosim

- **RMSE:** Penalizează erorile mari
- **MAE:** Robust la outlieri
- **MAPE:** Independent de scală (%)

Atenție

- MAPE nedefinit când $y_t = 0$
- Comparați pe **același** set test
- Raportați metrici **out-of-sample**



Bitcoin: definirea problemei

Întrebarea de Cercetare

- Putem prognoza **volatilitatea** Bitcoin folosind modele GARCH?

Caracteristicile Datelor

- Sursă: Yahoo Finance (BTC-USD)
- Perioadă: Ian 2019 – Ian 2025
- Frecvență: Zilnică
- Observații: ≈ 2.200 zile

Fapte stilizate

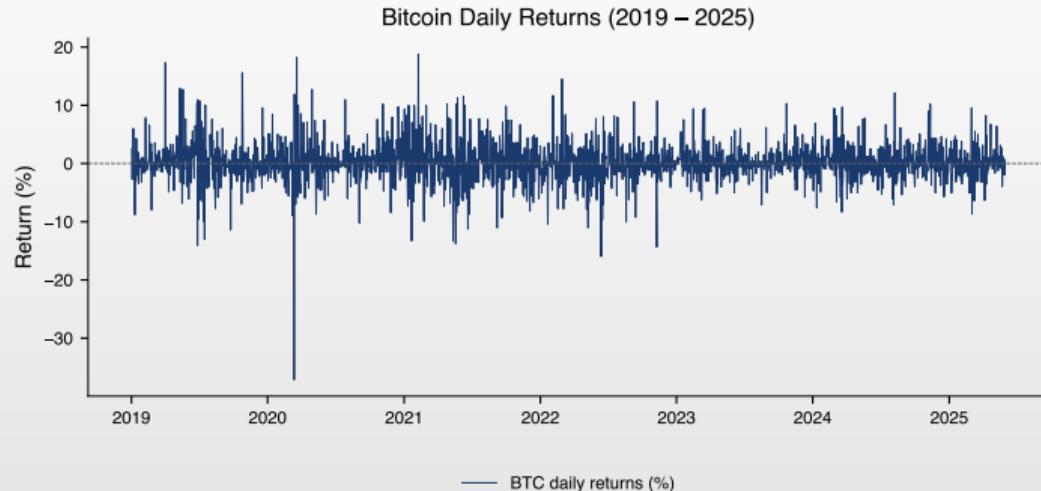
- Randamente: medie aproape zero
- Cozi groase (curtosis > 3)
- Clustering al volatilității

Insight Cheie

- Randamentele financiare sunt de obicei:
 - ▶ Impredictibile în medie
 - ▶ Predictibile în varianță
- ➔ Focus pe **prognoză volatilității**



Bitcoin: volatility clustering

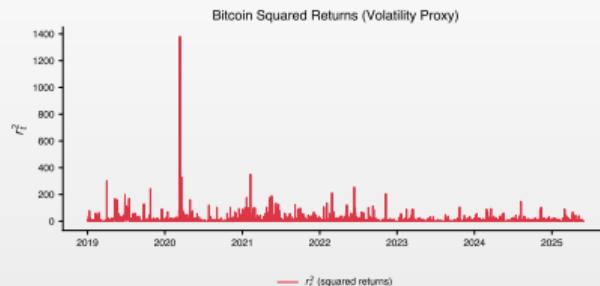


Observație

- Randamentele mari tind să urmeze randamente mari, cele mici urmează cele mici
- Acesta este **volatility clustering** ↘ fenomenul pe care GARCH îl captează

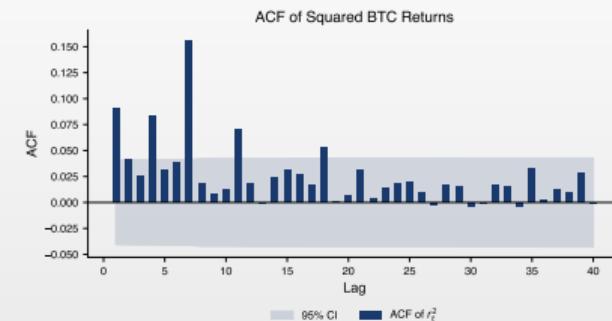


Bitcoin: dovezi pentru GARCH



Randamente pătrate

- r_t^2 sunt proxy pentru volatilitate σ_t^2
- Vârfurile se grupează



ACF

- Barele depășesc benzile albastre \succ autocorelație semnificativă

De ce GARCH?

- Dacă r_t^2 ar fi zgomot alb, ACF ar fi zero
- ACF semnificativ: **volatilitatea trecută prezice volatilitatea viitoare** \succ GARCH captează asta!



Specificarea modelului GARCH

Definiție 2 (Modelul GARCH(p,q))

- **Date:** Fie r_t randamentele. Modelul GARCH(p,q) este:

$$r_t = \mu + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t, \quad z_t \sim N(0, 1)$$

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2$$

- **Condiții:** $\omega > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, și $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1$

Variante de Model

- **GARCH(1,1):** Cel mai comun
- **GJR-GARCH:** Efect de levier
- **EGARCH:** Șouri asimetrice

Interpretare

- α : Impactul șourilor trecute
- β : Persistența volatilității
- $\alpha + \beta \approx 1$: Persistență înaltă



Bitcoin: împărțirea datelor și staționaritate

Împărțirea Datelor

Set	Perioadă	N
Antrenament (70%)	2019-01 – 2023-03	1.543
Validare (20%)	2023-03 – 2024-06	441
Test (10%)	2024-06 – 2025-01	221
Total	2.205	

Teste de Staționaritate

Serie	ADF	Rezultat
Prețuri	$p = 0.50$	Non-staționară
Randamente	$p < 0.01$	Staționară

➤ Modelăm **randamente**, nu prețuri

De ce Contează Staționaritatea

- GARCH**: necesită input slab staționar
- Prețuri vs Randamente**: Prețurile urmează random walk, randamentele sunt staționare



Bitcoin: selectarea modelului pe setul de validare

Metodologie

- Estimăm fiecare model pe datele de antrenament, evaluăm pe setul de validare

Model	AIC	BIC	Val MAE	Selectare
GARCH(1,1)	6.994,8	7.020,6	2,638	Cel mai bun
GARCH(2,1)	6.993,7	7.024,6	2,640	
GJR-GARCH(1,1)	6.983,7	7.014,6	2,669	
EGARCH(1,1)	—	—	—	Eșuat*

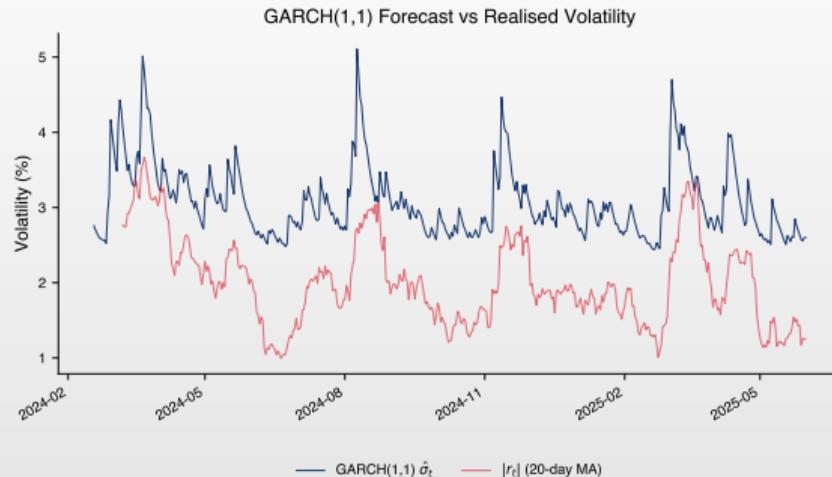
*Prognoze analitice indisponibile pentru $h > 1$

Rezultat

- GARCH(1,1)** selectat pe baza celui mai mic MAE de validare pentru prognozele de volatilitate



Bitcoin: evaluarea finală pe setul de test



Parametri

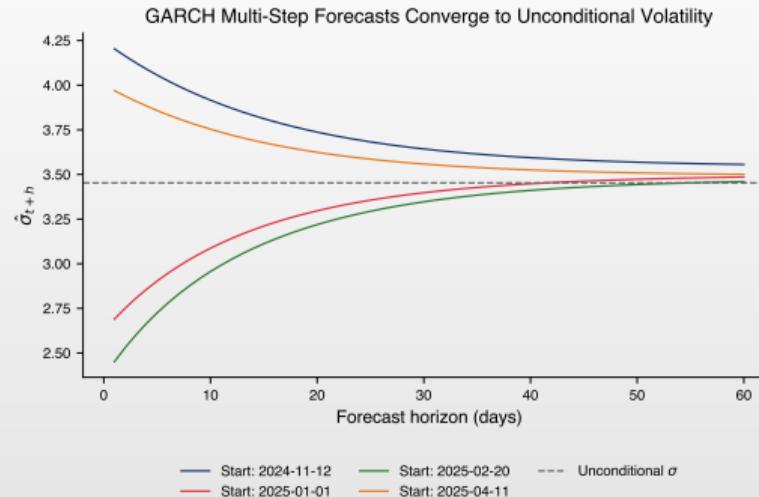
- $\omega = 0,87, \alpha = 0,09, \beta = 0,84$
- $\alpha + \beta = 0,93$ (persistență înaltă)

Performanță Test

- MAE = 1,82, RMSE = 2,14
- Prognoză urmărește bine volatilitatea realizată



GARCH: prognozele multi-step converg

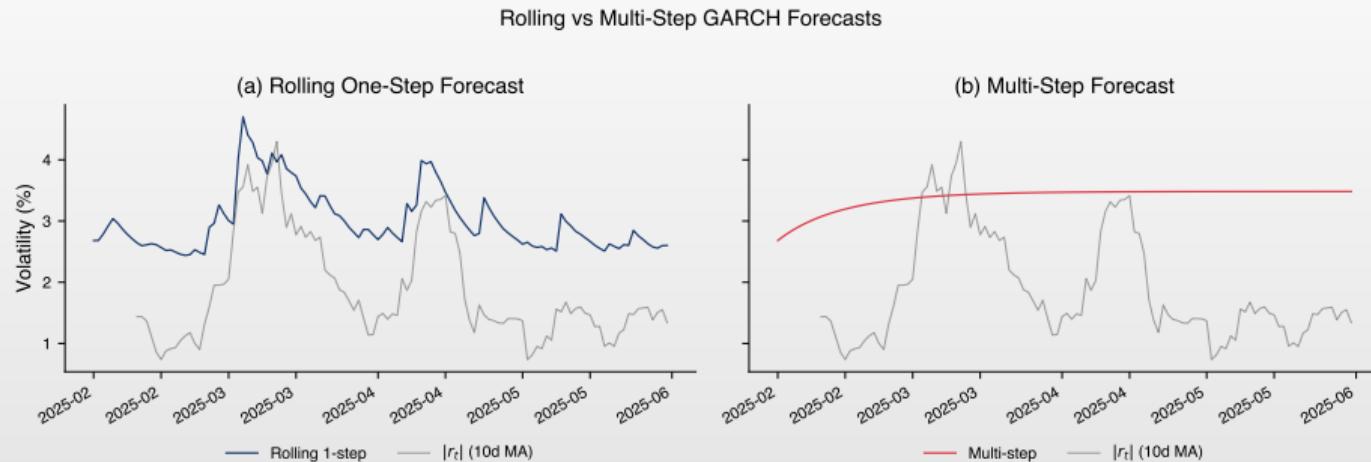


Insight Cheie

- Prognozele multi-step converg la $\bar{\sigma}^2 = \frac{\omega}{1-\alpha-\beta}$
- Soluția: prognoze rolling one-step-ahead



GARCH: soluția rolling one-step-ahead



Rolling 1-Step (Stânga)

- Re-estimare la fiecare t (dinamic)

Multi-Step (Dreapta)

- Converge la $\bar{\sigma}^2$ (plat)



Bitcoin: Concluzii cheie

Sumar

1. Rendamentele sunt staționare; prețurile nu
2. GARCH(1,1) depășește variantele mai complexe
3. Persistență înaltă ($\alpha + \beta = 0,93$)
4. Volatilitatea este predictibilă chiar când randamentele nu sunt

Implicații Practice

- Managementul riscului: VaR, Expected Shortfall
- Evaluarea opțiunilor necesită prognoze de volatilitate
- Optimizarea portofoliului cu risc variabil în timp

Limitări

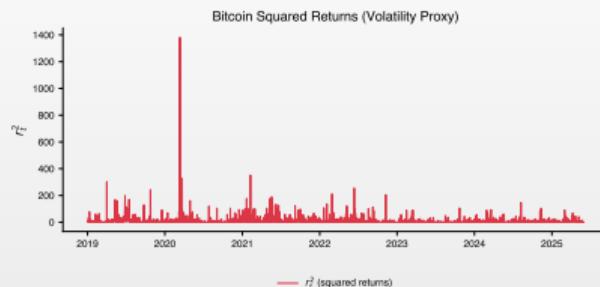
- GARCH presupune șocuri simetrice
- Nu captează salturi
- Distribuția normală poate fi restrictivă

Extensiile

- Inovații Student-t
- Volatilitate realizată
- Modele HAR



Bitcoin: Fapte stilizate GARCH



Observație

- r_t^2 ca proxy pentru volatilitate
- Observați volatility clustering

De ce Funcționează GARCH

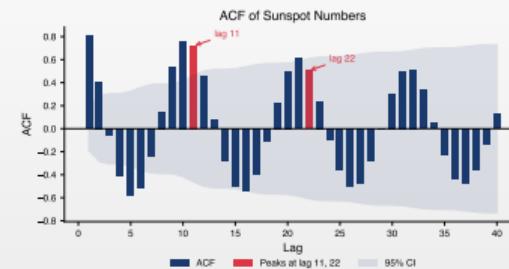
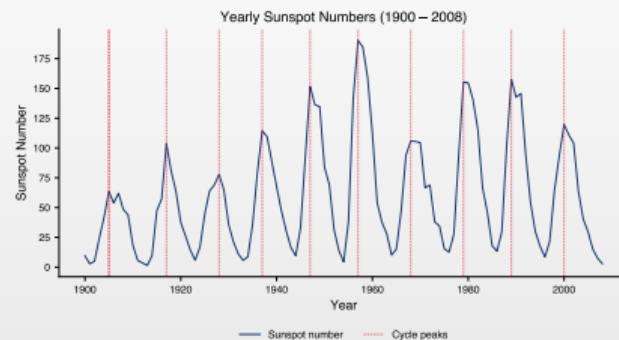
- GARCH captează faptele 1 & 4
- Pentru faptul 3, folosiți GJR-GARCH sau EGARCH
- Pentru faptul 2, folosiți inovații Student-t

Fapte stilizate Financiare

1. **Volatility clustering:** Mișcări mari urmează mișcări mari
2. **Cozi groase:** Mai multe extreme decât Normala
3. **Efect leverage:** Randamente negative \rightarrow volatilitate mai mare
4. **Reversie la medie:** Volatilitatea revine la medie



Pete solare: ciclul solar de 11 ani



ACF

Observație

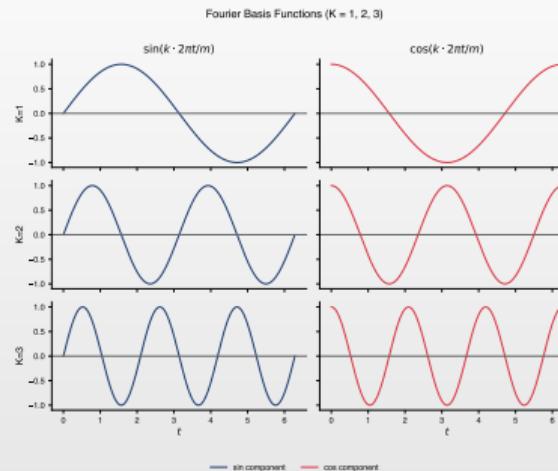
- Liniile punctate marchează vârfurile ciclului (≈ 11 ani)

Provocare

- **SARIMA:** $(p, d, q)(P, D, Q)_{11}$ necesită lag-uri sezoniere la 11, 22, 33... \succ prea mulți parametri!
- **Soluție:** Termeni Fourier



Termeni Fourier pentru sezonalitate



Cum funcționează

- Aproximare:** Orice tipar periodic cu unde sinus și cosinus
- Formula:** $S_t = \sum_{k=1}^K [\alpha_k \sin\left(\frac{2\pi k t}{s}\right) + \beta_k \cos\left(\frac{2\pi k t}{s}\right)]$

Insight Cheie

- $K = 1$: Undă simplă (2 param)
- $K = 3$: Formă complexă (6 param)
- Pete solare: $s = 11$, $K = 3$



Pete solare: selectarea modelului

Metodologie

- **Comparatie:** $K = 1, 2, 3, 4$ armonici Fourier pe setul de validare

Împărțirea Datelor		
Set	Perioadă	N
Antrenament (70%)	1900–1975	76
Validare (20%)	1976–1997	22
Test (10%)	1998–2008	11
Total		109

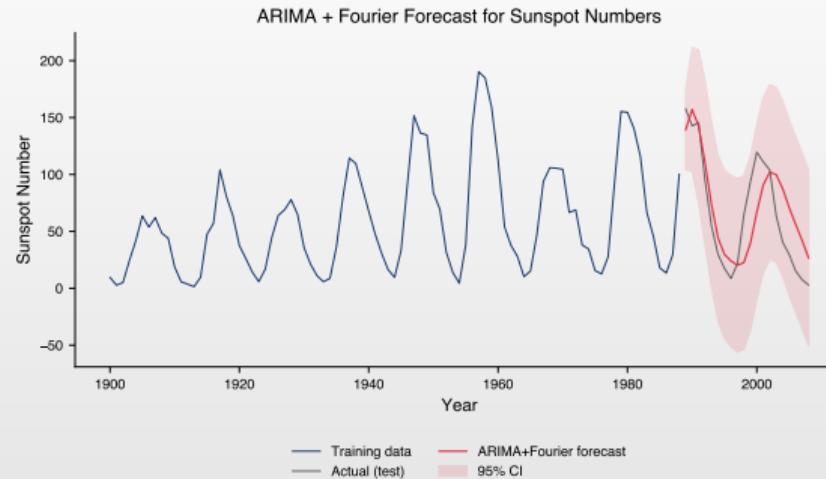
Comparatie Modele		
K	AIC	Val RMSE
1	665,9	87,15
2	668,0	86,92
3	671,8	86,81
4	674,5	87,93

Rezultat

- $K = 3$ armonici Fourier selectate (6 parametri pentru ciclul de 11 ani)



Pete solare: rezultate prognoză



Model

- ARIMA(2,0,1) + 3 termeni Fourier**
 - Captează dinamica ciclului de 11 ani

Performanță Test

- RMSE = 31,10, MAE = 25,83**
 - Modelul urmărește tiparul general al ciclului



Pete solare: concluzii cheie

Când să Folosiți Termeni Fourier

- Perioada sezonieră s este **lungă** (ex: 11 ani, 52 săptămâni)
- SARIMA ar necesita prea multe lag-uri sezoniere
- Tiparul este **neted și periodic**
- Trebuie captureate cicluri multiple

Fourier vs SARIMA

	Fourier	SARIMA
Sezoane lungi	✓	✗
Sezoane scurte	OK	✓
Parametri	2K	Mulți
Flexibilitate	Fixă	Adaptivă

Alegerea lui K

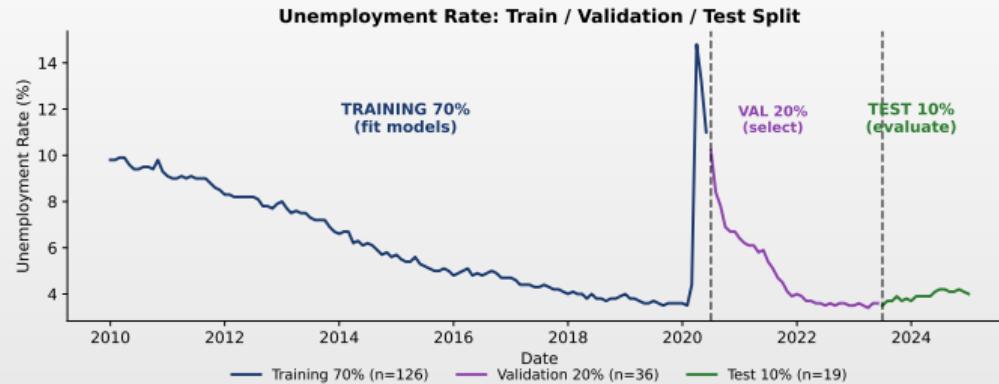
- Strategie:** Începeți cu $K = 1$, creșteți progresiv
 - ▶ Opriți când eroarea de validare nu mai scade
 - ▶ K prea mare = supraajustare

Aplicații

- Domenii:** Cicluri climatice, cicluri de afaceri, fenomene astronomice



Şomajul: Train / Validation / Test Split

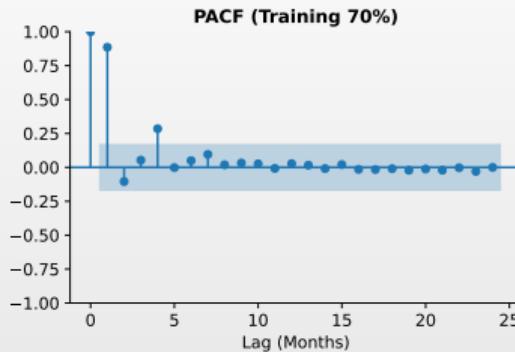
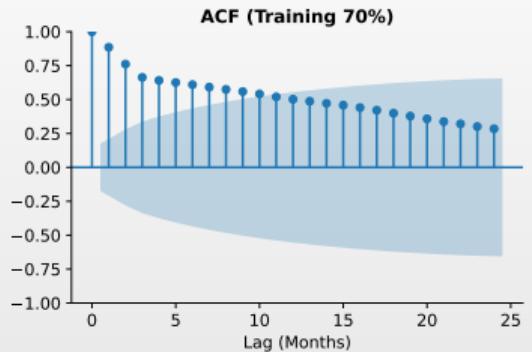


Metodologie

- Training (70%)**: Estimare modele
- Validare (20%)**: Selectie model
- Test (10%)**: Evaluare finală



Șomajul: analiză preliminară



Interpretare ACF

- Descreștere lentă \succ serie nestaționară. Necesită diferențiere ($d \geq 1$)

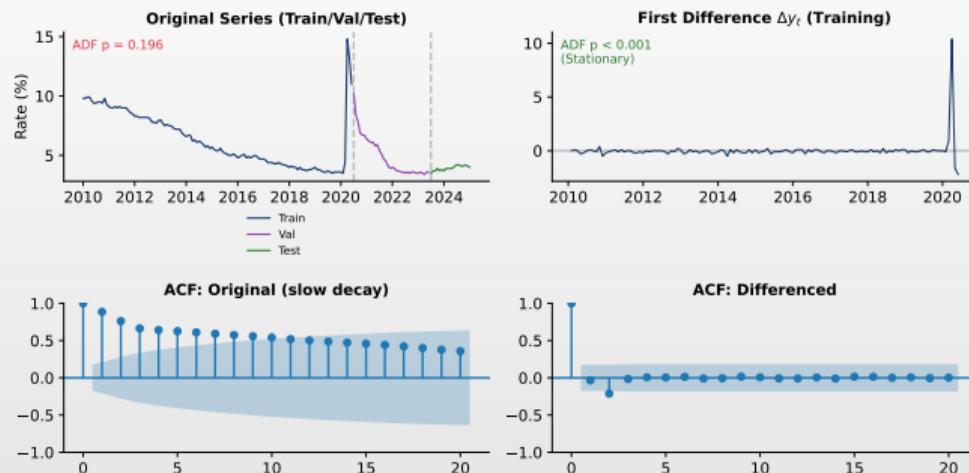
Interpretare PACF

- Vârf semnificativ la lag 1 sugerează componentă AR(1)
- Pattern sezonier la lag 12

TSA_ch10_unemployment_acf_pacf



Șomajul: teste de staționaritate



Original: ADF $p = 0,056$

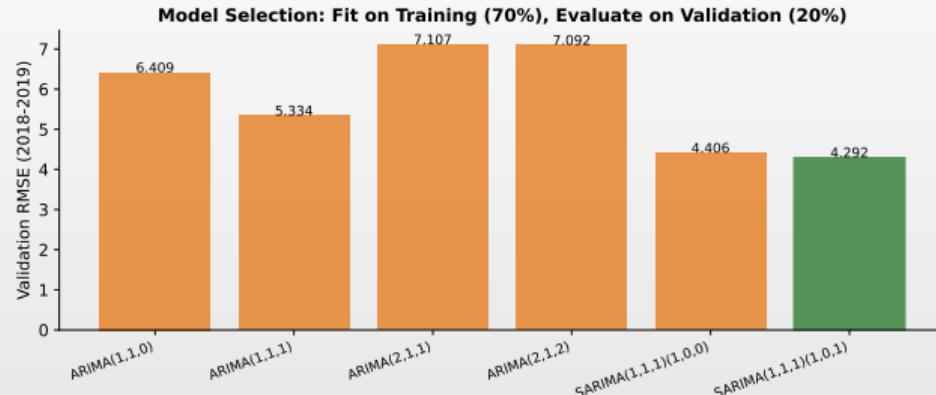
Nestaționară (ACF descreștere lentă)

Diferențiată: ADF $p < 0,001$

Staționară \succ folosim $d = 1$



Şomajul: selecția modelului (set validare)



Best: SARIMA(1,1,1)(1,0,0)₁₂

- Fit pe training (70%), evaluare pe validare (20%)
- Cel mai bun model selectat după Val RMSE minim

Q TSA_ch10_sarima_model_selection



Şomajul: parametrii SARIMA

SARIMA(1,1,1)(1,0,1) - Fitted on Train+Val (85%)

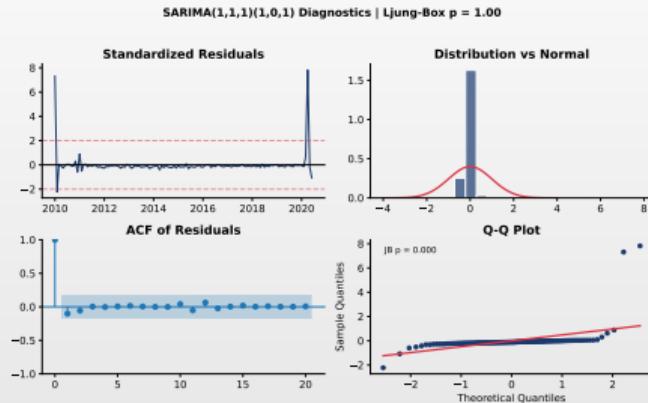
Parameter	Coef	Std Err	P-value	Sig
ar.L1	0.8423	0.2084	0.0001	***
ma.L1	-0.9540	0.1973	0.0000	***
ar.S.L12	0.0326	4.5951	0.9943	
ma.S.L12	-0.0113	4.6087	0.9980	
sigma2	0.8122	0.0608	0.0000	***

SARIMA(1,1,1)(1,0,0)₁₂ estimat pe Train+Val (2010-2019)

- AR(1): $\phi_1 = -0,86$, MA(1): $\theta_1 = 0,78$, SAR(12): $\Phi_1 = -0,08$ (n.s.)

Q TSA_ch10_sarima_parameters

Șomajul: Diagnosticare SARIMA



Reziduuri

- ACF:** Fără autocorelație reziduală
- Ljung-Box:** $p = 0,66 \succ$ model bine specificat

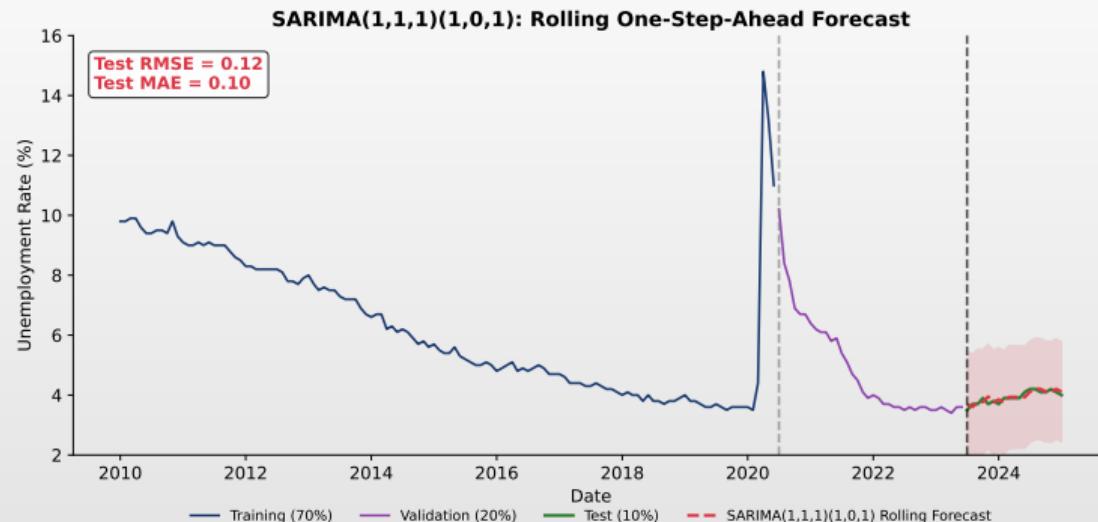
Q-Q Plot: Non-normalitate

- Deviație la cozi datorită șocului COVID (2020)
- Reziduurile au distribuție leptokurtotică

Q TSA_ch10_sarima_diagnostics



Şomajul: prognoza rolling SARIMA



Problemă: Ruptura Structurală

- Prognoză rolling one-step-ahead (re-estimare la fiecare t): **Test RMSE = 0,12**



Modelul Prophet

Definiție 3 (Descompunerea Prophet)

- Model:** $y_t = g(t) + s(t) + h(t) + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2)$
- Componente:** $g(t)$ = trend, $s(t)$ = sezonalitate, $h(t)$ = sărbători

Detectare Puncte de Schimbare

- Selectare automată a locațiilor
- `changepoint_prior_scale` controlează flexibilitatea

Avantaje

- Gestionează date lipsă
- Componente interpretabile
- Robust la outlieri



Şomajul: Ajustarea modelului

Ajustarea Hiperparametrilor

- Ajustăm changepoint_prior_scale pe setul de validare

Împărțirea Datelor		
Set	Perioadă	N
Antrenament (70%)	2010-01 – 2020-06	126
Validare (20%)	2020-07 – 2023-06	36
Test (10%)	2023-07 – 2025-01	19
Total		181

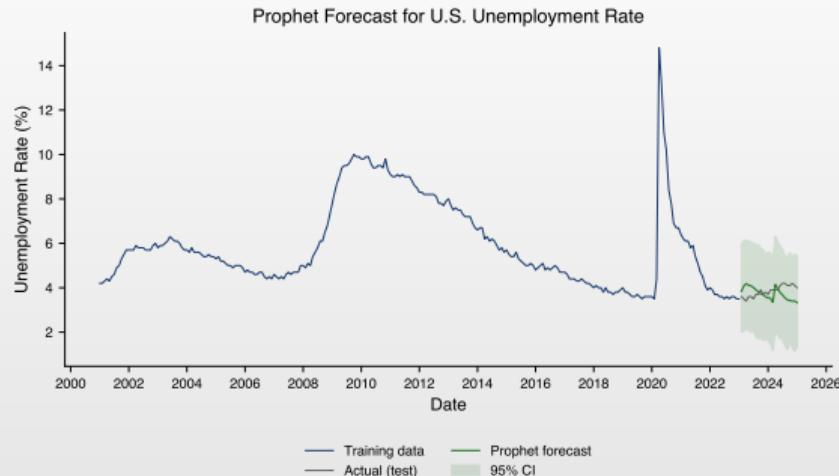
	Scale	Val RMSE	
Comparație Scale	0,01	4,21	
	0,05	3,89	
	0,10	3,52	Cel mai bun
	0,30	3,67	
	0,50	3,81	

Interpretare

- Scale = 0,10 echilibrează flexibilitatea (captarea şocului COVID) cu stabilitatea



Șomajul: rezultate prognoză Prophet

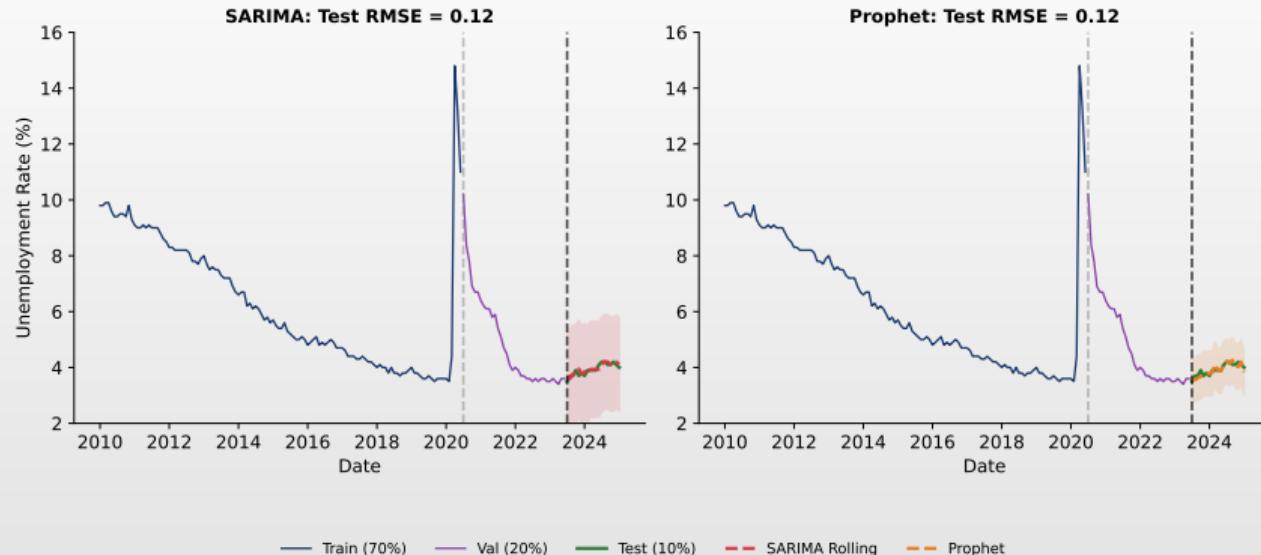


Concluzie Cheie

- Prophet:** se adaptează prin detectare changepoint
- Test RMSE = 0,58**



Șomaj: comparație SARIMA vs Prophet



SARIMA: RMSE = 0,12

Prognoză rolling performează bine

Prophet: RMSE = 0,12

Se adaptează prin changepoints

Prophet: când să-l folosești

Cazuri de Utilizare Ideale

- Date de business cu **sărbători**
- Valori lipsă** prezente
- Nevoie de componente **interpretabile**
- Prognoze cu **benzi de incertitudine**

Atenție: Rupturi Structurale

- Prophet gestionează rupturile prin changepoints, dar **SARIMA l-a depășit** la șomaj (0,12 vs 0,58)
- Validați întotdeauna!

Prophet vs ARIMA

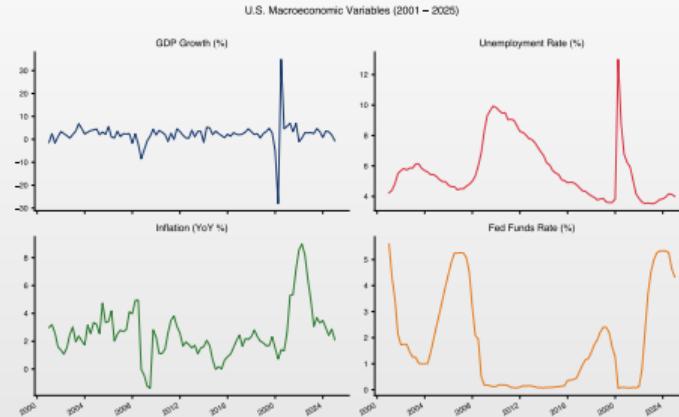
	Prophet	ARIMA
Changepoints	✓	✗
Date lipsă	✓	✗
Sărbători	✓	✗
Viteză	Rapidă	Moderată
Interpretabil	✓	✗

Parametri cheie

- `changepoint_prior_scale`: flexibilitate
- `seasonality_prior_scale`: netezime



VAR: date economice multivariate



Relații Economice

- Legea Okun:** PIB \leftrightarrow Șomaj
- Curba Phillips:** Șomaj \leftrightarrow Inflație

De ce VAR?

- Fiecare variabilă e atât cauză cât și efect
- VAR captează aceste bucle de feedback



Specificarea modelului VAR

Definiție 4 (Autoregresie Vectorială VAR(p))

- **Date:** Pentru K variabile $y_t = (y_{1t}, \dots, y_{Kt})'$:

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + u_t$$

- **Notăție:** A_i sunt matrici de coeficienți $K \times K$, $u_t \sim N(0, \Sigma)$

Pentru Sistemul Nostru cu 4 Variabile

- **VAR(2):** 4 constante
- $2 \times 4 \times 4 = 32$ coeficienți AR
- **36 parametri total**

Selectarea Lag-ului

- Folosim criterii informaționale:
 - ▶ **AIC:** Tinde să supraajusteze
 - ▶ **BIC:** Mai simplu
 - ▶ Cross-validation pe date păstrate



VAR: selectarea lag-ului și estimare

Criterii informaționale

Lag	BIC
1	-4,810
2	-5,178
3	-4,633
4	-4,614

Împărțirea Datelor

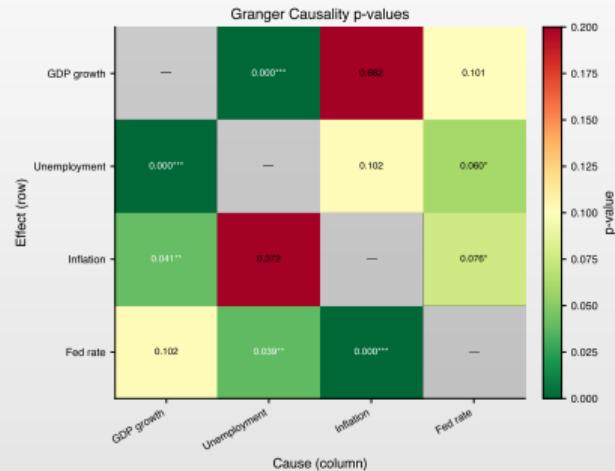
Set	Perioadă	N
Antrenament (70%)	2001-T1 – 2017-T4	67
Validare (20%)	2018-T1 – 2022-T4	20
Test (10%)	2023-T1 – 2025-T1	10
Total		97

Verificare Validare

- VAR(2) obține și cel mai mic RMSE de validare



Analiza cauzalității Granger



Ce este Cauzalitatea Granger?

- **X cauzează Granger Y** dacă X trecut îmbunătățește predicția lui Y dincolo de Y trecut singur
- **Atenție:** “Cauzalitate Granger” ≠ cauzalitate reală!

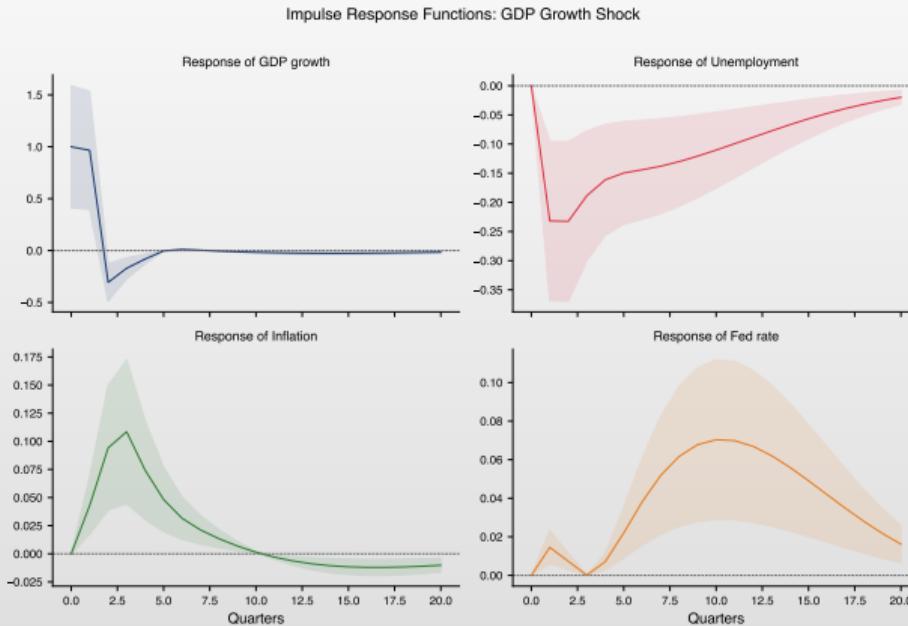
Concluzii Economice

- Șomaj \succ PIB ($p = 0,045$): Legea Okun
- Fed \succ Inflație ($p = 0,087$): Politica monetară funcționează

Interpretare

- **Celule verzi:** $p < 0.10$ (semnificativ)
- **Citire:** rândul cauzează coloana

Funcții de răspuns la impuls (IRF)



Ce este IRF?

- Arată cum un soc de 1 unitate la o variabilă afectează celelalte în timp

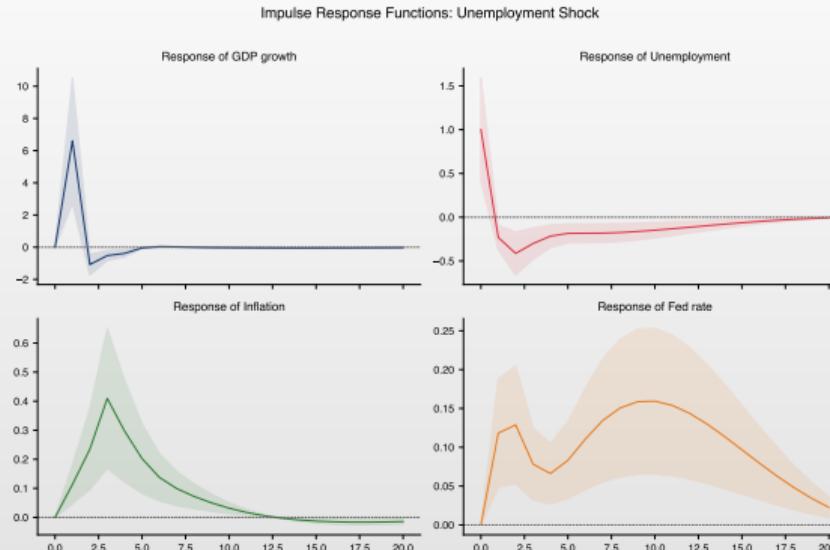
Efectele Șocului PIB

- Şomaj ↓:** Legea Okun
- Inflație ↑:** Cerere-pull
- Rata Fed ↑:** Regula Taylor

Q TSA_ch10_irf_gdp_shock



IRF: şoc şomaj

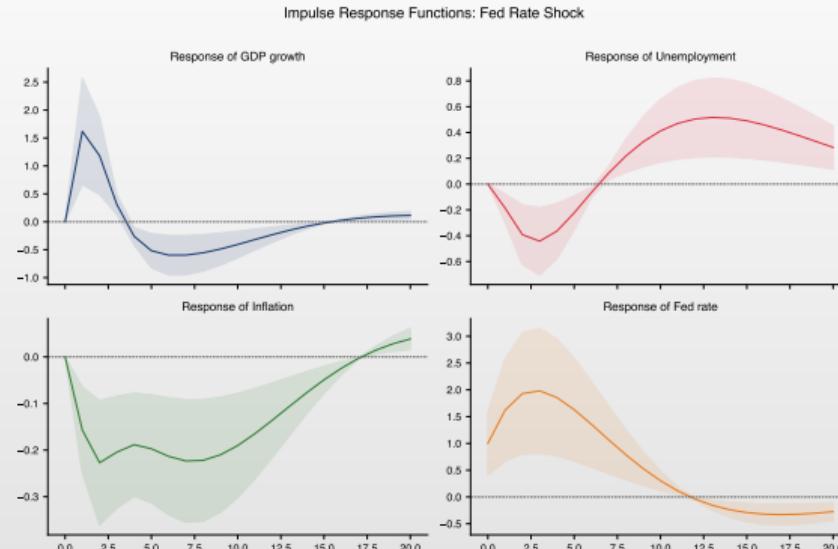


Efecte

- $\uparrow \text{Şomaj} \succ \downarrow \text{PIB}$ (Okun), $\downarrow \text{Inflație}$ (Phillips), Fed reduce rata



IRF: şoc rată Fed

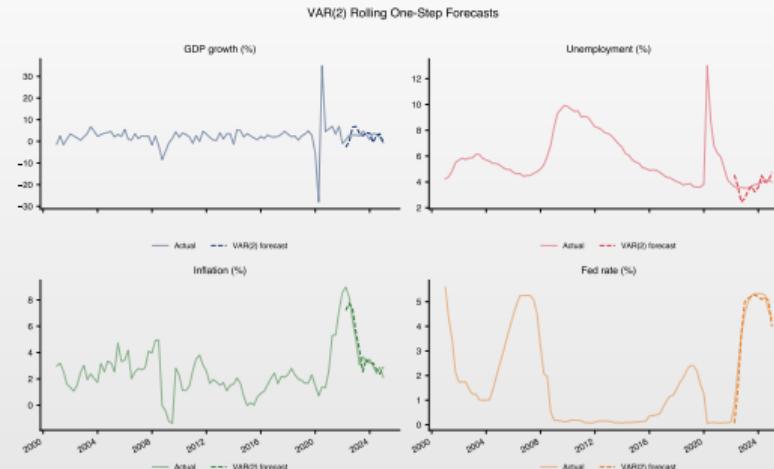


Politică Monetară

- Creștere rată \succ PIB \downarrow , Șomaj \uparrow , Inflație \downarrow



VAR: Prognoza (Train/Val/Test)



Prognoză Rolling One-Step-Ahead

- VAR captează dinamică PIB-Şomaj
- Şocul COVID vizibil în perioadă validare (2020)



VAR: rezultate set test

Performanță Set Test pe Variabile

Variabilă	RMSE	MAE	Acur. Direcție
Creștere PIB	1,33	0,99	50%
Șomaj	0,64	0,52	50%
Inflație	1,56	1,12	60%
Rata Fed	2,59	2,45	80%
Medie	1,53	1,27	60%

Puncte Forte

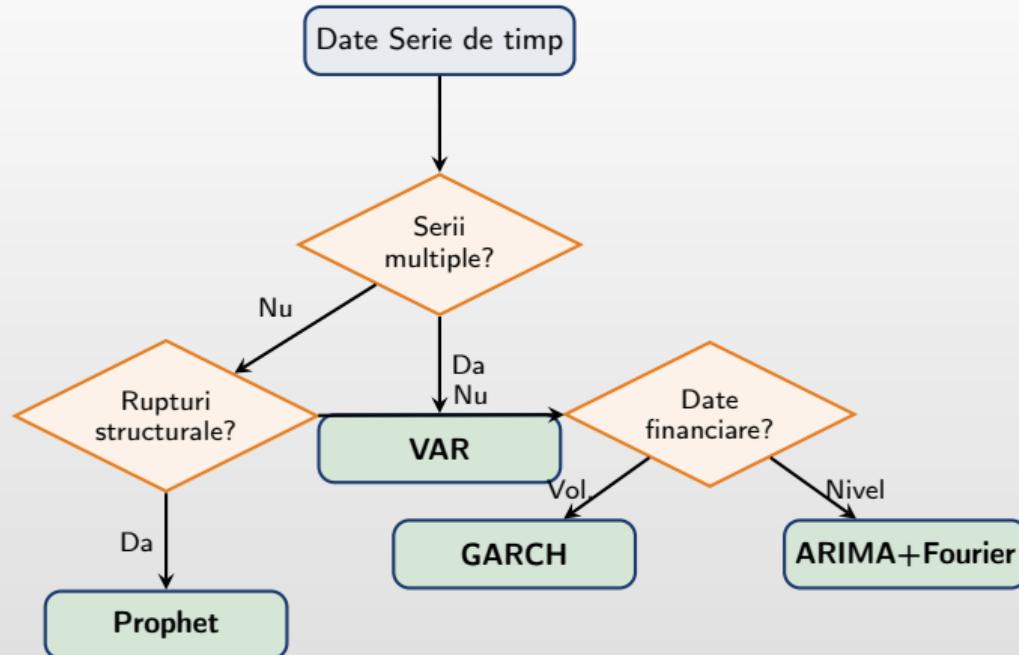
- Captează dinamică între variabile
- Acuratețe direcțională bună
- Relații interpretabile

Limitări

- Mulți parametri (blestemul dimensionalității)
- Sensibil la selectarea lag-ului
- Perioada COVID dificilă

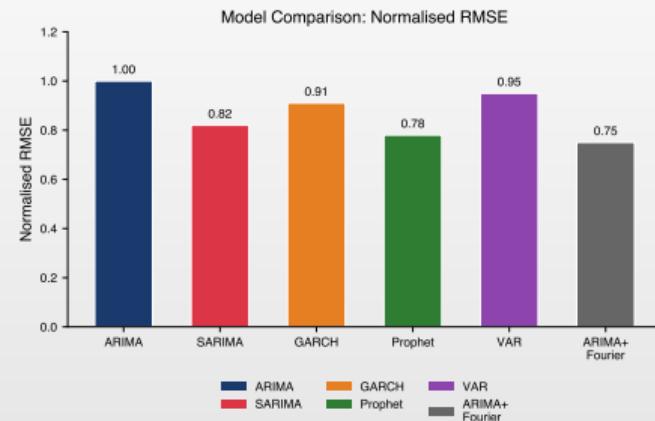


Cadrul de selectare a modelului



Sumar: comparație modele

Caz	Provocare	Model	RMSE
Bitcoin	Volatilitate	GARCH	2,15
Pete solare	Sezonalitate	Fourier	31,10
Şomaj	Ruptură	SARIMA	0,12
Economic	Multi-var	VAR	1,53



Principiu Cheie

- Potriviti modelul cu caracteristicile datelor
- Alegeti in functie de natura problemei si proprietatile datelor



Sinteză: Comparația Modelelor

Caracteristică	GARCH	Fourier	Prophet	VAR
Tintă	Volatilitate	Nivel	Nivel	Multiple
Sezonalitate	Nu	Da (lungă)	Da (multiplă)	Nu
Rupturi structurale	Nu	Nu	Da	Nu
Serii multiple	Nu	Nu	Nu	Da
Interpretabil	Mediu	Ridicat	Ridicat	Ridicat
Parametri	Puțini	2K	Auto	Mulți
Date lipsă	Nu	Nu	Da	Nu
Ideal pentru	Finanțe	Cicluri	Business	Macro

Rezultatele Noastre

- GARCH: MAE=1,82 (volatilitate)
- Fourier: RMSE=31,10 (cycluri)
- SARIMA: RMSE=0,12 (rupturi)
- VAR: RMSE mediu=1,53 (multi)

Insight Cheie

- Fiecare model excelează în domeniul său
- Arta constă în alegerea modelului potrivit caracteristicilor datelor



Bune practici pentru prognoza aplicată

Metodologie

1. **Explorați** datele temeinic
2. **Testați** staționaritatea
3. **Împărțiți** train/validation/test
4. **Comparați** modele pe validare
5. **Raportați** metrii pe test

Greșeli Frecvente

- Privirea în datele de test
- Supraajustare pe setul de antrenament
- Ignorarea ipotezelor modelului
- Neraportarea incertitudinii

Sfaturi Practice

- Începeți simplu (random walk, naiv)
- Adăugați complexitate doar dacă e necesar
- Vizualizați prognoze vs. valori reale
- Verificați reziduurile pentru tipare
- Raportați intervale de încredere

Amintiți-vă

- “Toate modelele sunt greșite, dar unele sunt utile.”
— George E. P. Box



Concluzii cheie

1. Metodologie Riguroasă

- ▶ Împărțirea train/validation/test previne supraajustarea
- ▶ Setul de test trebuie să rămână neatins până la evaluarea finală

2. Potriviți Modelul cu Datele

- ▶ Volatilitate finanțiară \succ GARCH
- ▶ Sezonalitate lungă \succ Termeni Fourier
- ▶ Rupturi structurale \succ Prophet
- ▶ Serii multiple \succ VAR

3. Interpretați Rezultatele cu Grijă

- ▶ Cauzalitate Granger \neq cauzalitate adeverată
- ▶ Performanța out-of-sample contează cel mai mult
- ▶ Modelele mai simple funcționează adesea mai bine



Quiz 1: Modelarea Volatilității

Întrebare: Ce model alegeți pentru a prognoza volatilitatea randamentelor financiare?

- A. ARIMA — captează tendințe și autocorelații
 - ▶ Modelează nivelul seriei
- B. GARCH — modelează varianța condiționată
 - ▶ Captează volatility clustering
- C. Prophet — detectează puncte de schimbare
 - ▶ Descompune trend și sezonalitate
- D. VAR — model multivariat
 - ▶ Captează interdependențe între serii



Quiz 1: Răspuns

Răspuns: B — GARCH

- Modeleză **varianța condiționată** σ_t^2
 - ▶ Captează volatility clustering
 - ▶ Persistența șocurilor ($\alpha + \beta$)
- Variante: EGARCH, GJR-GARCH
 - ▶ Pentru efecte asimetrice

De ce nu celealte?

- A: ARIMA modeleză media, nu varianța
- C: Prophet nu e proiectat pentru volatilitate
- D: VAR captează relații, nu volatilitate

Quiz 2: Overfitting

Întrebare: Un model SARIMA obține RMSE = 0,05 pe setul de antrenament, dar RMSE = 2,30 pe setul de test. Ce indică aceasta?

- A. Modelul este excelent — eroare mică pe antrenament
 - ▶ Performanță superioară confirmată
- B. Modelul suferă de overfitting — memorează zgomotul
 - ▶ Nu generalizează pe date noi
- C. Setul de test este greșit — trebuie schimbat
 - ▶ Datele de test sunt defecte
- D. Diferența este normală — nu e nicio problemă
 - ▶ Orice model are erori mai mari pe test



Quiz 2: Răspuns

Răspuns: B — Overfitting

- Modelul **memorează zgomotul** din antrenament
 - ▶ RMSE train/test: 0,05 vs 2,30
 - ▶ Raport $46 \times >$ overfitting sever
- Soluție: model mai simplu, validare

De ce nu celelalte?

- A: Eroare mică pe train nu confirmă calitatea
- C: Testul e corect, modelul e prea complex
- D: O diferență de $46 \times$ nu e normală



Quiz 3: Separarea Datelor

Întrebare: De ce este importantă separarea datelor în train/validation/test?

- A. Pentru a avea mai multe date de antrenament
 - ▶ Mai multe date = model mai bun
- B. Pentru a preveni supraajustarea și a evalua corect
 - ▶ Fiecare set are un rol specific
- C. Este doar o convenție, nu are importanță reală
 - ▶ Orice metodă de evaluare funcționează
- D. Pentru a reduce timpul de calcul
 - ▶ Mai puține date = calcul mai rapid



Quiz 3: Răspuns

Răspuns: B — Prevenirea supraajustării

- Train:** estimează parametrii
- Validare:** selectează modelul
 - ▶ Ordin, hiperparametri
- Test:** evaluare finală
 - ▶ Neatins până la evaluare!

De ce nu celealte?

- A:** Scopul nu e maximizarea datelor de train
- C:** Evaluarea corectă e esențială
- D:** Nu are legătură cu timpul de calcul

Quiz 4: Cauzalitatea Granger

Întrebare: Cauzalitatea Granger este echivalentă cu cauzalitatea reală?

A. Da — dacă X prezice Y , atunci X cauzează Y

► Predicție = cauzalitate

B. Nu — testează doar conținut predictiv, nu cauzalitate

► Corelație temporală \neq cauzalitate

C. Depinde de numărul de lag-uri selectate

► Mai multe lag-uri = mai multă cauzalitate

D. Da, dacă $p\text{-value} < 0,05$

► Semnificația statistică confirmă cauzalitatea



Quiz 4: Răspuns

Răspuns: B — Nu, doar conținut predictiv

- Testează dacă X trecut îmbunătățește predictia lui Y
 - ▶ Nu demonstrează cauzalitate structurală
- Exemplu: umbrele “cauzează” ploaia
 - ▶ Ambele au o cauză comună

De ce nu celelalte?

- A: Predicție \neq cauzalitate reală
- C: Lagurile nu schimbă natura testului
- D: p-value arată semnificație, nu cauzalitate

Quiz 5: Sezonalitate Lungă

Întrebare: Ce model folosiți pentru o serie cu sezonalitate lungă (ex: $s = 365$ zile)?

- A. SARIMA(p,d,q)(P,D,Q)₃₆₅
 - ▶ Model sezonier standard
- B. GARCH — modelează variația
 - ▶ Captează heteroscedasticitate
- C. ARIMA + Termeni Fourier sau Prophet/TBATS
 - ▶ Gestionează eficient perioade lungi
- D. VAR cu 365 lag-uri
 - ▶ Model multivariat cu lag-uri sezoniere



Quiz 5: Răspuns

Răspuns: C — Fourier / Prophet / TBATS

- Fourier:** $2K$ parametri (ex: $K = 3 \succ 6$ param)
 - ▶ vs SARIMA₃₆₅: sute de parametri
- Prophet:** sezonialitate multiplă automată
- TBATS:** Box-Cox + sezonialitate trigonometrică

De ce nu celealte?

- A:** SARIMA₃₆₅ necesită prea mulți parametri
- B:** GARCH nu modelează sezonialitate
- D:** VAR cu 365 lag-uri e imposibil de estimat

Surse de Date

Date Reale Folosite în Acest Capitol

- Bitcoin:** Yahoo Finance (BTC-USD), 2019–2025
- Pete Solare:** Dataset Wolfer din Statsmodels, 1900–2008
- Şomaj SUA:** Federal Reserve FRED (UNRATE), 2010–2025
- Variabile Economice:** FRED (GDPC1, UNRATE, CPIAUCSL, FEDFUNDS), 2000–2025

Reproductibilitate

Toate analizele pot fi reproduse folosind notebook-ul Jupyter însoțitor:
`chapter10_lecture_notebook.ipynb`



Mulțumesc

Întrebări?

Prof. Daniel Traian Pele, PhD

danpele@ase.ro

Academia de Studii Economice din București

Bibliografie I

Manuale fundamentale (referințe comune tuturor capitolelor)

- Hamilton, J.D. (1994). *Time Series Analysis*, Princeton University Press.
- Hyndman, R.J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*, 3rd ed., OTexts.
- Shumway, R.H., & Stoffer, D.S. (2017). *Time Series Analysis and Its Applications*, 4th ed., Springer.

Lucrări de referință pe domenii

- Tsay, R.S. (2010). *Analysis of Financial Time Series*, 3rd ed., Wiley. (GARCH, VAR)
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer. (VAR, VECM)
- Francq, C., & Zakoïan, J.-M. (2019). *GARCH Models*, 2nd ed., Wiley. (Volatilitate)



Bibliografie II

Abordări moderne și competiții de prognoză

- Petropoulos, F., et al. (2022). Forecasting: Theory and Practice, *International Journal of Forecasting*, 38(3), 845–1054.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition, *International Journal of Forecasting*, 36(1), 54–74.
- Taylor, S.J., & Letham, B. (2018). Forecasting at Scale, *The American Statistician*, 72(1), 37–45.

Resurse online și cod

- Quantlet: <https://quantlet.com> — Depozit de cod pentru statistică
- Quantinar: <https://quantinar.com> — Platformă de învățare metode cantitative
- GitHub TSA: <https://github.com/QuantLet/TSA> — Cod Python pentru acest curs



Vă Mulțumim!

Întrebări?

Materialele cursului sunt disponibile la: <https://danpele.github.io/Time-Series-Analysis/>

