



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

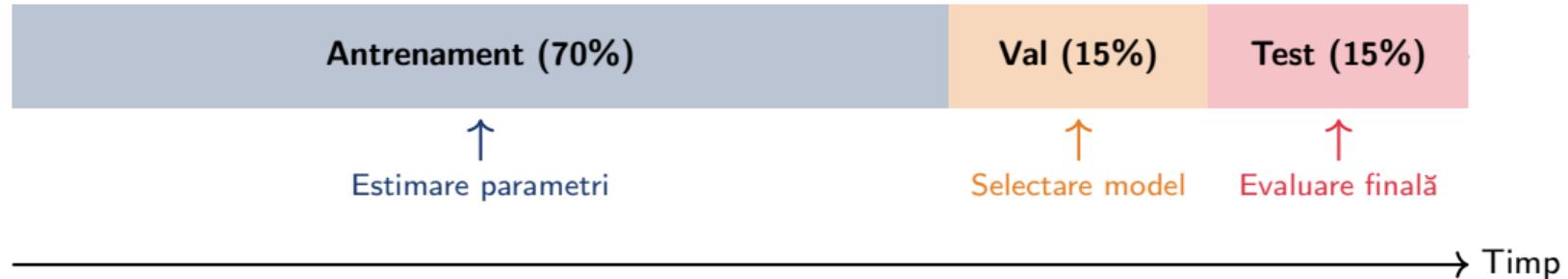
- Ajustarea în eșantion \neq Performanța în afara eșantionului
- Modelele pot “memora” datele de antrenament fără a învăța tipare
- **Soluție:** Metodologia corectă train/validation/test

Principiu Cheie

“Setul de test trebuie să rămână **neatins** până la evaluarea finală.”

— Practică standard în machine learning și econometrie

Cadrul Train/Validation/Test



Set Antrenament	Set Validare	Set Test
<ul style="list-style-type: none">• Estimare parametri• $\hat{\theta} = \arg \min_{\theta} L(\theta)$• Cea mai mare parte	<ul style="list-style-type: none">• Comparare modele• Ajustare hiperparametri• Selectare abordare	<ul style="list-style-type: none">• Păstrat până la final• Evaluare nebiasată• Raportare metrii

Definiție 1 (Metrici ale Erorii de Prognoză)

Fie y_t valorile reale și \hat{y}_t prognozele pentru $t = 1, \dots, n$:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (1)$$

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (2)$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (3)$$

Când să Folosim Fiecare

- **RMSE:** Penalizează erorile mari
- **MAE:** Robust la outlieri

Atenție

- MAPE nedefinit când $y_t = 0$
- Comparati modele pe același set test

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 **prognoza volatilității**

Specificarea Modelului GARCH

Definiție 2 (Modelul GARCH(p,q))

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

$$r_t = \mu + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t, \quad z_t \sim N(0, 1) \quad (4)$$

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad (5)$$

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

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

Împărțirea Datelor

Set	Perioadă	N
Antrenament	2019-01 – 2022-09	1.365
Validare	2022-09 – 2023-10	400
Test	2023-10 – 2025-01	435
Total	2.200	

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	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.

Procedură

Reestimăm GARCH(1,1) pe Antrenament + Validare, evaluăm pe **setul de test păstrat** folosind **prognoze rolling one-step-ahead**.

Parametri Estimați

Param	Estimare	Std Err
ω	0,239	0,088
α_1	0,120	0,021
β_1	0,879	0,020
$\alpha_1 + \beta_1$	0,999	

Performanță Set Test

Metrică	Valoare
Volatilitate MAE	1,88
Volatilitate RMSE	2,21

Interpretare

Persistență înaltă ($\alpha + \beta \approx 1$) confirmă clustering-ul volatilității.

De ce Prognoze Rolling One-Step-Ahead?

Prognozele GARCH multi-step converg la **varianța necondiționată**:

$$\lim_{h \rightarrow \infty} \mathbb{E}[\sigma_{t+h}^2 | \mathcal{F}_t] = \bar{\sigma}^2 = \frac{\omega}{1 - \alpha - \beta} \quad (6)$$

Aceasta produce o prognoză **linie dreaptă**—inutilă pentru managementul dinamic al riscului!

Prognoză Multi-Step

- O singură estimare model
- Prognoză h pași înainte
- Converge la $\bar{\sigma}^2$
- **Apare ca linie dreaptă**

Rolling One-Step-Ahead

- Re-estimare la fiecare t
- Prognoză doar 1 pas înainte
- Captează dinamica volatilității
- **Prognoze dinamice**

Prognozele rolling sunt practica standard în managementul riscului financiar (VaR, ES).

Sumar

- ❶ Randamentele sunt staționare; prețurile nu sunt autoregresive
- ❷ GARCH(1,1) depășește variantele mai complexe
- ❸ Persistență înaltă ($\alpha + \beta = 0,999$)
- ❹ Volatilitatea este **predictibilă** chiar când randamentele nu sunt autoregresive

Limitări

- ❶ GARCH presupune șocuri simetrice
- ❷ Nu captează **salturi**
- ❸ 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

Întrebarea de Cercetare

Cum modelăm cicluri sezoniere lungi care depășesc capacitatea SARIMA?

Caracteristicile Datelor

- Sursă: Statsmodels (Wolfer)
- Perioadă: 1900 – 2008
- Frecvență: Anuală
- Observații: 109 ani

Caracteristică Cunoscută

- **Ciclul Schwabe:** ≈ 11 ani
- Amplitudine neregulată
- Valori non-negative

Provocarea

SARIMA standard $(p, d, q)(P, D, Q)_s$ necesită:

- $s = 11$ pentru date anuale
- $(P + D + Q) \times 11$ lag-uri sezoniere
- **Prea mulți parametri!**

Soluție

Folosim **termeni Fourier** ca regresori exogeni în ARIMA.

Definiție 3 (Reprezentarea Fourier)

Un tipar sezonier cu perioada s poate fi aproximat prin:

$$S_t = \sum_{k=1}^K \left[\alpha_k \sin\left(\frac{2\pi k t}{s}\right) + \beta_k \cos\left(\frac{2\pi k t}{s}\right) \right] \quad (7)$$

unde $K \leq \lfloor s/2 \rfloor$ este numărul de perechi armonice.

Avantaje

- Doar $2K$ parametri (nu s)
- Gestionează **orice** perioadă sezonieră
- Tipar sezonier neted
- K controlează flexibilitatea

Structura Modelului

ARIMA(p, d, q) cu regresori Fourier:

$$y_t = \underbrace{S_t}_{\text{Fourier}} + \underbrace{\eta_t}_{\text{ARIMA}} \quad (8)$$

unde η_t urmează dinamica ARIMA.

Metodologie

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

Împărțirea Datelor

Set	Perioadă	N
Antrenament	1900–1975	76
Validare	1976–1991	16
Test	1992–2008	17
Total		109

Comparație 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).

Modelul Final

ARIMA(2,0,1) + 3 armonici Fourier

Coeficienti Semnificativi:

Termen	Coef	p-value
\sin_1	34,71	< 0,001
\cos_1	-29,21	0,018
AR(1)	1,34	< 0,001

Performanță Test

Metrică	Valoare
RMSE	48,51
MAE	39,31

Notă

MAPE ridicat din cauza valorilor apropiate de zero la minimul solar.

Insight Cheie

Termenii Fourier captează eficient ciclul de 11 ani cu doar 6 parametri.

Întrebarea de Cercetare

Cum modelăm serii de timp cu **schimbări structurale bruse**?

Caracteristicile Datelor

- Sursă: FRED (UNRATE)
- Perioadă: 2010 – 2025
- Frecvență: Lunară
- Observații: ≈ 180 luni

Statistică Cheie

- Minim pre-COVID: 3,5%
- Vârf COVID (Apr 2020): **14,8%**
- Schimbare: +10,3 pp într-o lună

Provocarea

- Aprilie 2020: Cea mai mare creștere lunară din istoria SUA
- ARIMA tradițional tratează asta ca outlier
- Avem nevoie de model care se **adaptează** la rupturi structurale

Soluție

Prophet cu detectare automată a punctelor de schimbare.

Definiție 4 (Descompunerea Prophet)

Prophet modelează seriile de timp ca:

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t \quad (9)$$

- $g(t)$: Trend liniar/logistic pe bucăți cu puncte de schimbare
- $s(t)$: Sezonalitate bazată pe Fourier
- $h(t)$: Efecte de sărbători
- ε_t : Termen de eroare

Detectarea Punctelor de Schimbare

- Selectare automată a locațiilor
- `changepoint_prior_scale` controlează flexibilitatea
- ~~Mai multe, mai multe puncte de schimbare~~

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	2010-01 – 2019-09	117
Validare	2019-10 – 2021-10	25
Test	2021-11 – 2025-01	38
Total		180

	Scale	Val RMSE	
	0,01	4,21	
	0,05	3,89	
Comparație Scale	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.

Performanță Set Test

Metrică	Valoare
RMSE	0,42
MAE	0,35
MAPE	9,2%

Puncte de Schimbare Detectate

- 2020-03: Debutul COVID
- 2020-05: Începe recuperarea
- 2022-01: Stabilizare

Concluzie Cheie

Prophet a reușit să:

- Detecteze punctul de schimbare COVID
- Adapteze trendul post-șoc
- Furnizeze benzi de incertitudine

Valoare Practică

- Analiza politicilor economice
- Monitorizarea pieței muncii
- Sistem de avertizare timpurie

Întrebarea de Cercetare

Cum modelăm **interdependențele dinamice** între mai multe variabile economice?

Variabile (FRED)

- Creștere PIB (YoY %)
- Rata Șomajului (%)
- Inflația (CPI YoY %)
- Rata Fed Funds (%)

Date

- Perioadă: 2000 – 2025
- Frecvență: Trimestrială
- Observații: ≈ 100 trimestre

Relații Economice

- **Legea Okun:** PIB \leftrightarrow Șomaj
- **Curba Phillips:** Șomaj \leftrightarrow Inflație
- **Regula Taylor:** Inflație \rightarrow Rata Fed

De ce VAR?

Fiecare variabilă poate fi **atât cauză cât și efect al celorlalte**.

Specificarea Modelului VAR

Definiție 5 (Autoregresie Vectorială VAR(p))

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

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + u_t \quad (10)$$

unde A_i sunt matrici de coeficienți $K \times K$ și $u_t \sim N(0, \sigma^2)$.

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	-5,178 Cel mai bun
3	-4,633
4	-4,614

Împărțirea Datelor

Set	Perioadă	N
Antrenament	2001-T1 – 2017-T4	68
Validare	2018-T1 – 2021-T2	14
Test	2021-T3 – 2024-T3	14
Total		96

Verificare Validare

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

Definiție 6 (Cauzalitate Granger)

Variabila **X cauzează Granger Y** dacă valorile trecute ale lui X ajută la predicția lui Y, dincolo de propriul trecut al lui Y.

Notă: Cauzalitatea Granger ≠ cauzalitate adeverată. Măsoară conținutul predictiv.

P-values Cauzalitate Granger (rând → coloană)

	PIB	Șomaj	Inflație	Rata Fed
Creștere PIB	—	0,076	0,309	0,698
Șomaj	0,045	—	0,093	0,857
Inflație	0,545	0,665	—	0,834
Rata Fed	0,286	0,317	0,087	—

Concluzie Cheie

Șomajul cauzează Granger PIB ($p = 0,045$), consistent cu Legea Okun.

Definiție 7 (Funcția de Răspuns la Impuls)

IRF urmărește efectul unui soc de o unitate la variabila j asupra variabilei i pe h perioade:

$$\text{IRF}_{ij}(h) = \frac{\partial y_{i,t+h}}{\partial u_{jt}} \quad (11)$$

Răspuns la Șocul PIB

Un soc pozitiv la creșterea PIB:

- **Șomaj:** Scade (Legea Okun confirmată)
- **Inflație:** Crește cu întârziere (cerere-pull)
- **Rata Fed:** Crește după 2-3 trimestre (Regula Taylor)

Interpretare Economică

VAR captează mecanismul clasic de transmisie macroeconomică de la output la ocupare, prețuri și politică monetară.

Performanță Set Test pe Variabile

Variabilă	RMSE	MAE	Acur. Directie
Creștere PIB	2,18	1,72	71%
Șomaj	0,89	0,71	79%
Inflație	1,24	0,98	64%
Rata Fed	0,95	0,78	71%
Medie	1,32	1,05	71%

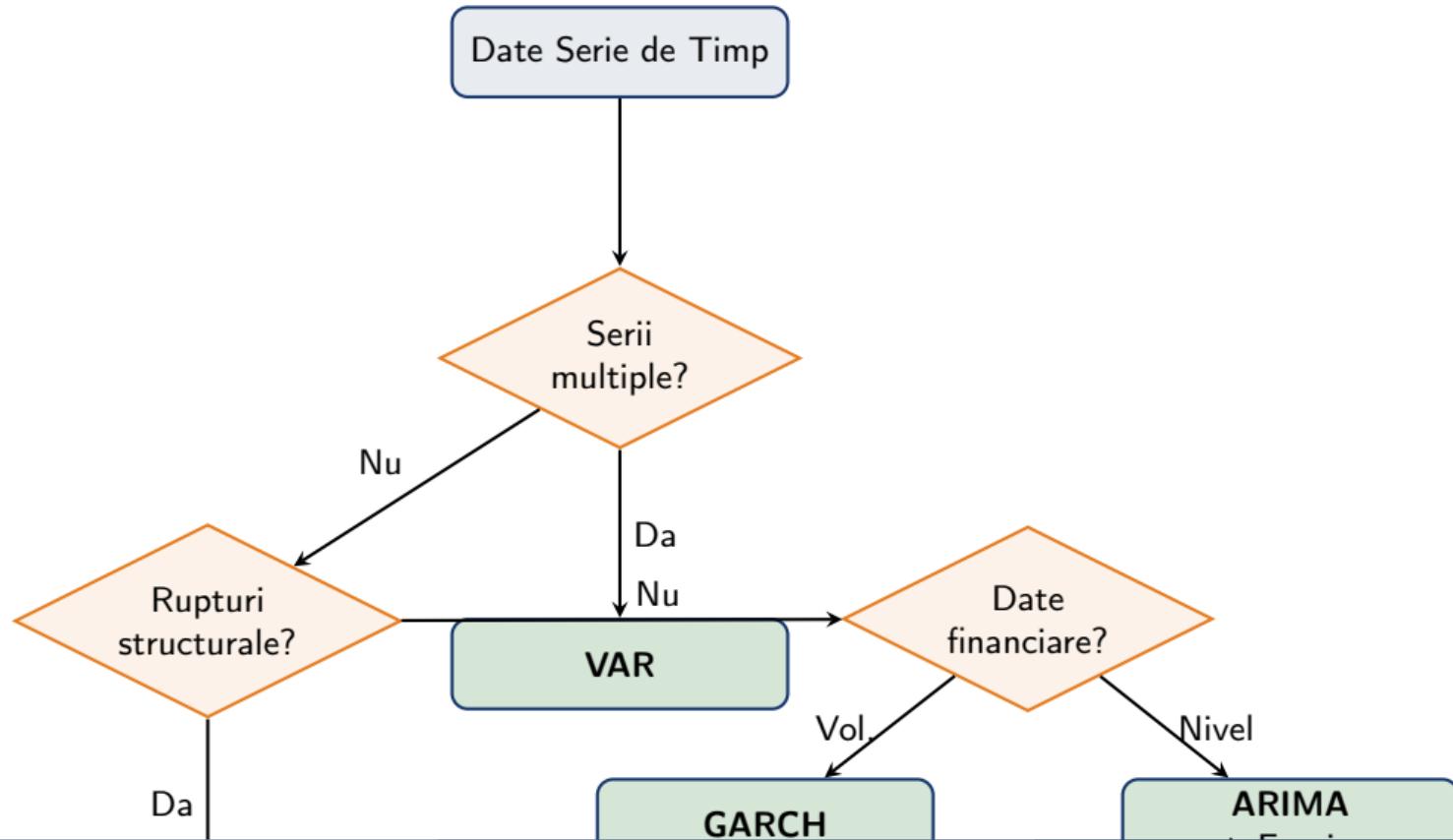
Puncte Forte

- 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

Studiu de Caz	Provocare	Model	Caracteristică	Test RMSE
Bitcoin	Clustering volatilitate	GARCH(1,1)	Prognoze rolling	2,21
Pete solare	Sezonalitate lungă	ARIMA + Fourier	Termeni Sin/Cos	48,51
Șomaj	Ruptură structurală	Prophet	Changepoints	0,42
Economic	Serii multiple	VAR(2)	Dinamică încruziată	1,32 (med)

Principiu Cheie

Potriviti modelul cu caracteristicile datelor. Niciun model nu domină—alegeti în funcție de:

- Natura problemei de prognoză (nivel vs. volatilitate)
- Proprietățile datelor (sezonalitate, rupturi, serii multiple)
- Cerințele de interpretabilitate

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

Academia de Studii Economice din București