



Analiza și Prognoza Seriilor de Timp

## Capitolul 10: Recapitulare Completă

Studii de Caz Aplicate cu Metodologie Riguroasă



# Cuprins

- 1 Metodologia Prognozei
- 2 Studiu de Caz 1: Volatilitatea Bitcoin (GARCH)
- 3 Studiu de Caz 2: Ciclurile Petelor Solare (Fourier)
- 4 Studiu de Caz 3: Șomajul (Prophet)
- 5 Studiu de Caz 4: Analiză Multivariată (VAR)
- 6 Sinteză și Ghid

## 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

## Time Series Train/Validation/Test Split



### Set Antrenament

- Estimare parametri
- Cea mai mare parte

### Set Validare

- Comparare modele
- Ajustare hiperparam

### Set Test

- Păstrat
- Metrici finale

## Definiție 1 (Metrici ale Erorii de Prognoză)

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**

## 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:  $\approx 2.200$  zile

### Fapte Stilizate

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

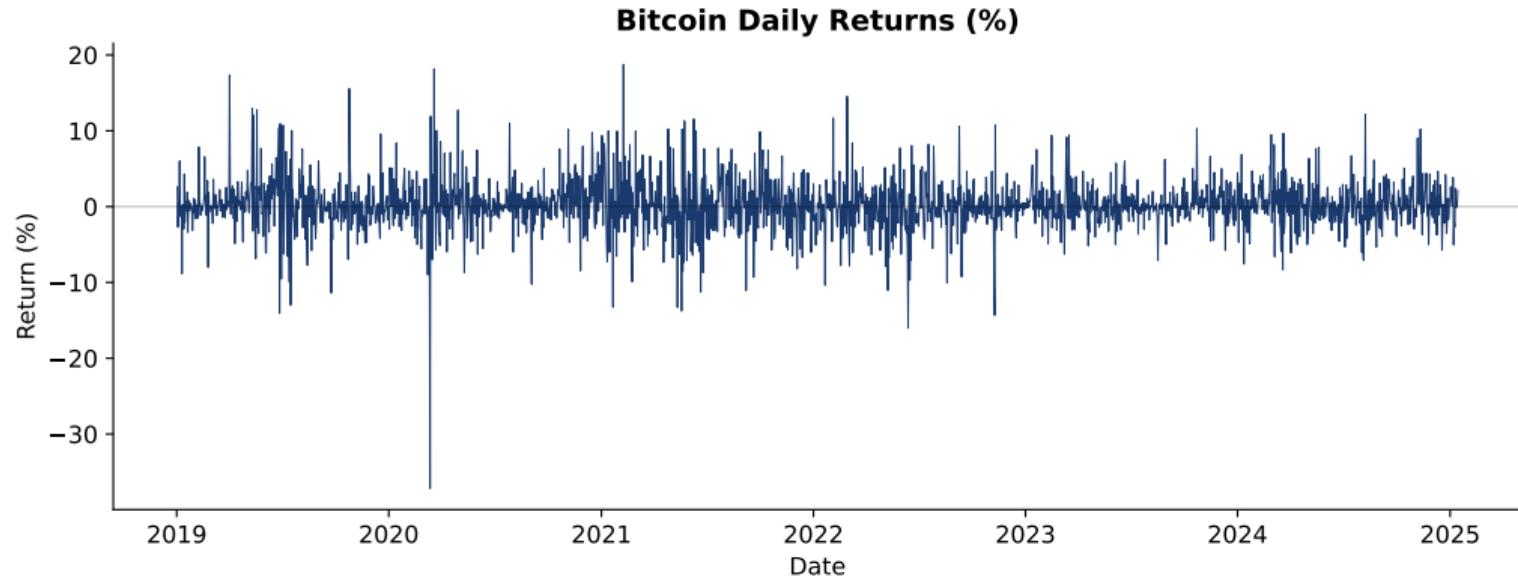
### Insight Cheie

Randamentele financiare sunt de obicei:

- **Impredictibile** în medie
- **Predictibile** în varianță

⇒ Focus pe **prognoza volatilității**

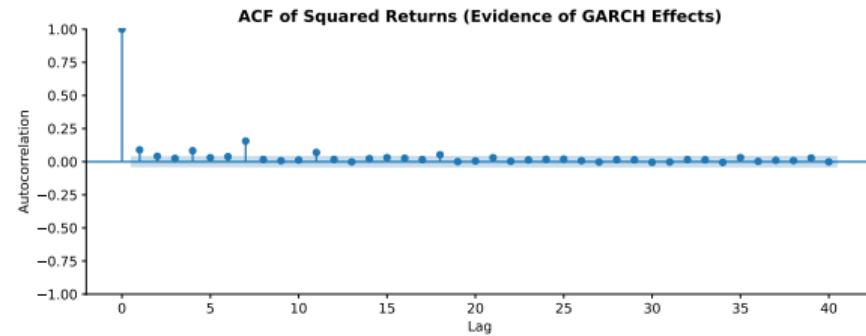
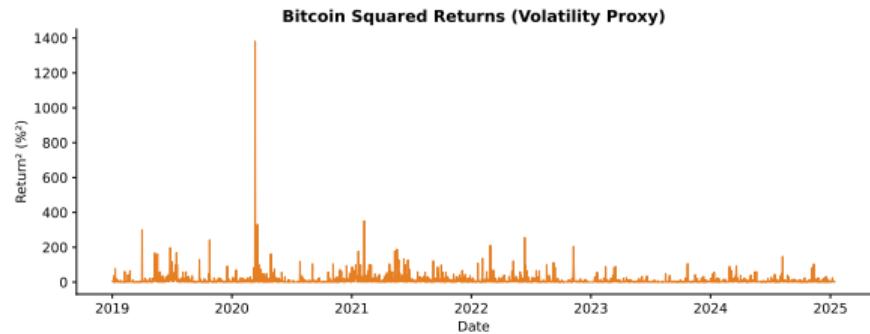
## Bitcoin: Clustering-ul Volatilității



### Observație

Randamentele mari tind să urmeze randamente mari, cele mici urmează cele mici. Acesta este **clustering-ul volatilității**—fenomenul pe care GARCH îl captează.

# Bitcoin: Dovezi pentru GARCH



Randamentele pătrate  $r_t^2$  sunt proxy pentru volatilitate  $\sigma_t^2$ .  
Vârfurile se grupează.

Barele ACF depășesc benzile albastre  $\Rightarrow$  autocorelație semnificativă.

## De ce GARCH?

Dacă  $r_t^2$  ar fi zgomot alb, ACF ar fi zero. ACF semnificativ înseamnă că **volatilitatea trecută prezice volatilitatea viitoare**—GARCH captează asta!

### Definiție 2 (Modelul GARCH(p,q))

Fie  $r_t$  randamentele. Modelul GARCH(p,q) este:

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

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

unde  $\omega > 0$ ,  $\alpha_i \geq 0$ ,  $\beta_j \geq 0$ , 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**: Șocuri asimetrice

#### Interpretare

- $\alpha$ : Impactul șocurilor trecute
- $\beta$ : 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
<b>Total</b>	<b>2.205</b>	

## 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. 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.

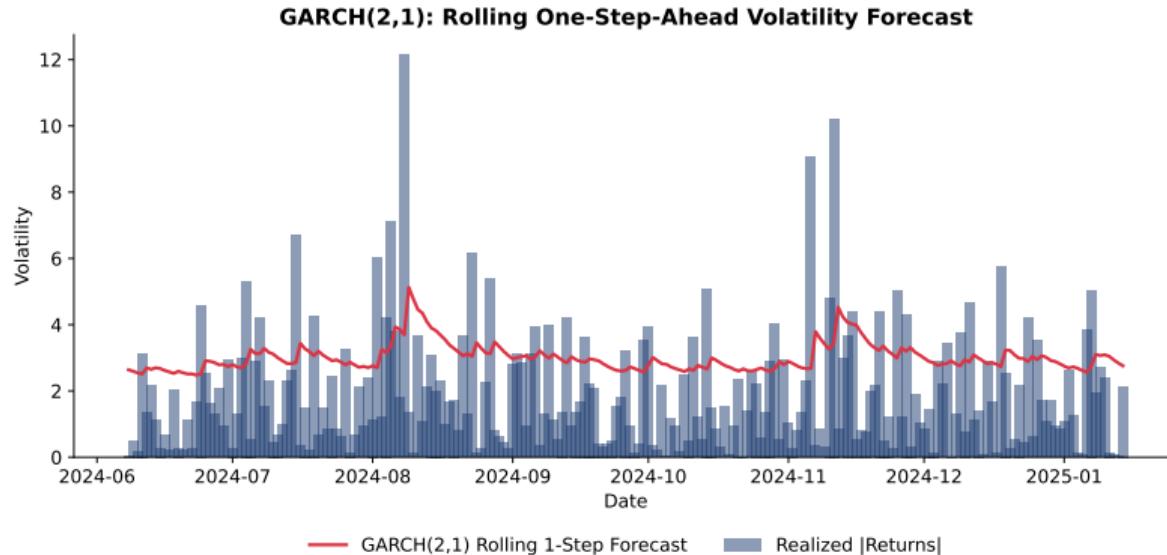
Model	AIC	BIC	Val MAE	Selectare
GARCH(1,1)	6.994,8	7.020,6	<b>2,638</b>	Cel mai bun
GARCH(2,1)	6.993,7	7.024,6	2,640	
GJR-GARCH(1,1)	6.983,7	7.014,6	2,669	
EGARCH(1,1)	—	—	—	Eșuat*

\* 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

$$\omega = 0,87, \alpha = 0,09, \beta = 0,84$$

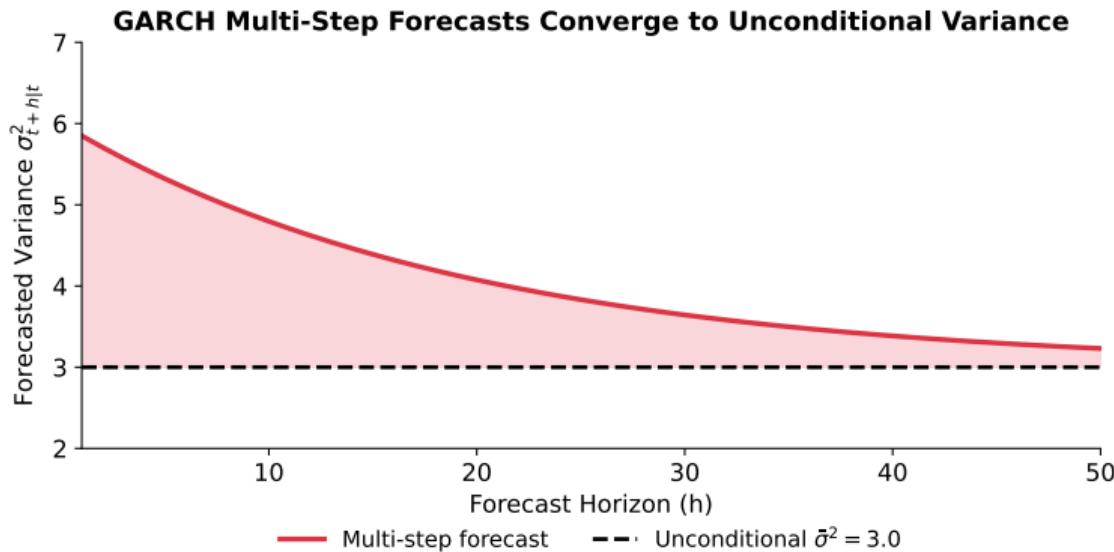
$$\alpha + \beta = 0,93 \text{ (persistență înaltă)}$$

### Performanță Test

$$MAE = 1,82, RMSE = 2,14$$

Prognoza urmărește bine volatilitatea realizată.

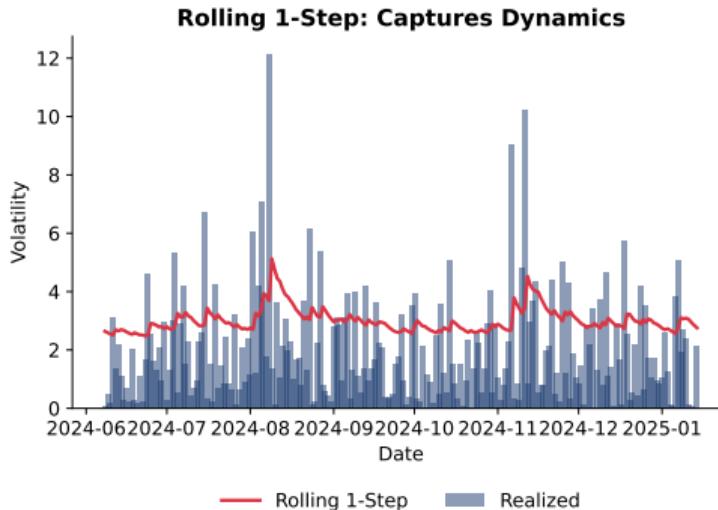
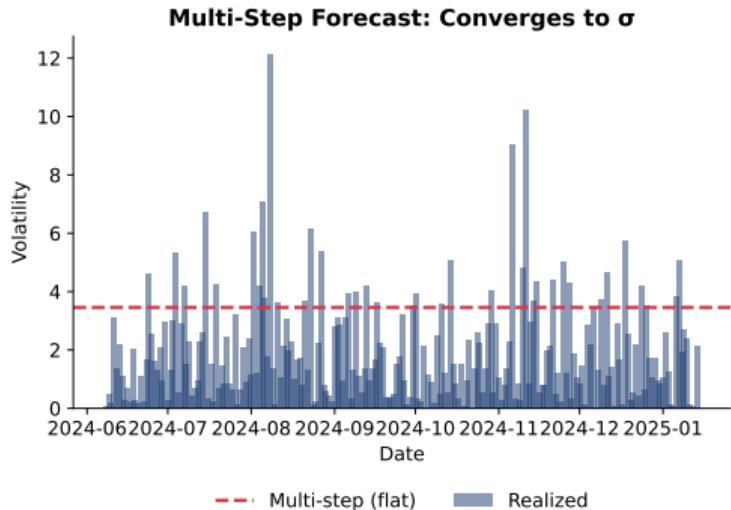
## GARCH: Prognozele Multi-Step Converg



### Insight Cheie

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

# GARCH: Soluția Rolling One-Step-Ahead



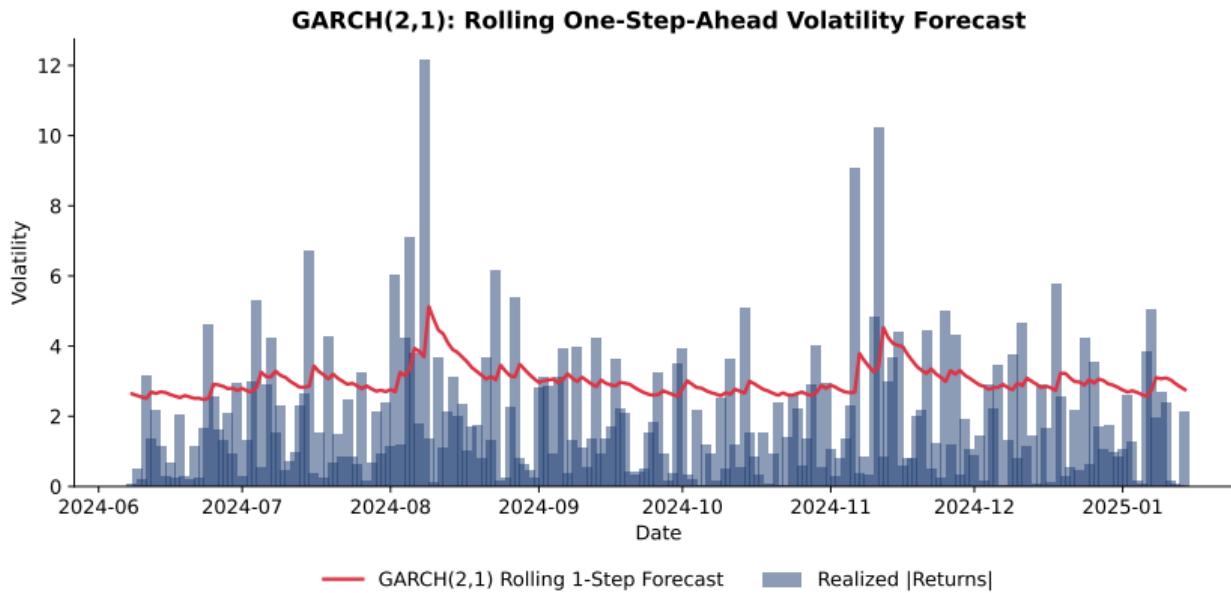
Multi-Step (Stânga)

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

Rolling 1-Step (Dreapta)

Re-estimare la fiecare  $t$  (dinamic)

## Bitcoin: Prognoza Volatilității GARCH (Set Test)



### Rezultat

Prognozele rolling one-step-ahead GARCH(1,1) captează **tiparele dinamice ale volatilității**. Linia roșie urmărește volatilitatea realizată (zona albastră).

## Sumar

- ❶ Randamentele sunt staționare; prețurile nu sunt autoregresive
- ❷ GARCH(1,1) depășește variantele mai complexe
- ❸ Persistență înaltă ( $\alpha + \beta = 0,93$ )
- ❹ Volatilitatea este **predictibilă** chiar când randamentele nu sunt autoregresive

## Limitări

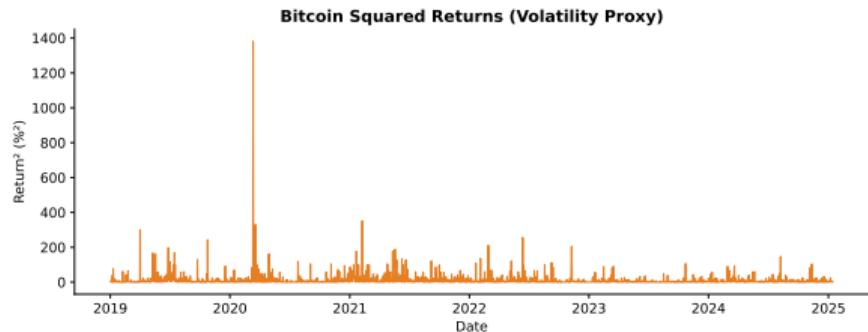
- GARCH presupune șocuri simetrice
- Nu captează **salturi**
- Distribuția normală poate fi restrictivă

## Implicații Practice

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

## Extensiile

- Inovații Student-t
- Volatilitate realizată
- Modele HAR



Randamente pătrate  $r_t^2$  ca proxy pentru volatilitate.  
Observați clustering-ul perioadelor de volatilitate înaltă.

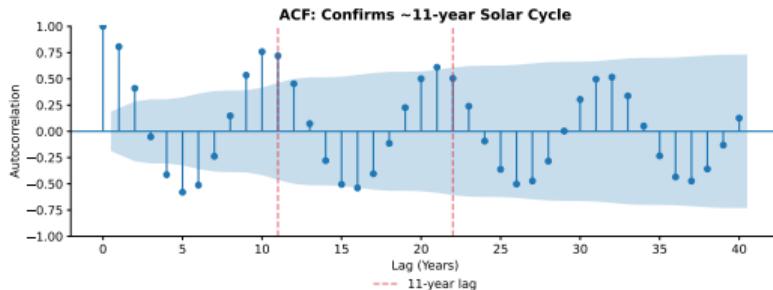
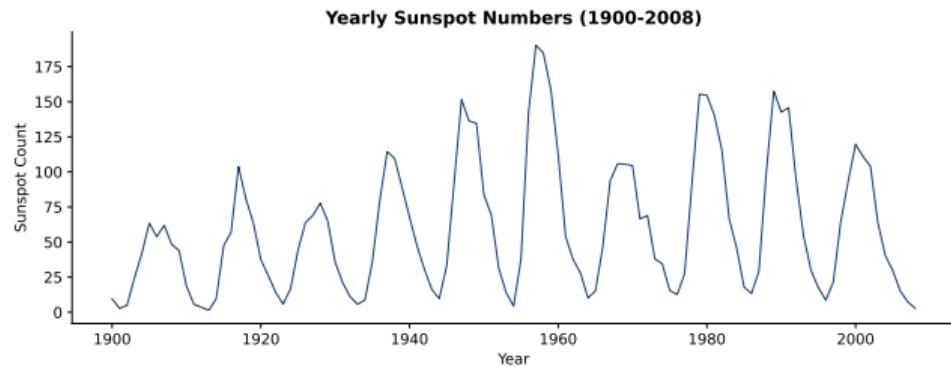
## Fapte Stilizate Financiare

- ❶ **Clustering volatilitate:** Mișcări mari urmează mișcări mari
- ❷ **Cozi groase:** Mai multe evenimente extreme decât prezice Normala
- ❸ **Efect leverage:** Randamente negative → volatilitate mai mare
- ❹ **Reversie la medie:** Volatilitatea revine la nivelul pe termen lung

## De Ce Funcționează GARCH

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

# Pete Solare: Ciclul Solar de 11 Ani



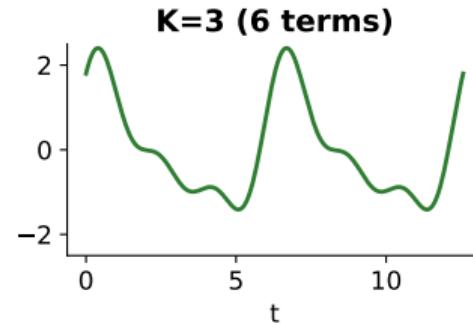
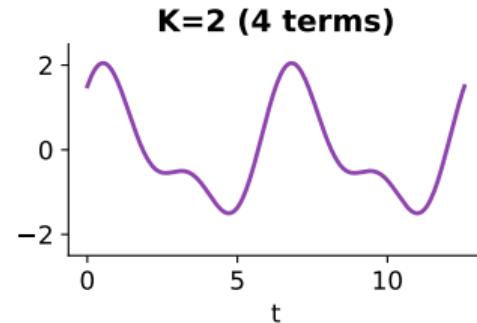
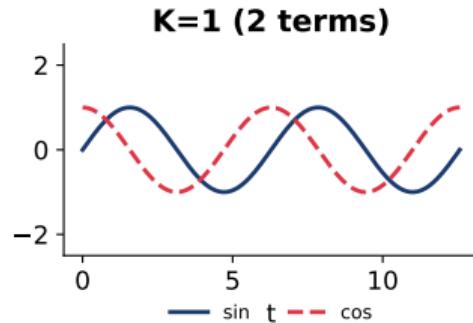
ACF are vârfuri la lag 11 și 22, confirmând periodicitatea ciclului solar.

Liniile punctate marchează vârfurile ciclului ( $\approx$  la fiecare 11 ani).  
Amplitudinea variază.

## Provocare

SARIMA( $p, d, q$ )( $P, D, Q$ )<sub>11</sub> necesită estimarea lag-urilor sezoniere la 11, 22, 33... Prea mulți parametri!  
**Soluție:** Termeni Fourier.

## Fourier Terms: More K = More Flexibility



### Cum Funcționează

Aproximăm orice tipar periodic folosind unde sinus și cosinus:  
$$S_t = \sum_{k=1}^K [\alpha_k \sin\left(\frac{2\pi k t}{s}\right) + \beta_k \cos\left(\frac{2\pi k t}{s}\right)]$$

### Insight Cheie

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

## Metodologie

Comparăm  $K = 1, 2, 3, 4$  armonici Fourier pe setul de validare.

Împărțirea Datelor

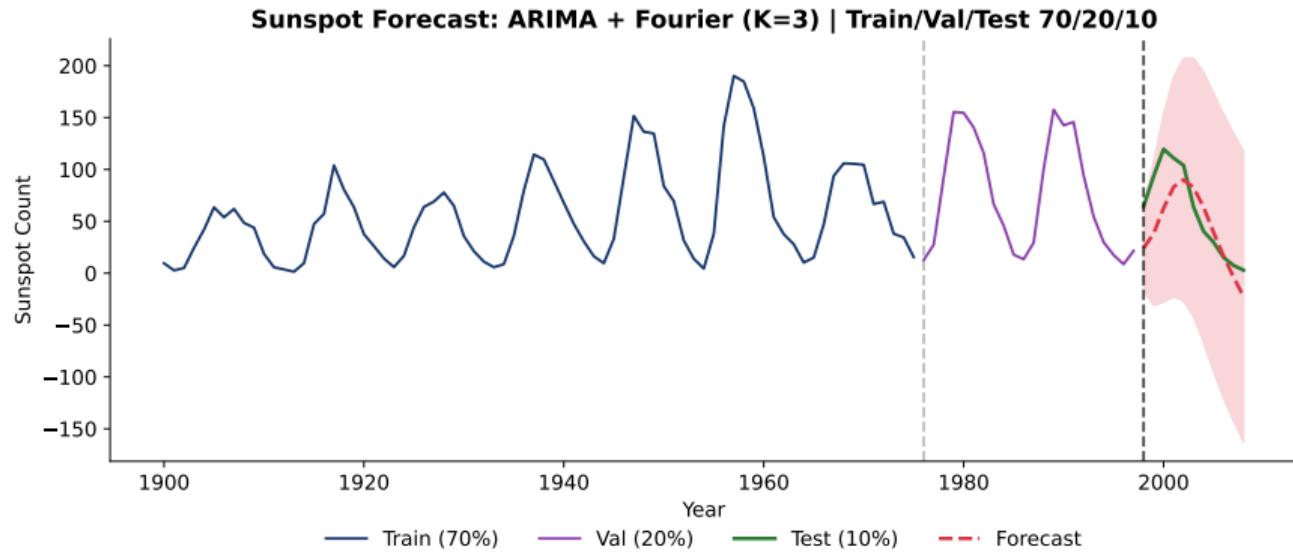
Set	Perioadă	N
Antrenament (70%)	1900–1975	76
Validare (20%)	1976–1997	22
Test (10%)	1998–2008	11
<b>Total</b>	<b>109</b>	

Comparație Modele

K	AIC	Val RMSE	
1	665,9	87,15	
2	668,0	86,92	
3	671,8	<b>86,81</b>	Cel mai bun
4	674,5	87,93	

## Rezultat

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



## 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.

## 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 capturate cicluri multiple

## Fourier vs SARIMA

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

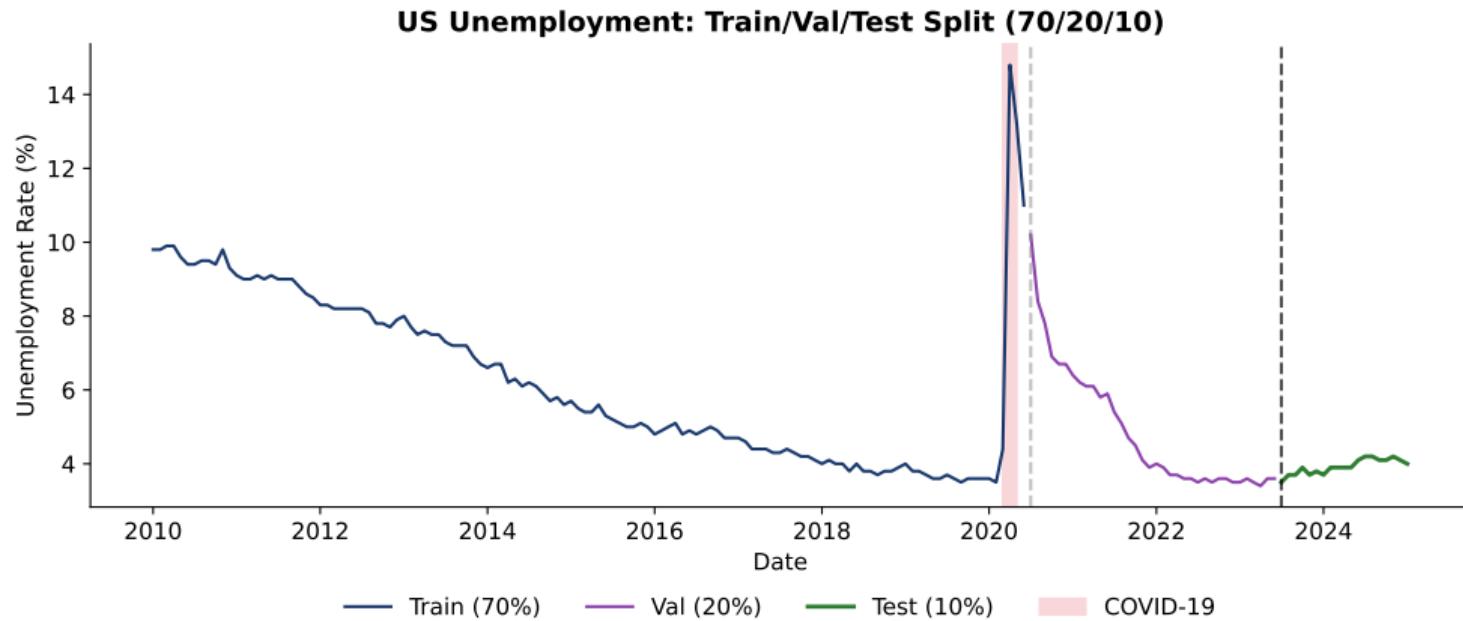
## Alegerea lui K

Începeți cu  $K = 1$ , creșteți până când eroarea de validare nu mai scade. K prea mare = overfitting.

## Aplicații

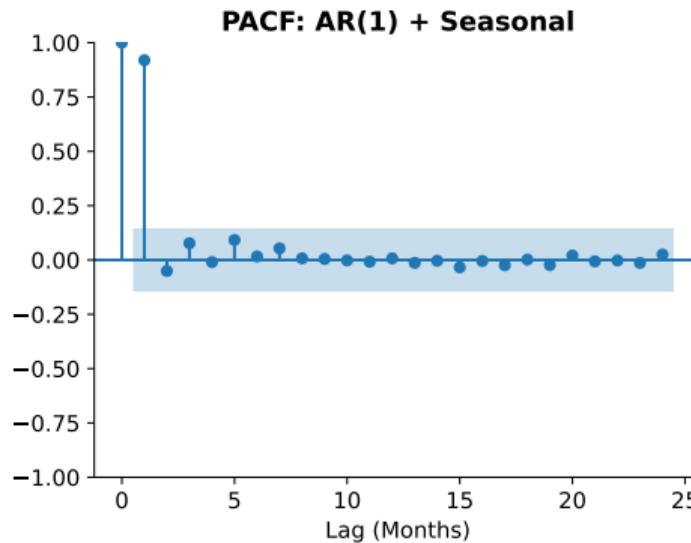
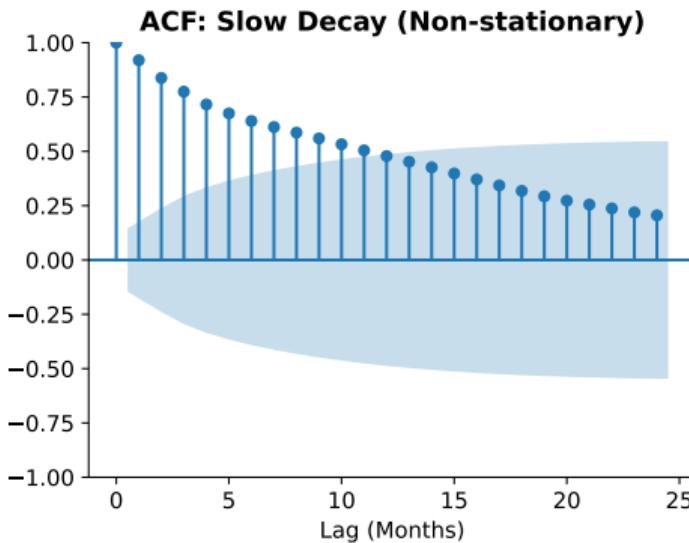
Cicluri climatice, cicluri de afaceri, fenomene astronomice

## Şomajul: Train / Validation / Test Split



### Metodologie

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



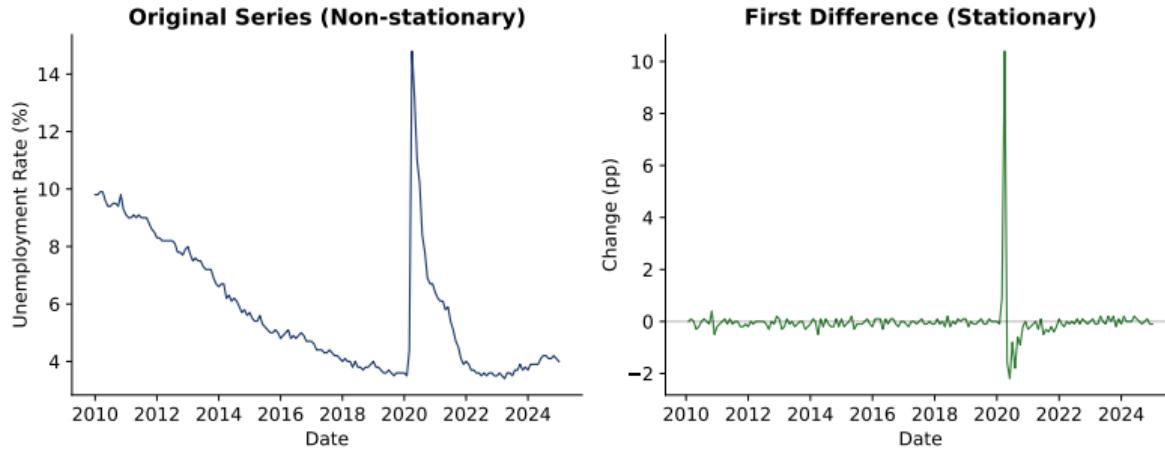
## Interpretare ACF

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

## Interpretare PACF

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

## Şomajul: Teste de Staţionaritate



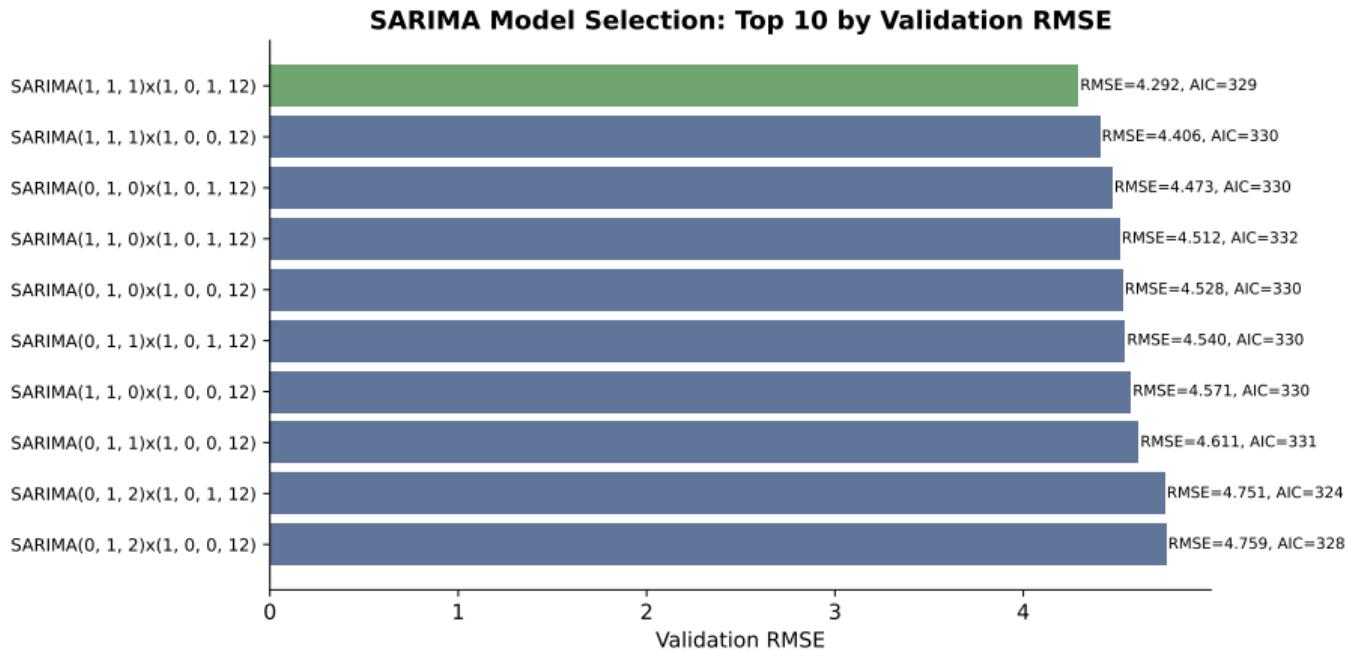
Original: ADF  $p = 0,056$

Nestaţionară (ACF descreştere lentă)

Diferenţiată: ADF  $p < 0,001$

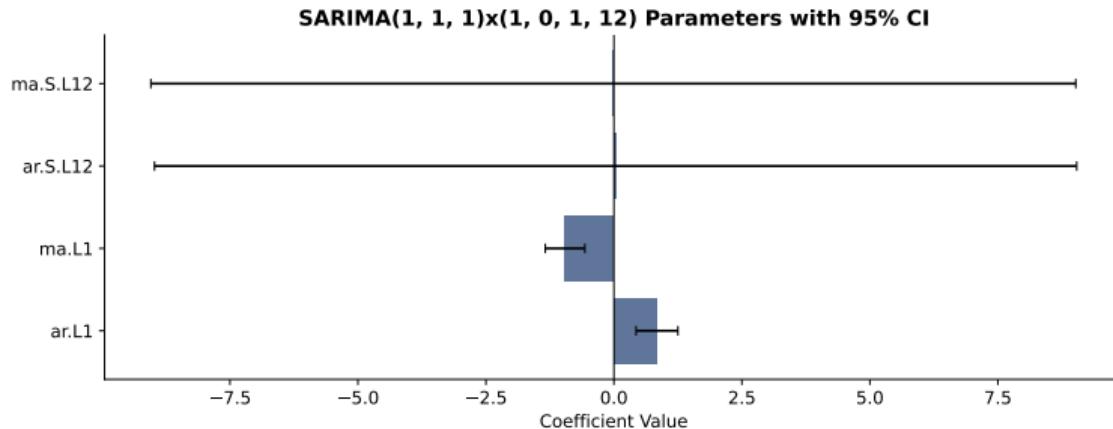
Staţionară  $\Rightarrow$  folosim  $d = 1$

## Şomajul: Selectia Modelului (Set Validare)



Best: SARIMA(1,1,1)(1,0,0)<sub>12</sub>

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

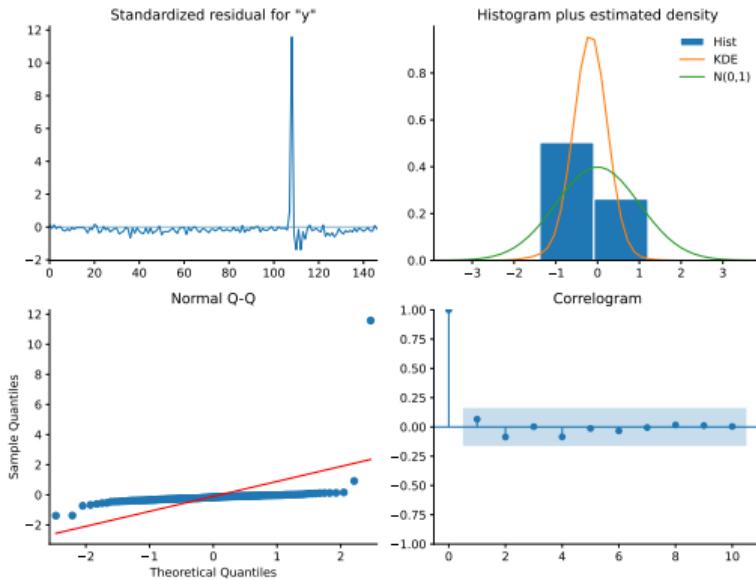


SARIMA(1,1,1)(1,0,0)<sub>12</sub> estimat pe Train+Val (2010-2019)

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

# Şomajul: Diagnosticare SARIMA

SARIMA(1, 1, 1)x(1, 0, 1, 12) Residual Diagnostics



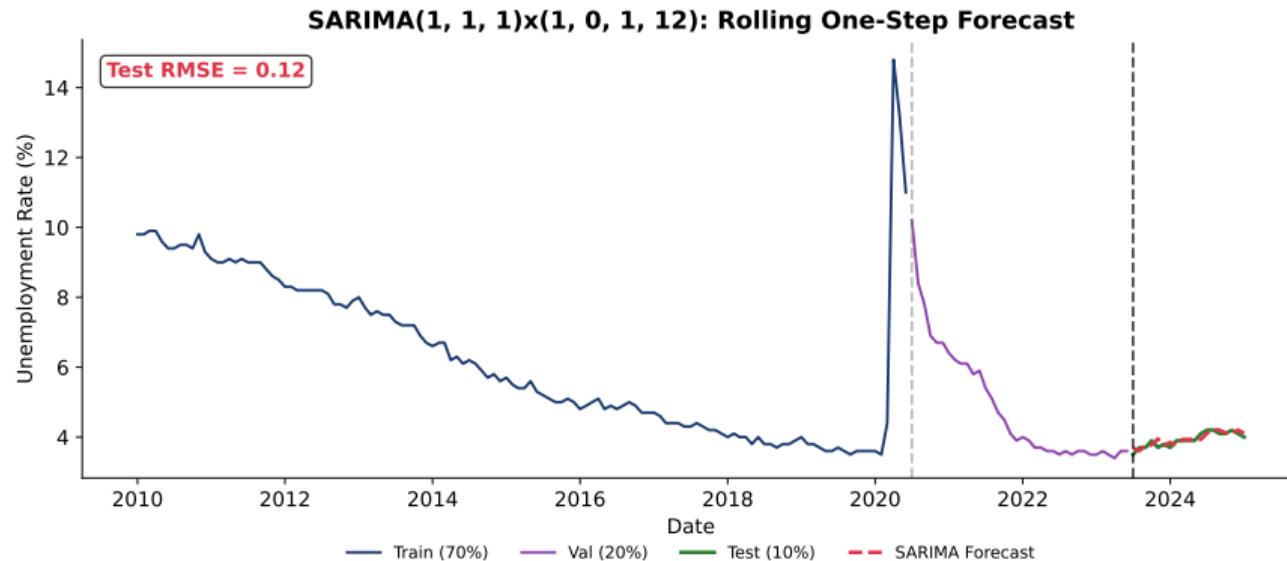
Reziduuri

Rez. std., histogramă, ACF, Q-Q plot.

Ljung-Box  $p = 0,66$

Fără autocorelație. Model bine specificat.

## Şomajul: Prognoza Rolling SARIMA



### Problemă: Ruptura Structurală

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

## Definiție 3 (Descompunerea Prophet)

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2)$$

unde  $g(t)$  = trend,  $s(t)$  = sezonalitate,  $h(t)$  = sărbători,  $\sigma^2$  = varianța zgomotului (estimată).

### Detectare Puncte de Schimbare

- Selectare automată a locațiilor
- `changepoint_prior_scale` controlează flexibilitatea

### Avantaje

- Gestionează date lipsă
- Componente interpretabile
- Robust la outlieri

## 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		<b>181</b>

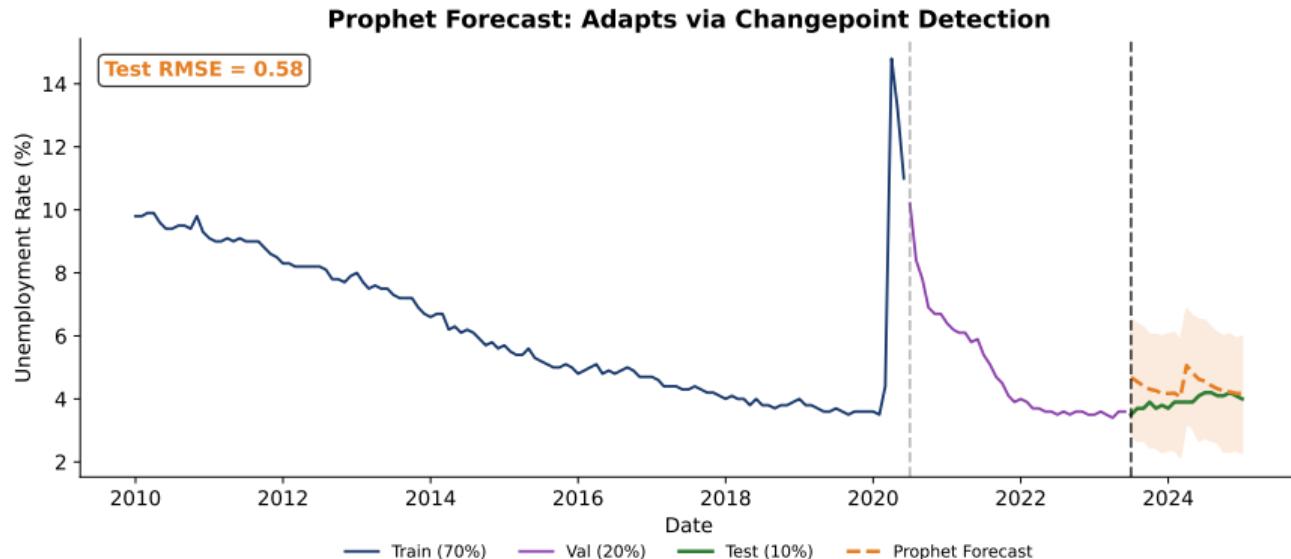
### Comparație Scale

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

## Interpretare

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

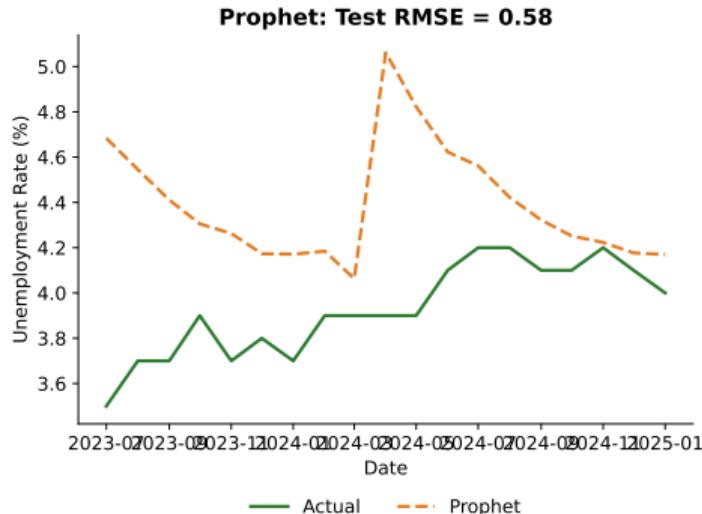
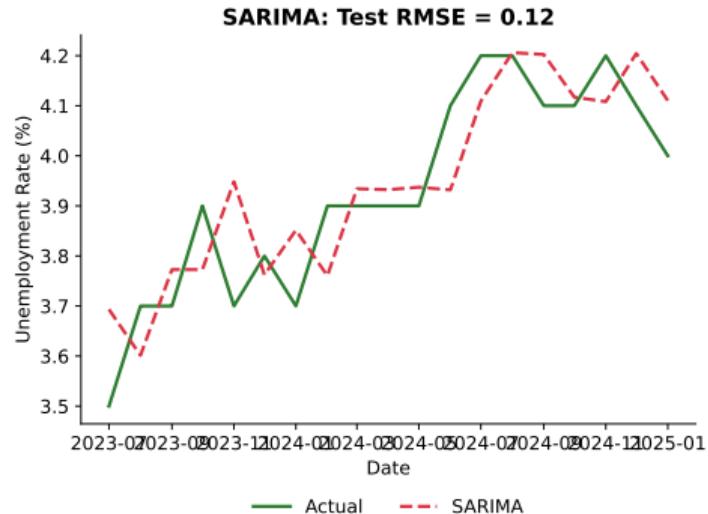
## Şomajul: Rezultate Prognoză Prophet



### Concluzie Cheie

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

## Şomaj: Comparație SARIMA vs Prophet



SARIMA: RMSE = 0,12

Prognoză rolling performează bine.

Prophet: RMSE = 0,58

Eroare mai mare din cauza rupturii structurale.

# Prophet: Când să-l Folosești

## Cazuri de Utilizare Ideale

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

## Prophet vs ARIMA

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

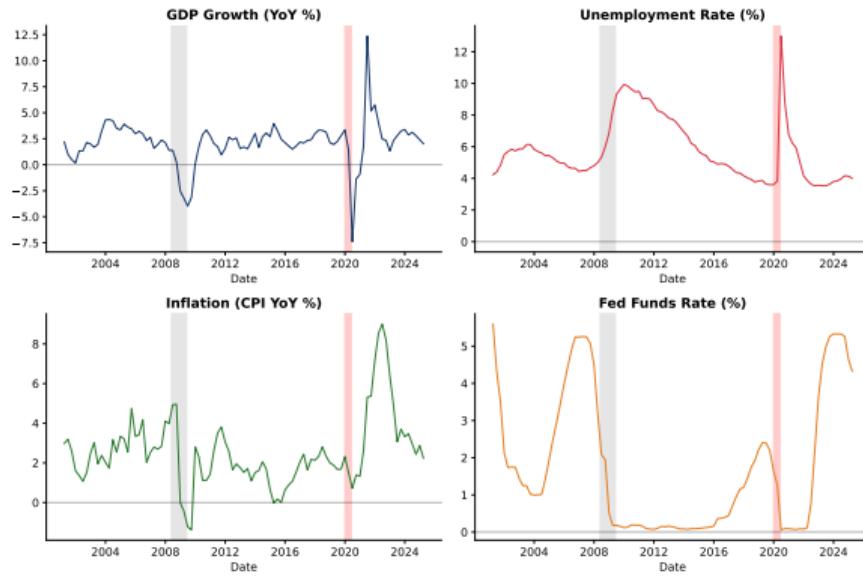
## Nu Este Ideal Pentru

- Date financiare de înaltă frecvență
- Date fără trend/sezonalitate clară
- Serii de timp foarte scurte

## 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.

### Definiție 4 (Autoregresie Vectorială VAR(p))

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

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

unde  $A_i$  sunt matrici de coeficienți  $K \times K$ ,  $u_t \sim N(0, \Sigma)$ ,  $\Sigma$  = matricea de covariantă.

#### Pentru Sistemul Nostru cu 4 Variabile

VAR(2) are:

- 4 intercepte
- $2 \times 4 \times 4 = 32$  coeficienți AR
- **36 parametri total**

#### Selectarea Lag-ului

Folosim criterii informaționale:

- **AIC**: Tinde să supraajusteze
- **BIC**: Mai parsimonios
- Cross-validation pe date păstrate

## VAR: Selectarea Lag-ului și Estimare

### Criterii Informaționale

Lag	BIC
1	-4,810
2	<b>-5,178</b> Cel mai bun
3	-4,633
4	-4,614

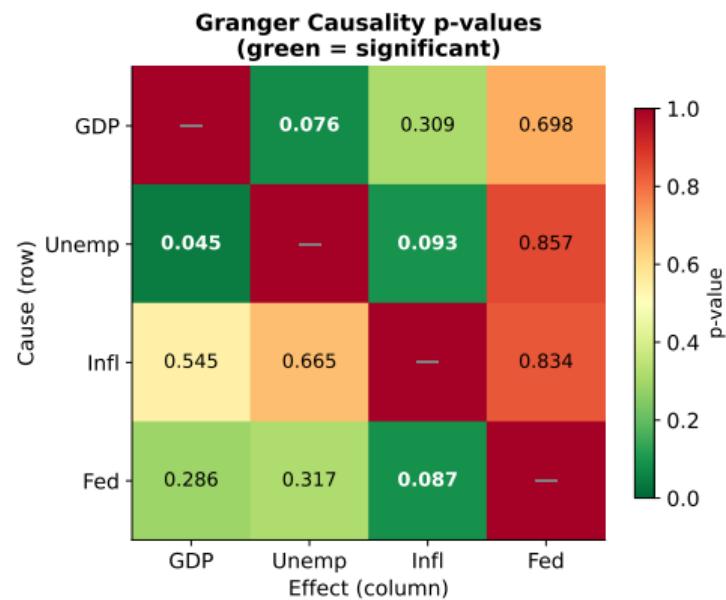
### Împărțirea Datelor

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

### 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.

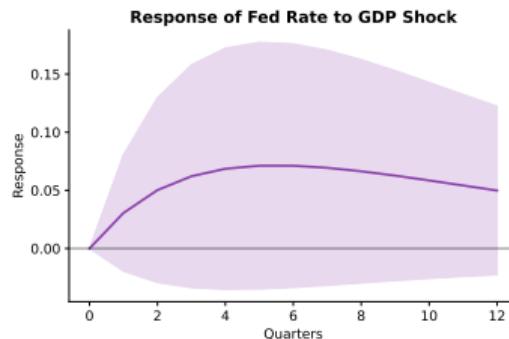
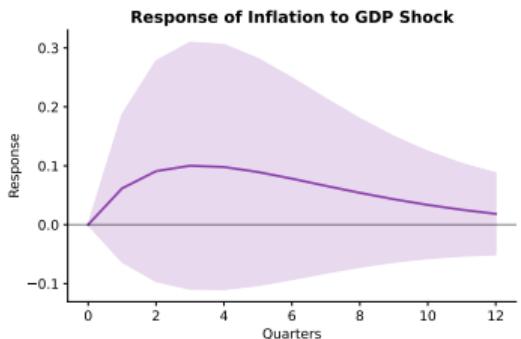
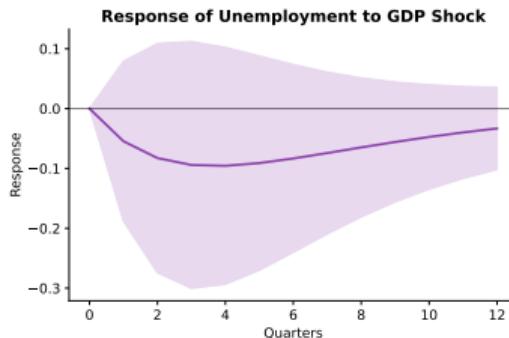
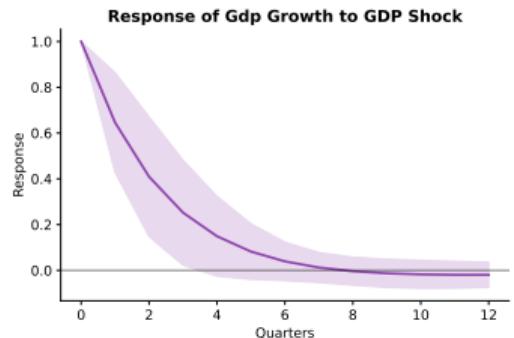
Atenție: “Cauzalitate Granger”  $\neq$  cauzalitate reală!

## Concluzii Economice

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

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

# Funcții de Răspuns la Impuls (IRF)

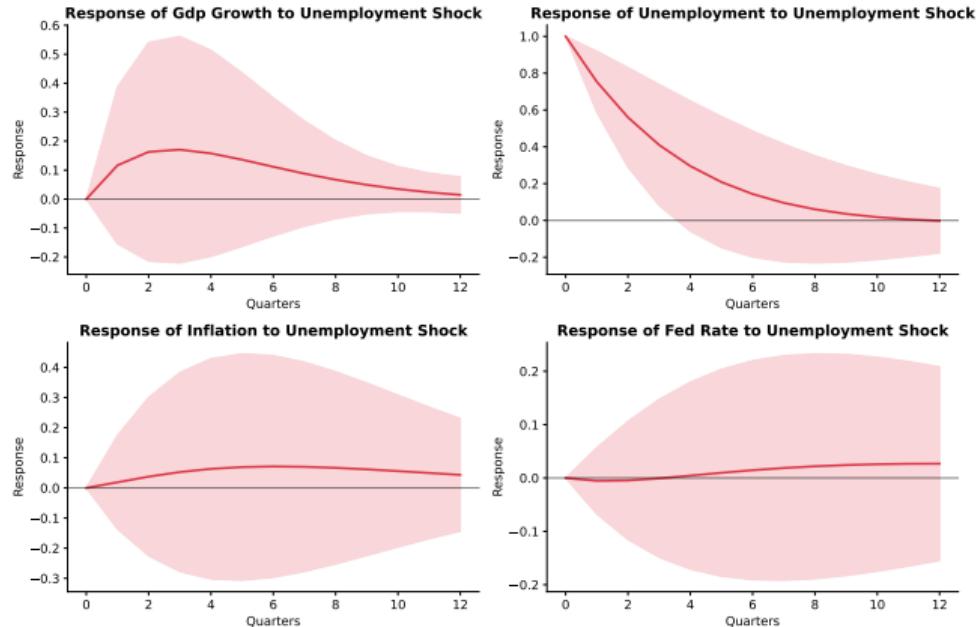


## Ce este IRF?

Arată cum un șoc de 1 unitate la o variabilă afectează celelalte în timp.

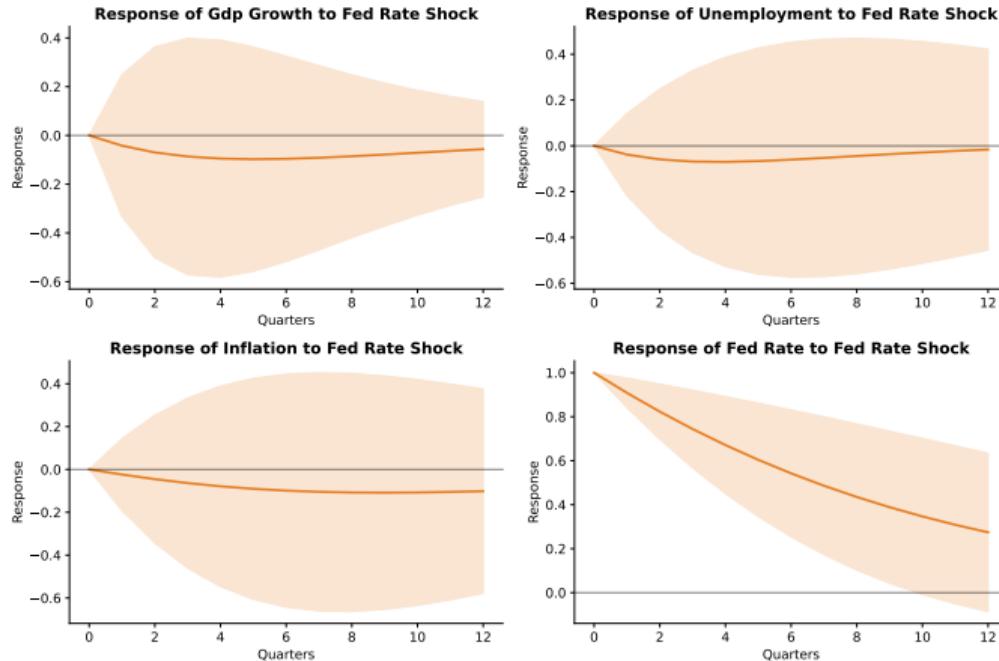
## Efectele Șocului PIB

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



## Efecte

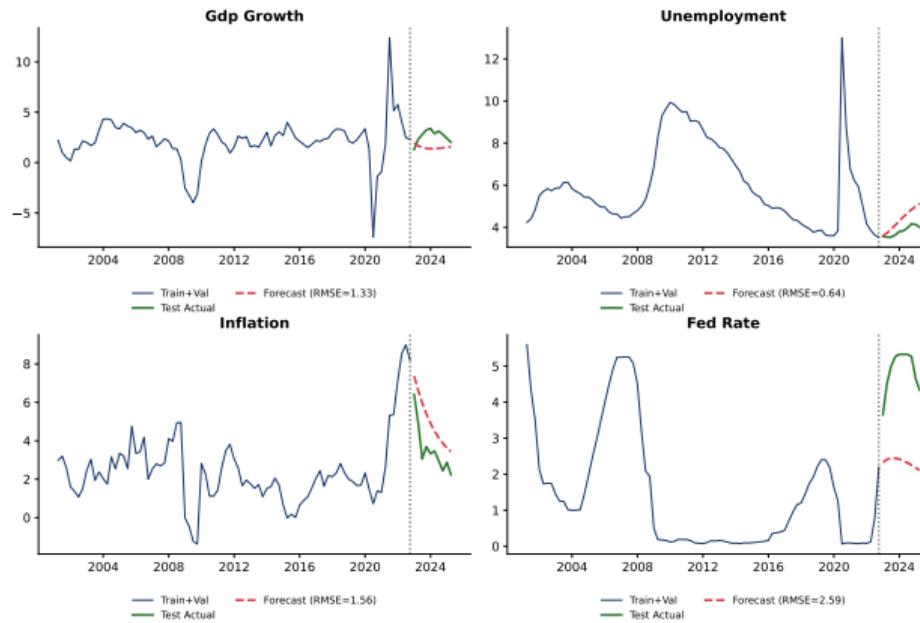
↑ Șomaj  $\Rightarrow$  ↓ PIB (Okun), ↓ Inflație (Phillips), Fed reduce rata.



## Politică Monetară

Creștere rată  $\Rightarrow$  PIB  $\downarrow$ , Șomaj  $\uparrow$ , Inflație  $\downarrow$ .

## VAR: Prognoză (Train/Val/Test)



## Prognoză Rolling One-Step-Ahead

VAR captează dinamica PIB-Şomaj. řoul COVID vizibil în perioada 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%	
<b>Medie</b>	<b>1,53</b>	<b>1,27</b>	<b>60%</b>	

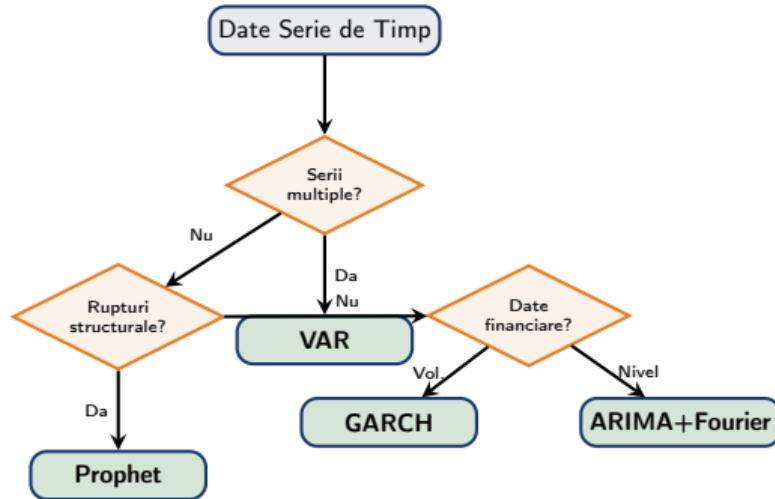
## Puncte Forte

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

## Limitări

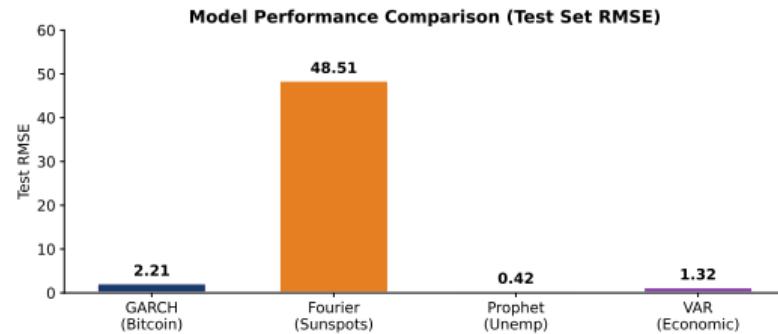
- Mulți parametri (blestemul dimensionalității)
- Sensibil la selectarea lag-ului
- Perioada COVID dificilă

# Cadrul de Selectare a Modelului



## Sumar: Comparație Modele

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



### Principiu Cheie

Potrivii modelul cu caracteristicile datelor. Alegeti în funcție de natura problemei și proprietățile datelor.

# Comparație Cuprinzătoare a 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 (cycli)
- SARIMA: RMSE=0,12 (rupturi)
- VAR: RMSE mediu=1,53 (multi)

## Insight Cheie

Fiecare model excelează în domeniul său. Arta constă în potrivirea modelului cu caracteristicile datelor.

## Metodologie

- ❶ Explorați datele temeinic
- ❷ Testați staționaritatea
- ❸ Împărțiți train/validation/test
- ❹ Comparați modele pe validare
- ❺ Raportați metriki pe test

## 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

## Greșeli Frecvente

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

## Amintiți-vă

“Toate modelele sunt greșite, dar unele sunt utile.”

— George E. P. Box

## 1 Metodologie Riguroasă

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

## 2 Potriviti Modelul cu Datele

- Volatilitate finanțiară → GARCH
- Sezonalitate lungă → Termini Fourier
- Rupturi structurale → Prophet
- Serii multiple → 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

## Referințe

-  Box, G.E.P., Jenkins, G.M., Reinsel, G.C., & Ljung, G.M. (2015). *Time Series Analysis: Forecasting and Control*. Ed. 5, Wiley.
-  Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.
-  Tsay, R.S. (2010). *Analysis of Financial Time Series*. Ed. 3, Wiley.
-  Hyndman, R.J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. Ed. 3, OTexts.
-  Taylor, S.J., & Letham, B. (2018). Forecasting at Scale. *The American Statistician*, 72(1), 37-45.
-  Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3), 307-327.
-  Sims, C.A. (1980). Macroeconomics and Reality. *Econometrica*, 48(1), 1-48.

### Date Reale Folosite în Acest Capitol

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

### Reproductibilitate

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

# Mulțumesc

Întrebări?

Prof. Daniel Traian Pele, PhD

[danpele@ase.ro](mailto:danpele@ase.ro)

Academia de Studii Economice din București