Analiza și Prognoza seriilor de timp

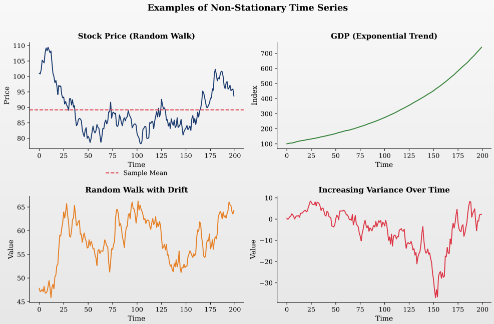
Capitolul 3: Modele ARIMA pentru Date Nestaționare

Daniel Traian PELE

Academia de Studii Economice din București
IDA Institute Digital Assets
Blockchain Research Center
AI4EFin Artificial Intelligence for Energy Finance
Academia Română, Institutul de Prognoză Economică
MSCA Digital Finance

Structura cursului

De ce ARIMA? Datele nestaționare sunt pretutindeni



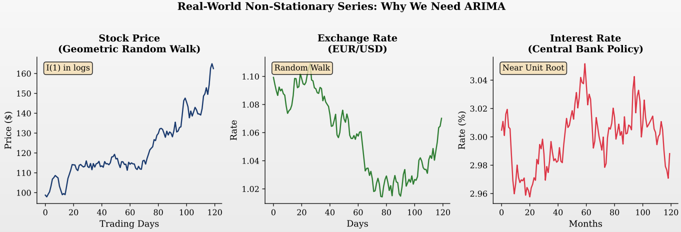
- Prețurile acțiunilor, PIB, cursurile de schimb prezintă **trenduri** sau **comportament rătăcitor**
- Media din eșantion (linia roșie) este lipsită de sens pentru un mers aleatoriu ($\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$)
- Modelele ARMA standard **nu pot** gestiona aceste serii direct

TSA_ch3_motivation_nonstationary

Aplicații practice

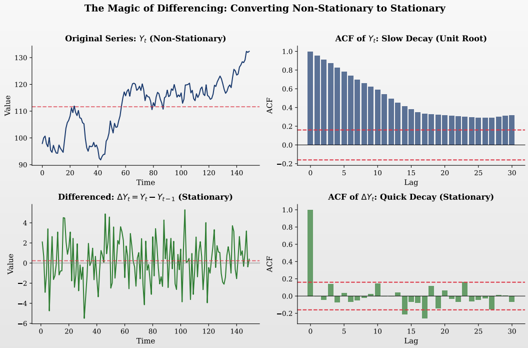
Provocarea

- Prețuri de acțiuni: mers aleatoriu în logaritmi
- Cursuri de schimb: mers aleatoriu
- Rate ale dobânzii: foarte persistente
  - Aproape de rădăcină unitate



TSA_ch3_motivation_realworld

Soluția: diferențierea



Observație Cheie

Diferențierea transformă o serie nestaționară într-una staționară: $\Delta Y_t = Y_t - Y_{t-1}$. ACF se schimbă de la descreștere lentă la descreștere rapidă!

TSA_ch3_motivation_differencing

Ce vom învăța astăzi

Concepte Fundamentale

1. **Nestaționaritatea**: De ce contează și cum o detectăm
2. **Teste de rădăcină unitate**: Testele ADF, PP, KPSS
3. **Diferențierea**: Transformarea cheie
4. **Modele ARIMA**: Combinarea diferențierii cu ARMA
5. **Metodologia Box-Jenkins**: Identificare $\rightarrow$ Estimare $\rightarrow$ Validare

La sfârșitul acestui curs

Veți fi capabili să modelați și să prognozați serii de timp nestaționare precum prețurile acțiunilor, PIB și cursurile de schimb folosind modele ARIMA.

De ce contează nestaționaritatea

Problema

Multe serii de timp economice și financiare sunt **nestaționare**:

- PIB, prețuri de acțiuni, cursuri de schimb, indici de inflație
- Prezintă trenduri, medii în schimbare sau varianță în creștere

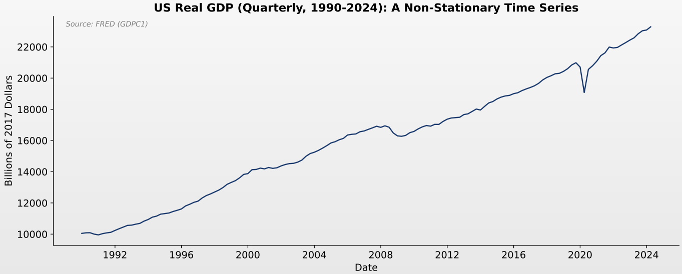
Consecințele Nestaționarității

- Modelele ARMA standard presupun staționaritate
- Regresia OLS cu date nestaționare duce la **regresie falsă**
- Momentele din eșantion (medie, varianță, ACF) nu sunt estimatori consistenți
- Inferența statistică devine invalidă

Exemplu: PIB real SUA

Observații

- **Trend** ascendent clar – media nu este constantă
- Exemplu clasic de serie **nestaționară**
- Nu putem aplica modele ARMA direct



TSA_ch3_gdp_levels

Tipuri de nestaționaritate

Trend determinist

$$Y_t = \alpha + \beta t + \varepsilon_t$$

- Trendul este o funcție deterministă de timp
- Poate fi eliminat prin **regresie**
- Șocurile au efecte temporare

Trend stochastic (Rădăcină Unitate)

$$Y_t = Y_{t-1} + \varepsilon_t$$

- Proces de mers aleatoriu
- Trebuie eliminat prin **diferențiere**
- Șocurile au efecte permanente

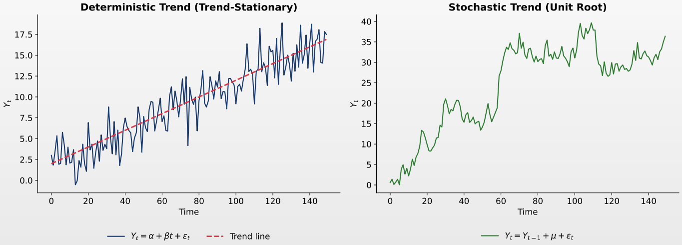
Distincție Cheie

Identificarea corectă este crucială: eliminarea trendului prin regresie pentru un proces cu rădăcină unitară sau diferențierea unui proces staționar în trend duc ambele la specificare greșită!

Vizualizarea diferenței

Observații

- **Stânga**: Trend determinist
  - Abaterile de la trend sunt temporare
- **Dreapta**: Trend stochastic
  - Șocurile se acumulează permanent
- Ambele arată similar, dar necesită tratamente **diferite**!



TSA_ch3_trend_comparison

Procesul de mers aleatoriu

Definiție 1 (Mers aleatoriu)

Un **mers aleatoriu** este definit ca:

$$Y_t = Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2)$$

Cu condiția inițială $Y_0 = 0$, avem: $Y_t = \sum_{i=1}^{t} \varepsilon_i$

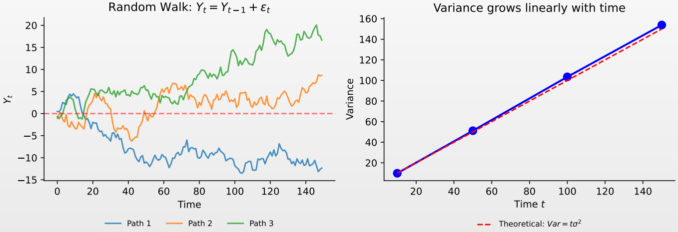
Proprietățile Mersului Aleatoriu

- $\mathbb{E}[Y_t] = 0$ (medie constantă)
- $\text{Var}(Y_t) = t\sigma^2$ (varianța crește în timp!)
- $\text{Cov}(Y_t, Y_{t-k}) = (t-k)\sigma^2$ pentru $k \leq t$
- ACF: $\rho_k = \sqrt{\frac{t-k}{t}} \to 1$ când $t \to \infty$

Mers aleatoriu: ilustrație vizuală

Proprietăți cheie

- **Stânga**: Traiectorii rătăcesc imprevizibil, fără revenire la medie
- **Dreapta**: $\text{Var}(Y_t) = t\sigma^2$ crește liniar
  - Caracteristica definitorie a nestaționarității



TSA_ch3_def_random_walk

Demonstrație: varianța mersului aleatoriu

Afirmație: Pentru $Y_t = Y_{t-1} + \varepsilon_t$ cu $Y_0 = 0$: $\text{Var}(Y_t) = t\sigma^2$

Demonstrație: Prin substituție recursivă:

$$Y_t = Y_{t-1} + \varepsilon_t = Y_{t-2} + \varepsilon_{t-1} + \varepsilon_t = \cdots = \sum_{i=1}^{t} \varepsilon_i$$

Luând varianța:

$$\text{Var}(Y_t) = \text{Var}\left(\sum_{i=1}^{t} \varepsilon_i\right) = \sum_{i=1}^{t} \text{Var}(\varepsilon_i) + 2\sum_{i<j} \text{Cov}(\varepsilon_i, \varepsilon_j)$$

Deoarece ε_t sunt independente (zgomot alb), toate covarianțele sunt zero: $\text{Var}(Y_t) = \sum_{i=1}^{t} \sigma^2 = \boxed{t\sigma^2}$

Nestaționaritate

Varianța depinde de $t \Rightarrow$ încalcă cerința staționarității ($\text{Var}(Y_t) = \gamma(0)$ constant).

Demonstrație: autocovarianța mersului aleatoriu

Afirmație: $\text{Cov}(Y_t, Y_{t-k}) = (t-k)\sigma^2$ pentru $k \leq t$

Demonstrație: Folosind $Y_t = \sum_{i=1}^{t} \varepsilon_i$ și $Y_{t-k} = \sum_{i=1}^{t-k} \varepsilon_i$:

$$\text{Cov}(Y_t, Y_{t-k}) = \text{Cov}\left(\sum_{i=1}^{t} \varepsilon_i, \sum_{j=1}^{t-k} \varepsilon_j\right)$$

$$= \sum_{i=1}^{t} \sum_{j=1}^{t-k} \text{Cov}(\varepsilon_i, \varepsilon_j) = \sum_{i=1}^{t-k} \text{Var}(\varepsilon_i) = \boxed{(t-k)\sigma^2}$$

Doar termenii cu $i = j$ supraviețuiesc (când $i \leq t-k$).

ACF:

$$\rho(k) = \frac{\text{Cov}(Y_t, Y_{t-k})}{\sqrt{\text{Var}(Y_t)\text{Var}(Y_{t-k})}} = \frac{(t-k)\sigma^2}{\sqrt{t\sigma^2 \cdot (t-k)\sigma^2}} = \sqrt{\frac{t-k}{t}}$$

Mers aleatoriu cu drift

Definiție 2 (Mers aleatoriu cu drift)

Un mers aleatoriu cu drift include un termen constant:

$$Y_t = \mu + Y_{t-1} + \varepsilon_t$$

Echivalent: $Y_t = Y_0 + \mu t + \sum_{i=1}^{t} \varepsilon_i$

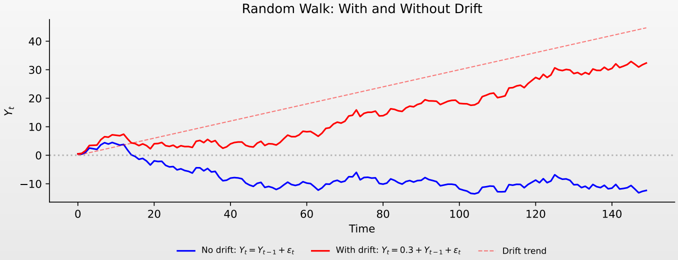
Proprietăți

- $\mathbb{E}[Y_t] = Y_0 + \mu t$ (media crește liniar)
- $\mathrm{Var}(Y_t) = t\sigma^2$ (varianța tot crește)
- Drift-ul μ creează un trend ascendent sau descendent
- Tot nestaționar în ciuda faptului că are un "trend"

Mers aleatoriu cu drift: ilustrație vizuală

Comparație

- **Fără drift** (albastru): rătăcește în jurul lui zero
- **Cu drift** $\mu > 0$ (roșu): trend ascendent sistematic
- Ambele sunt nestaționare
  - Drift-ul adaugă trend determinist la rătăcirea stocastică

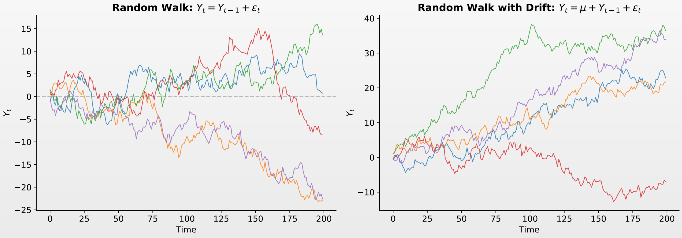


TSA_ch3_def_random_walk_drift

Simularea mersurilor aleatorii

Observații

- **Stânga**: Mersuri aleatorii pure – fără drift, rătăcesc imprevizibil
- **Dreapta**: Cu drift – trend ascendent în medie
- Fiecare traiectorie este unică
  - Incertitudinea crește în timp

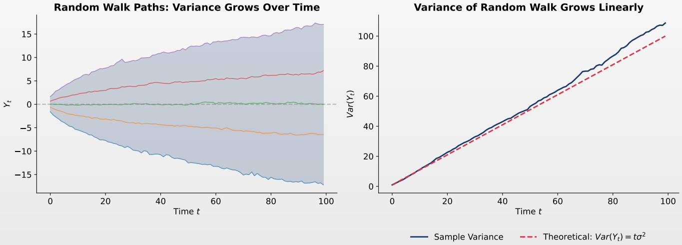


TSA_ch3_random_walk

Creșterea varianței: de ce mersurile aleatorii sunt nestaționare

Observații

- **Stânga**: Evantaiul de traiectorii arată incertitudinea crescând
- **Dreapta**: Varianța crește liniar: $\text{Var}(Y_t) = t\sigma^2$
- Aceasta violează staționaritatea
  - Varianța ar trebui să fie constantă



TSA_ch3_variance_growth

Procese integrate

Definiție 3 (Proces Integrat de Ordin d)

O serie de timp $\{Y_t\}$ este **integrată de ordin** d, scrisă $Y_t \sim I(d)$, dacă:

- Y_t este nestaționară
- $(1-L)^d Y_t = \Delta^d Y_t$ este staționară
- $(1-L)^{d-1} Y_t$ este încă nestaționară

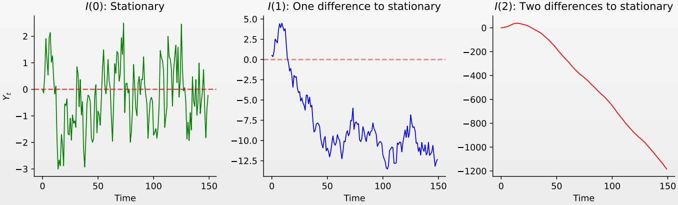
Cazuri Comune

- $I(0)$: Proces staționar (de ex., ARMA)
- $I(1)$: Prima diferență este staționară (cel mai frecvent pentru date economice)
- $I(2)$: A doua diferență este staționară (mai rar)

Proces integrat: ilustrație vizuală

Ordinul de integrare

- $I(0)$: Staționar — nicio diferențiere necesară
- $I(1)$: O diferență necesară (mers aleatoriu)
- $I(2)$: Două diferențe necesare
- Majoritatea seriilor economice sunt $I(0)$ sau $I(1)$



TSA_ch3_def_integrated

Operatorul diferență

Definiție 4 (Prima Diferență)

Operatorul primei diferențe Δ este definit ca: $\Delta Y_t = Y_t - Y_{t-1} = (1-L)Y_t$, unde L este operatorul lag ($LY_t = Y_{t-1}$).

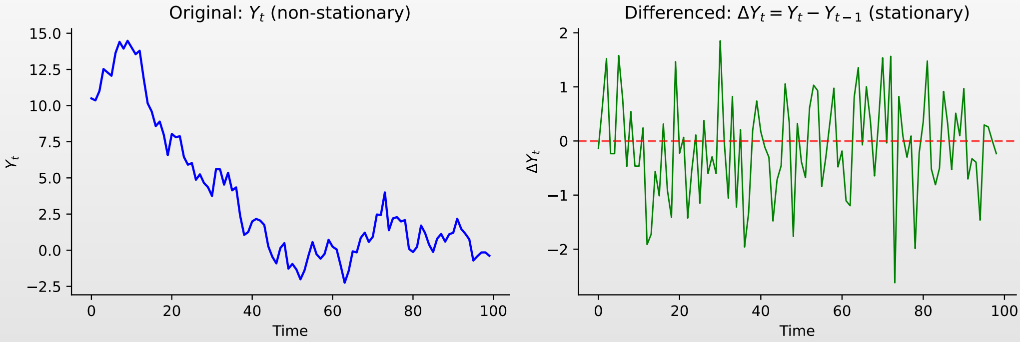
Diferențe de Ordin Superior

- A două diferență: $\Delta^2 Y_t = \Delta(\Delta Y_t) = (1-L)^2 Y_t$
- $\Delta^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$
- Diferența de ordin d: $\Delta^d Y_t = (1-L)^d Y_t$

Rezultat Cheie

Dacă $Y_t \sim I(d)$, atunci $\Delta^d Y_t \sim I(0)$ (staționar).

Prima diferență: ilustrație vizuală



Stânga: serie nestaționară. Dreapta: după prima diferență, seria devine staționară.

TSA_ch3_def_difference

Exemplu: diferențierea unui mers aleatoriu

Mers aleatoriu la Zgomot Alb

Fie $Y_t = Y_{t-1} + \varepsilon_t$ (mers aleatoriu). Luând prima diferență:

$$\Delta Y_t = Y_t - Y_{t-1} = \varepsilon_t$$

Prima diferență este zgomot alb – un proces staționar!

Interpretare

- Un mers aleatoriu este $I(1)$
- O diferență îl transformă în $I(0)$
- "Schimbările" într-un mers aleatoriu sunt staționare

Demonstrație: diferențierea induce staționaritatea

Afirmație: Dacă $Y_t \sim I(1)$, atunci $\Delta Y_t = Y_t - Y_{t-1}$ este staționar.

Demonstrație pentru Mers aleatoriu cu drift: $Y_t = \mu + Y_{t-1} + \varepsilon_t$
Prima diferență este:

$$\Delta Y_t = Y_t - Y_{t-1} = \mu + \varepsilon_t$$

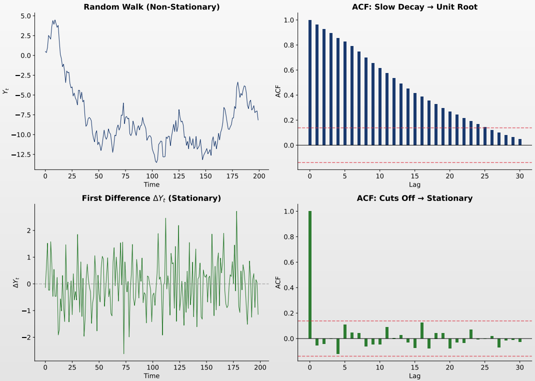
Verificăm condițiile de staționaritate:

1. **Media:** $\mathbb{E}[\Delta Y_t] = \mu$ (constantă, nu depinde de t) ✓
2. **Varianța:** $\text{Var}(\Delta Y_t) = \text{Var}(\varepsilon_t) = \sigma^2$ (constantă) ✓
3. **Autocovarianța:** $\text{Cov}(\Delta Y_t, \Delta Y_{t-k}) = \text{Cov}(\varepsilon_t, \varepsilon_{t-k}) = 0$ pentru $k \neq 0$ ✓

Principiu General

Diferențierea elimină "memoria" care face varianța să se acumuleze. Pentru procese $I(d)$, sunt necesare d diferențe.

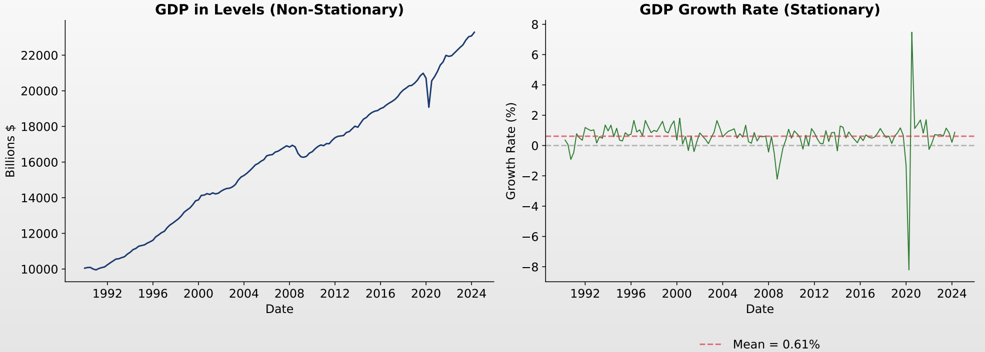
ACF: detectarea nestaționarității



- **Sus**: ACF mers aleatoriu scade foarte lent $\Rightarrow$ rădăcină unitate
- **Jos**: După diferențiere, ACF se întrerupe $\Rightarrow$ staționar

TSA_ch3_acf_nonstationary

Diferențierea în practică: exemplul PIB



- **Stânga**: PIB în niveluri – trend ascendent clar (nestaționar)
- **Dreapta**: Rata de creștere PIB (diferența logaritmică) – fluctuează în jurul mediei (staționar)
- Diferențierea elimină trendul și obține staționaritate

TSA_ch3_differencing

Supra-diferențierea

Avertisment: Supra-diferențierea

Diferențierea mai mult decât este necesar introduce probleme:

- Creează autocorelație negativă artificială
- Inflează varianța
- Pierde informație

Exemplu

Dacă $Y_t \sim I(1)$, atunci $\Delta Y_t \sim I(0)$. Dar dacă diferențiem din nou:

$$\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1} = \varepsilon_t - \varepsilon_{t-1}$$

Acesta este un MA(1) cu $\theta_1 = -1$ (la granița non-invertibilității)!

Definiția ARIMA

Definiție 5 (ARIMA(p,d,q))

O serie de timp $\{Y_t\}$ urmează un proces **ARIMA(p,d,q)** dacă:

$$\phi(L)(1-L)^d Y_t = c + \theta(L)\varepsilon_t$$

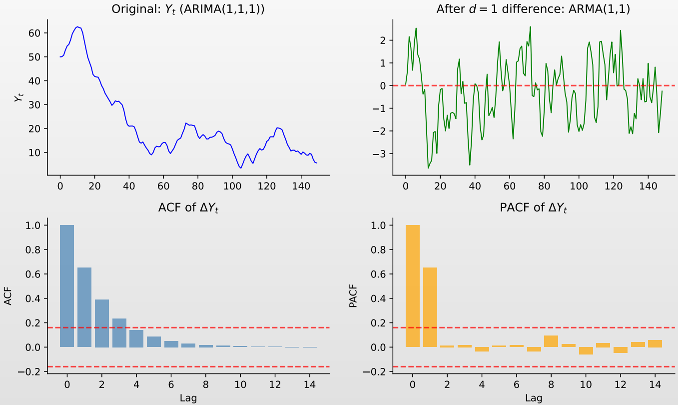
unde:

- $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ (polinomul AR)
- $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$ (polinomul MA)
- d este ordinul de integrare (numărul de diferențe)
- $\varepsilon_t \sim WN(0, \sigma^2)$

ARIMA: ilustrație vizuală

Interpretare

- **Sus**: seria ARIMA originală (nestaționară)
- **Jos**: după diferențiere de d ori
  - ACF/PACF dezvăluie ordinele AR și MA



TSA_ch3_def_arima

Componentele ARIMA

AR(p)	I(d)	MA(q)
Autoregresiv Memorie	Integrare Diferențiere	Medie Mobilă Șocuri

Cazuri Speciale

- ARIMA(p,0,q) = ARMA(p,q) – staționar
- ARIMA(0,1,0) = Mers aleatoriu
- ARIMA(0,1,1) = IMA(1,1) – netezire exponențială
- ARIMA(1,1,0) = ARI(1,1) – AR(1) diferențiat

Exemplu ARIMA(1,1,0)

Model ARI(1,1)

$$\Delta Y_t = c + \phi_1 \Delta Y_{t-1} + \varepsilon_t$$

Echivalent: $(1 - \phi_1 L)(1 - L) Y_t = c + \varepsilon_t$

Interpretare

- **Schimbările** în Y_t urmează un proces AR(1)
- Dacă $|\phi_1| < 1$, schimbările sunt staționare
- Y_t în sine are un trend stochastic
- Model comun pentru multe serii de timp economice

Exemplu ARIMA(0,1,1)

Model IMA(1,1)

$$\Delta Y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Echivalent: $(1 - L)Y_t = c + (1 + \theta_1 L)\varepsilon_t$

Conexiunea cu Netezirea exponențială

Modelul IMA(1,1) este echivalent cu **netezirea exponențială simplă**:

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha)\hat{Y}_t$$

unde $\alpha = 1 + \theta_1$ (pentru $-1 < \theta_1 < 0$).

Rolul constantei în ARIMA

Termenul Constant în ARIMA(p,d,q)

Când $d > 0$, constanta c are o interpretare diferită: $\phi(L)(1-L)^d Y_t = c + \theta(L)\varepsilon_t$

Implicații Importante

- Pentru $d = 1$: c reprezintă **drift-ul** (schimbarea medie): $\mathbb{E}[\Delta Y_t] = \frac{c}{1-\phi_1-\cdots-\phi_p}$
- Pentru $d = 2$: c afectează **curbura** trendului
- Adesea se presupune $c = 0$ când $d \geq 1$

Testarea pentru rădăcini unitate

De ce Testăm?

Înainte de a potrivi un model ARIMA, trebuie să determinăm:

1. Este seria staționară? (Este $d = 0$?)
2. Dacă nu, câte diferențe sunt necesare? (Care este d?)

Teste comune de rădăcină unitate

- **Dickey-Fuller (DF)** și **Augmented Dickey-Fuller (ADF)**
- **Phillips-Perron (PP)**
- **KPSS** (test de staționaritate – ipoteză nulă inversată)

Testul Dickey-Fuller

Configurare

Considerăm modelul AR(1): $Y_t = \phi Y_{t-1} + \varepsilon_t$. Scădem Y_{t-1}: $\Delta Y_t = (\phi - 1)Y_{t-1} + \varepsilon_t = \gamma Y_{t-1} + \varepsilon_t$, unde $\gamma = \phi - 1$.

Ipoteze

- H_0: $\gamma = 0$ (rădăcină unitate, $\phi = 1$, nestaționar)
- H_1: $\gamma < 0$ (staționar, $|\phi| < 1$)

Problemă Cheie

Sub H_0, statistică t **nu** urmează o distribuție t standard! Trebuie folosite valorile critice Dickey-Fuller.

Variante ale testului Dickey-Fuller

Trei Specificări

1. **Fără constantă, fără trend**: $\Delta Y_t = \gamma Y_{t-1} + \varepsilon_t$
2. **Cu constantă (drift)**: $\Delta Y_t = \alpha + \gamma Y_{t-1} + \varepsilon_t$
3. **Cu constantă și trend**: $\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \varepsilon_t$

Alegerea Specificării Corecte

- Examinați datele: au un trend vizibil?
- Includerea termenilor inutili reduce puterea
- Excluderea termenilor necesari duce la inferență incorectă

Testul Augmented Dickey-Fuller (ADF)

Problema cu DF Simplu

Dacă există dinamică AR dincolo de AR(1), reziduurile DF vor fi autocorelate.

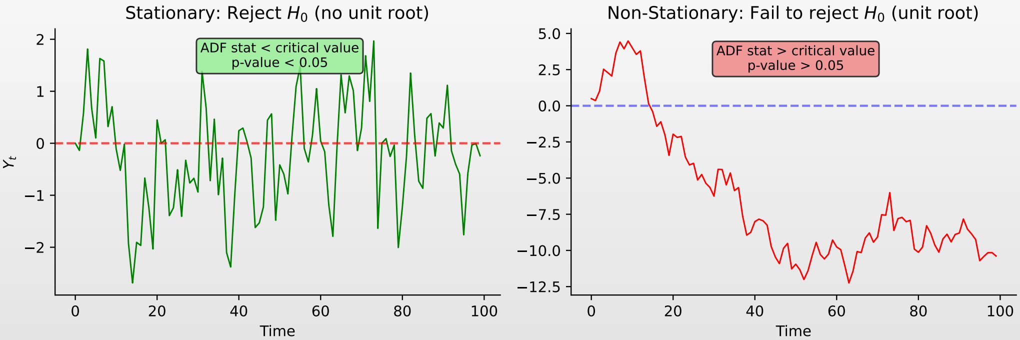
Definiție 6 (Testul ADF)

Adăugați diferențe întârziate: $\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \sum_{j=1}^{k} \delta_j \Delta Y_{t-j} + \varepsilon_t$
Testați $H_0 : \gamma = 0$ folosind valorile critice ADF.

Alegerea Lungimii Lag-ului k

- Folosiți criterii informaționale (AIC, BIC)
- Începeți cu k_{max}, reduceți până ultimul lag este semnificativ

Testul ADF: ilustrație vizuală



Stânga: serie staționară – ADF respinge rădăcina unitate. Dreapta: nestaționară – ADF nu respinge.

TSA_ch3_def_adf

Valori critice ADF

Model	**1%**	**5%**	**10%**
Fără constantă, fără trend	-2.58	-1.95	-1.62
Cu constantă	-3.43	-2.86	-2.57
Cu constantă și trend	-3.96	-3.41	-3.13

Regula de Decizie

- Statistică de test $<$ valoare critică $\Rightarrow$ Respingem H_0 (staționar)
- Statistică de test $\geq$ valoare critică $\Rightarrow$ Nu respingem (rădăcină unitate)

Testul Phillips-Perron (PP)

Motivație

Ca și ADF, testează H_0: Rădăcină unitate vs H_1: Staționar, dar folosește o **corecție non-parametrică** pentru corelația serială în loc de adăugarea diferențelor întârziate.

Statistică de Test

Testul PP modifică statistică t DF:

$$Z_t = t_{\hat{\gamma}} \cdot \sqrt{\frac{\hat{\sigma}^2}{\hat{\lambda}^2}} - \frac{T(\hat{\lambda}^2 - \hat{\sigma}^2)(se(\hat{\gamma}))}{2\hat{\lambda}^2 \cdot s}$$

unde $\hat{\lambda}^2$ este o estimare consistentă a varianței pe termen lung folosind Newey-West.

Avantaje față de ADF

- Robust la heteroscedasticitate și corelație serială
- Nu necesită selectarea lungimii lag-ului (folosește lățime de bandă)

Testul KPSS

Ipoteze Inversate

Spre deosebire de ADF: H_0: Staționar vs H_1: Rădăcină unitate

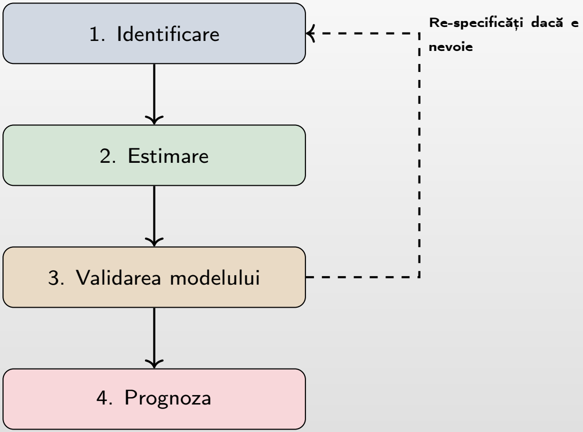
Procedura KPSS

Descompunem: $Y_t = \xi t + r_t + \varepsilon_t$ unde $r_t = r_{t-1} + u_t$. Testăm dacă $\text{Var}(u_t) = 0$.

Utilizare Complementară cu ADF

- ADF respinge, KPSS nu respinge $\Rightarrow$ Staționar
- ADF nu respinge, KPSS respinge $\Rightarrow$ Rădăcină unitate
- Ambele resping sau niciunul $\Rightarrow$ Neconcludent

Metodologia Box-Jenkins



Pasul 1: Determinarea lui d

Procedură

1. Reprezentați grafic seria de timp – căutați trenduri, varianță în schimbare
2. Examinați ACF – descreștere lentă sugerează nestaționaritate
3. Aplicați teste de rădăcină unitate (ADF, KPSS)
4. Dacă nestaționară, diferențiați și repetăți

Ghiduri Practice

- Majoritatea seriilor economice: $d = 1$ este suficient
- Rar avem nevoie de $d > 2$
- Dacă ACF al ΔY_t tot scade lent, încercați $d = 2$
- Atenție la supra-diferențiere (ACF cu $\rho_1 \approx -0.5$)

Pasul 2: Determinarea lui p și q

După Diferențiere

Odată ce $W_t = \Delta^d Y_t$ este staționar, folosiți ACF/PACF pentru a identifica ARMA(p,q):

Model	**ACF**	**PACF**
AR(p)	Scade exponențial	Se întrerupe după lag p
MA(q)	Se întrerupe după lag q	Scade exponențial
ARMA(p,q)	Scade	Scade

Criterii informaționale

Când tiparele sunt neclare, comparați modelele folosind:

- $\text{AIC} = -2\ln(L) + 2k$; $\text{BIC} = -2\ln(L) + k\ln(n)$

Mai mic este mai bun. BIC penalizează complexitatea mai mult.

Algoritmi Auto-ARIMA

Selecție automată a modelului

Software-ul modern poate selecta automat (p, d, q):

- Python: `pmdarima.auto_arima()`
- R: `forecast::auto.arima()`

Cum funcționează Auto-ARIMA

1. Folosește teste de rădăcină unitate pentru a determina d
2. Potrivește modele pentru diverse combinații (p, q)
3. Selectează modelul cu cel mai mic AIC/BIC
4. Opțional folosește căutare pas cu pas pentru eficiență

Atenție

Selecția automată este utilă dar nu infailibilă. Verificați întotdeauna validitatea modelului!

Metode de estimare

Estimarea prin Maximum de Verosimilitate (MLE)

Abordarea standard pentru ARIMA:

- Presupune $\varepsilon_t \sim N(0, \sigma^2)$
- Maximizează funcția de verosimilitate
- Oferă estimatori consistenți, eficienți
- Furnizează erori standard pentru inferență

MLE Condiționată vs Exactă

- **MLE Condiționată**: Condiționează pe valorile inițiale
- **MLE Exactă**: Tratează valorile inițiale ca necunoscute
- Diferența diminuează pe măsură ce dimensiunea eșantionului crește

Log-verosimilitatea condiționată

Funcția de log-verosimilitate gaussiană

Pentru ARIMA(p, d, q) cu $\varepsilon_t \sim N(0, \sigma^2)$: $\ell(\boldsymbol{\theta}, \sigma^2) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T} e_t^2(\boldsymbol{\theta})$ unde $e_t(\boldsymbol{\theta}) = X_t - \hat{X}_{t|t-1}$ sunt **erorile de predicție la un pas**, iar $\boldsymbol{\theta} = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, c)$.

Exemplu: ARIMA(1,1,1)

Erorile de predicție se calculează recursiv: $e_t = \Delta X_t - \phi_1 \Delta X_{t-1} - \theta_1 e_{t-1} - c$ MLE condiționată: fixează $e_0 = 0$, calculează $e_1, e_2, \ldots, e_T$, maximizează ℓ.

Estimarea lui σ^2

La parametrii optimi $\hat{\boldsymbol{\theta}}$: $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T} e_t^2(\hat{\boldsymbol{\theta}})$

Restricții asupra parametrilor

Staționaritate și invertibilitate

Modelul ARIMA estimat ar trebui să satisfacă:

- **Staționaritate AR**: Rădăcinile lui $\phi(z) = 0$ în afara cercului unitate
- **Invertibilitate MA**: Rădăcinile lui $\theta(z) = 0$ în afara cercului unitate

Verificare în Practică

Majoritatea software-ului raportează:

- Coeficienți estimați cu erori standard
- Rădăcinile polinoamelor AR și MA
- Avertisment dacă este detectată aproape-rădăcină-unitate

Diagnosticul modelului ARIMA

Verificări esențiale (aceleași ca pentru ARMA, cf. Cap. 2)

Dacă modelul este corect, reziduurile $\hat{\varepsilon}_t$ trebuie să fie zgomot alb:

1. **ACF/PACF rezidual**: fără vârfuri semnificative
2. **Testul Ljung-Box**: p-value $> 0.05 \Rightarrow$ fără autocorelare
3. **Graficul Q-Q**: verificarea normalității
4. **Heteroscedasticitate**: varianță constantă a reziduurilor

Aspecte specifice ARIMA

- Testul Ljung-Box: alegeți $m \approx \ln(n)$ sau $m = 10$ (trimestrial), $m = 20$ (lunar)
- Grade de libertate: $\chi^2(m - p - q)$, ajustate pentru p și q estimați
- Dacă testul eșuează: adăugați termeni AR/MA sau verificați rupturi structurale

Prognoze punctuale

Prognoza cu MSE Minim

Prognoză optimă la h pași înainte este speranța condiționată: $\hat{Y}_{T+h|T} = \mathbb{E}[Y_{T+h}|Y_T, Y_{T-1}, \ldots]$

Prognoza ARIMA(1,1,1)

Model: $(1 - \phi_1 L)(1 - L)Y_t = c + (1 + \theta_1 L)\varepsilon_t$

Prognoză un pas: $\hat{Y}_{T+1|T} = c + Y_T + \phi_1(Y_T - Y_{T-1}) + \theta_1 \hat{\varepsilon}_T$

Pentru $h > 1$: înlocuiți ε_{T+j} necunoscut cu 0, Y_{T+j} necunoscut cu $\hat{Y}_{T+j|T}$

Intervale de prognoză

Incertitudinea prognozei

Varianța erorii de prognoză la h pași: $\text{Var}(e_{T+h}) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2$, unde ψ_j sunt coeficienții MA(∞).

Intervale de încredere

Sub normalitate, interval $(1-\alpha)\%$: $\hat{Y}_{T+h|T} \pm z_{\alpha/2}\sqrt{\text{Var}(e_{T+h})}$

Proprietate Cheie pentru Serii I(1)

Pentru procese integrate, varianța prognozei crește nelimitat când $h \to \infty$. Intervalele se lărgesc în timp!

Prognoze pe termen lung pentru ARIMA

Comportament când $h \to \infty$

Pentru ARIMA(p,1,q) cu drift c:

- Prognoze punctuale: Trend liniar cu pantă = drift
- Intervale de prognoză: Lățimea crește cu $\sqrt{h}$

Pentru ARIMA(p,1,q) fără drift:

- Prognoze punctuale: Converg la ultimul nivel
- Intervale de prognoză: Tot cresc nelimitat

Implicație Practică

Prognozele ARIMA sunt cele mai fiabile pentru orizonturi scurte. Prognozele pe termen lung au benzi de incertitudine foarte largi.

Prognoza rollingă: concept

Ce este Prognoza Rollingă?

O tehnică pentru evaluarea acurateții prognozei în afara eșantionului:

1. Fixăm o **fereastră de antrenament** de dimensiune w
2. Estimăm modelul pe observațiile $t = 1, \ldots, w$
3. Prognozăm h pași înainte: $\hat{Y}_{w+h|w}$
4. **Deplasăm** fereastra înainte cu o perioadă
5. Repetăm până la sfârșitul eșantionului

De ce Prognoze Rollinge?

- Mimează scenariul de prognoză în timp real
- Oferă multiple erori de prognoză pentru evaluare
- Evită supraajustarea pe întregul eșantion

Prognoza rollingă: exemplu pas cu pas

Configurare: ARIMA(1,1,0) cu $\phi_1 = 0.6$

Model: $\Delta Y_t = \phi_1 \Delta Y_{t-1} + \varepsilon_t$ unde $\Delta Y_t = Y_t - Y_{t-1}$

Date la Momentul T

$Y_{T-2} = 100, \quad Y_{T-1} = 103, \quad Y_T = 108 \quad \Rightarrow \quad \Delta Y_{T-1} = 3, \quad \Delta Y_T = 5$

Prognoza Punctuală la 1 Pas

$$\hat{\Delta Y}_{T+1|T} = \phi_1 \cdot \Delta Y_T = 0.6 \times 5 = 3$$
$$\hat{Y}_{T+1|T} = Y_T + \hat{\Delta Y}_{T+1|T} = 108 + 3 = \boxed{111}$$

Prognoze punctuale multi-pas

Prognoza la 2 Pași

$$\Delta\hat{Y}_{T+2|T} = \phi_1 \cdot \Delta\hat{Y}_{T+1|T} = 0.6 \times 3 = 1.8$$
$$\hat{Y}_{T+2|T} = \hat{Y}_{T+1|T} + \Delta\hat{Y}_{T+2|T} = 111 + 1.8 = \boxed{112.8}$$

Formula Generală pentru Prognoza la h Pași (ARIMA(1,1,0))

$$\Delta\hat{Y}_{T+h|T} = \phi_1^h \cdot \Delta Y_T$$
$$\hat{Y}_{T+h|T} = Y_T + \Delta Y_T \cdot \frac{\phi_1(1-\phi_1^h)}{1-\phi_1}$$

Numeric: Prognoza la 3 Pași

$$\hat{Y}_{T+3|T} = 108 + 5 \times \frac{0.6(1-0.6^3)}{1-0.6} = 108 + 5 \times 1.176 = \boxed{113.88}$$

Intervale de încredere: formule

Varianța erorii de prognoză

Pentru ARIMA(1,1,0), varianța erorii de prognoză la h pași:

$$\text{Var}(e_{T+h|T}) = \sigma^2 \left(1 + \sum_{j=1}^{h-1} \psi_j^2\right)$$

unde $\psi_j = 1 + \phi_1 + \cdots + \phi_1^j = \frac{1-\phi_1^{j+1}}{1-\phi_1}$ pentru $j \geq 0$

Interval de Încredere $(1-\alpha)\%$

$$\hat{Y}_{T+h|T} \pm z_{\alpha/2} \cdot \sqrt{\text{Var}(e_{T+h|T})}$$

Pentru IC 95%: $z_{0.025} = 1.96$

Interval de încredere: exemplu numeric

Date: $\sigma^2 = 4$, $\phi_1 = 0.6$, $\hat{Y}_{T+1|T} = 111$

IC la 1 Pas

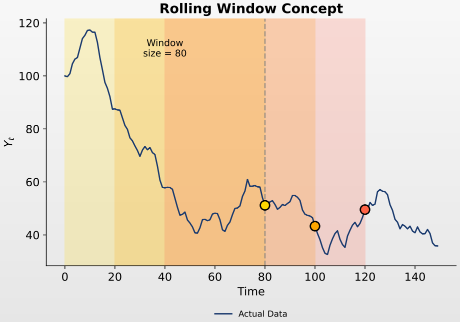
$$\text{Var}(e_{T+1|T}) = \sigma^2 = 4$$
$$\text{IC } 95\% = 111 \pm 1.96 \times \sqrt{4} = 111 \pm 3.92 = \boxed{[107.08,\ 114.92]}$$

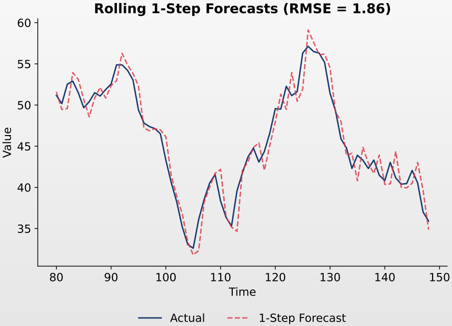
IC la 2 Pași (pentru $\hat{Y}_{T+2|T} = 112.8$)

$$\psi_1 = 1 + \phi_1 = 1.6, \quad \text{Var}(e_{T+2|T}) = 4(1 + 1.6^2) = 14.24$$
$$\text{IC } 95\% = 112.8 \pm 1.96 \times \sqrt{14.24} = 112.8 \pm 7.40 = \boxed{[105.40,\ 120.20]}$$

Notă: IC se lărgește pe măsură ce orizontul crește!

Ilustrație fereastră rollingă





- Fiecare fereastră produce o prognoză la 1 pas
- Comparăm prognozele cu valorile reale pentru a calcula RMSE, MAE
- Fereastra rollingă menține estimarea modelului actualizată

TSA_ch3_rolling_forecast

Prognoza Rollingă: Cod Python

Implementare

```
from statsmodels.tsa.arima.model import ARIMA

window_size = 100
forecasts, actuals = [], []

for t in range(window_size, len(y) - 1):
    train = y[:t]                      # Fereastra rollinga
    model = ARIMA(train, order=(1,1,0)).fit()
    forecast = model.forecast(steps=1)[0]
    forecasts.append(forecast)
    actuals.append(y[t])

rmse = np.sqrt(np.mean((np.array(forecasts) - np.array(actuals))**2))
```

Studiu de caz: analiză ARIMA completă

Obiectiv

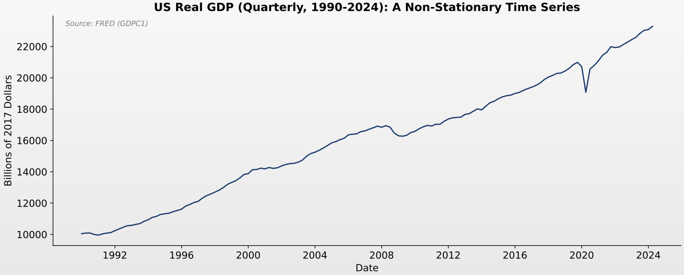
Prognoză PIB Real al SUA folosind metodologia Box-Jenkins

1. **Pasul 1**: vizualizarea datelor și verificarea staționarității
2. **Pasul 2**: Aplicarea testelor de rădăcină unitate (ADF, KPSS)
3. **Pasul 3**: Diferențiere dacă e necesar, identificare p și q
4. **Pasul 4**: Estimarea modelului ARIMA
5. **Pasul 5**: Diagnosticul modelului
6. **Pasul 6**: Generarea prognozelor cu intervale de încredere
7. **Pasul 7**: Evaluarea acurateții prognozei

Date

PIB Real SUA (FRED: GDPC1), Trimestrial, 1990T1–2024T2, $n = 138$ observații

Pasul 1: analiza inițială a datelor



Observații

- Trend ascendent clar $\Rightarrow$ medie neconstantă
- Varianța pare relativ stabilă (după transformare log)
- Scădere notabilă în 2020 (pandemia COVID-19)
- **Concluzie**: Seria este nestaționară, necesită diferențiere

Pasul 2: testarea rădăcinii unitate

Test ADF pe Log PIB în Niveluri

- Statistică test: -0.91
- Valori critice: -3.48 (1%), -2.88 (5%), -2.58 (10%)
- p-value: 0.79
- **Rezultat**: Nu putem respinge $H_0 \Rightarrow$ Rădăcină unitate prezentă

Test ADF pe Prima Diferență (Rata de Creștere)

- Statistică test: -13.24
- p-value: < 0.001
- **Rezultat**: Respingem H_0 la 1% $\Rightarrow$ Staționar după diferențiere

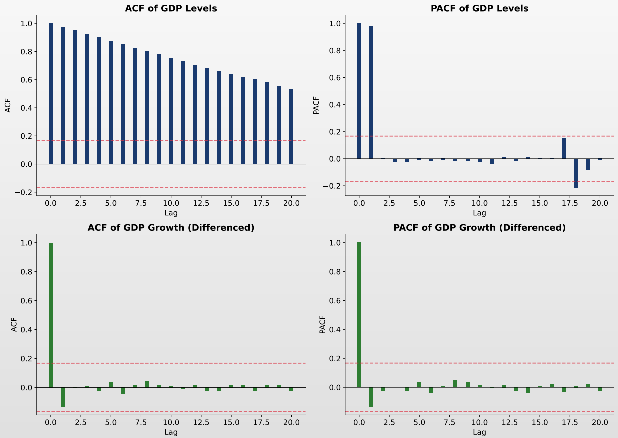
Concluzie

PIB este $I(1) \Rightarrow$ Folosim $d = 1$ în modelul ARIMA

Pasul 3: Identificarea modelului prin ACF/PACF

Analiza seriei diferențiate

- **ACF**: Vârf la lag 1, apoi se întrerupe
  - Sugerează MA(1)
- **PACF**: Vârf la lag 1, scade
  - Sugerează AR(1)
- **Candidate**: ARIMA(1,1,0), ARIMA(0,1,1), ARIMA(1,1,1)



Pasul 4: Estimarea modelului

Compararea modelelor folosind Criterii informaționale

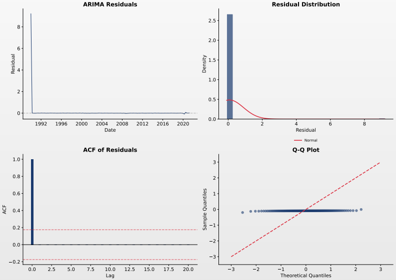
Model	**AIC**	**BIC**	**Log-Lik**
ARIMA(1,1,0)	−725.2	−719.5	364.6
ARIMA(0,1,1)	−724.8	−719.2	364.4
ARIMA(1,1,1)	−747.0	−738.5	376.5

Model Selectat: ARIMA(1,1,1)

$$(1 - 0.35L)(1 - L)Y_t = (1 + 0.58L)\varepsilon_t, \quad \hat{\sigma}^2 = 0.000156$$

- $\hat{\phi}_1 = 0.35$ (SE = 0.09), semnificativ la 1%
- $\hat{\theta}_1 = 0.58$ (SE = 0.08), semnificativ la 1%

Pasul 5: diagnosticul modelului



Analiza reziduurilor

- Test Ljung-Box: $Q(10) = 5.8$, p-value $= 0.83 \Rightarrow$ Fără autocorelare
- Test Jarque-Bera: $JB = 156.4$, p-value $< 0.001 \Rightarrow$ Non-normal (outlier COVID)
- **Concluzie**: Modelul trece verificările de autocorelare; outlierii sunt așteptați

Pasul 6: Prognoza cu intervale de încredere

Ultimele Valori Observate (Log PIB)

$Y_T = 9.973$ (2024T2), $Y_{T-1} = 9.956$ (2024T1)
$\Delta Y_T = 0.017$, $\hat{\varepsilon}_T = 0.004$

Prognoza la 1 Pas (2024T3)

$$\Delta \hat{Y}_{T+1} = \hat{\phi}_1 \Delta Y_T + \hat{\theta}_1 \hat{\varepsilon}_T = 0.35(0.017) + 0.58(0.004) = 0.0083$$

$$\hat{Y}_{T+1} = 9.973 + 0.0083 = \boxed{9.981}$$

Interval de Încredere 95%

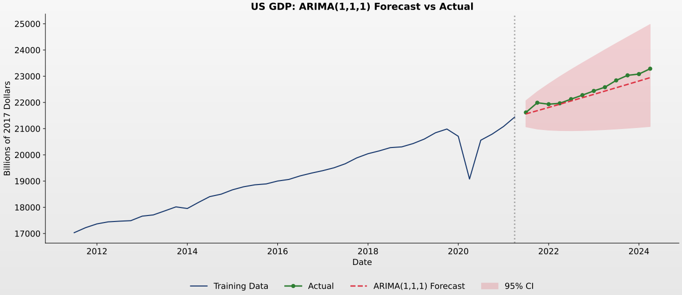
$$\text{IC} = 9.981 \pm 1.96 \times \sqrt{0.000156} = [9.957,\ 10.006]$$

În niveluri: Prognoză PIB = \$21,652 mld, IC = [\$21,142 mld, \$22,175 mld]

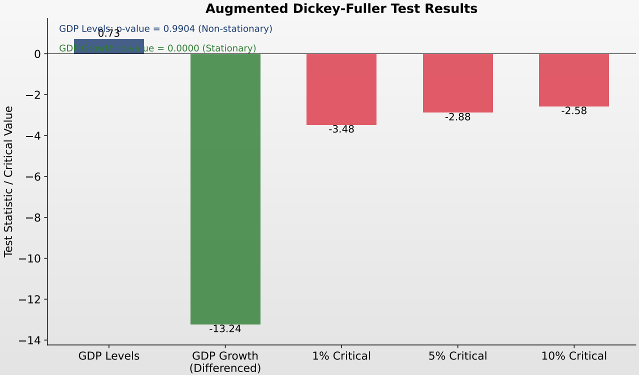
Pasul 7: Evaluarea prognozei

Performanță out-of-sample (ultimele 12 trimestre)

- RMSE = 0.0486 $\approx$ 4.86% eroare
- MAE = 0.0430 $\approx$ 4.30% eroare
- Acuratețe direcție = 91%
  - A prezis corect creștere/scădere

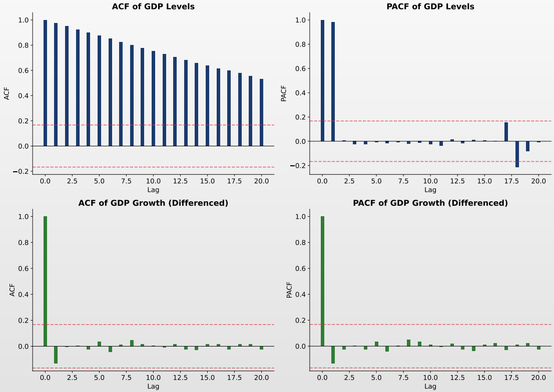


Rezultatele testului de rădăcină unitate



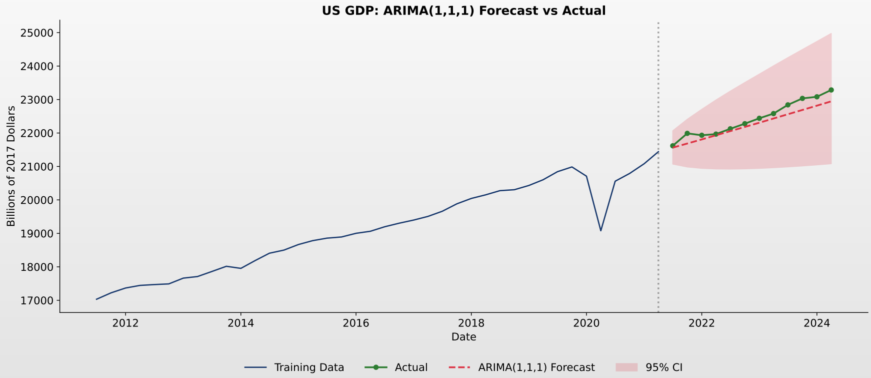
- PIB în niveluri: Nu putem respinge rădăcina unitate (nestaționar)
- Creștere PIB: Respingem rădăcina unitate la nivel de 1% (staționar)

ACF/PACF: niveluri vs diferențiat



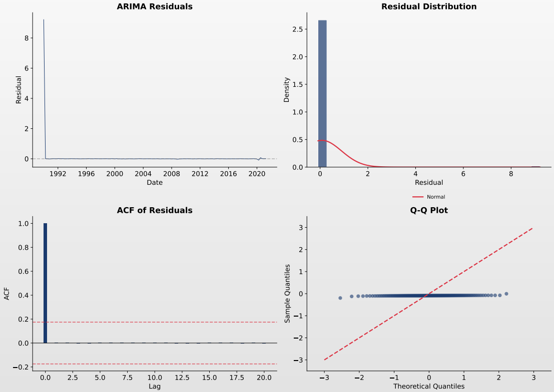
- **Sus**: Descreștere lentă ACF în niveluri sugerează nestaționaritate
- **Jos**: După diferențiere, ACF/PACF ajută la identificarea lui p și q

Prognoza ARIMA: real vs prezis



- ARIMA(1,1,1) captează dinamică trendului
- Intervalele de încredere se lărgesc cu orizontul de prognoză

Diagnosticul modelului



- Reziduurile par aleatorii; ACF în limitele benzilor
- Graficul Q-Q arată normalitate aproximativă

Implementare Python

Biblioteci Cheie

```
import pandas as pd
import numpy as np
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.tsa.stattools import adfuller, kpss
import pmdarima as pm
```

Exemplu Auto-ARIMA

```
model = pm.auto_arima(y, start_p=0, start_q=0,
                      max_p=3, max_q=3, d=None,
                      seasonal=False, trace=True)
print(model.summary())
```

Concluzii cheie

Puncte Principale

1. **Nestaționaritatea** este frecventă în datele economice – trebuie abordată
2. **Diferențierea** transformă I(d) în I(0)
3. **ARIMA(p,d,q)** combină diferențierea cu modelarea ARMA
4. **Testele de rădăcină unitate** (ADF, KPSS) ajută la determinarea lui d
5. **Metodologia Box-Jenkins**: Identificare $\rightarrow$ Estimare $\rightarrow$ Validare
6. **Prognozele** pentru serii I(1) au incertitudine în creștere

Pașii Următori

Capitolul 4 va extinde ARIMA pentru a gestiona sezonalitatea: modele SARIMA.

Întrebarea 1

Întrebare

O serie de timp Y_t urmează un mers aleatoriu: $Y_t = Y_{t-1} + \varepsilon_t$. Care este Var($Y_t$)?

(A) σ^2 (constantă)

(B) $t \cdot \sigma^2$ (crește liniar în timp)

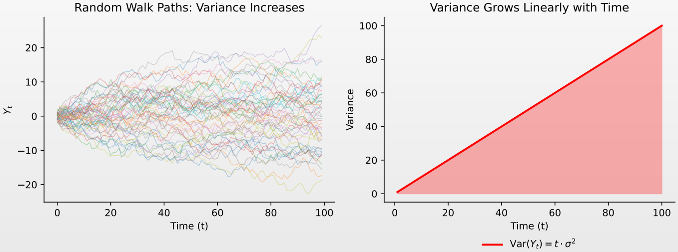
(C) σ^2/t (scade în timp)

(D) σ^{2t} (crește exponențial)

Întrebarea 1: Răspuns

Răspuns Corect: (B)
$\mathrm{Var}(Y_t) = t \cdot \sigma^2$

Varianța mersului aleatoriu crește liniar în timp — de aceea mersurile aleatorii sunt nestaționare.



TSA_ch3_quiz1_rw_variance

Întrebarea 2

Întrebare

Dacă o serie Y_t este I(2), de câte ori trebuie diferențiată pentru a atinge staționaritatea?

(A) 0 ori (deja staționară)

(B) 1 dată

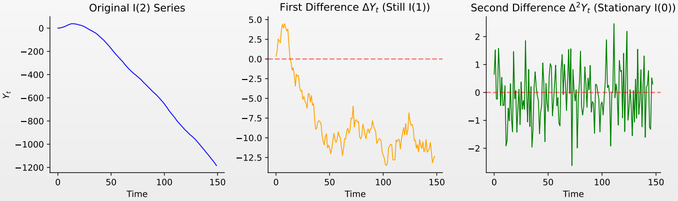
(C) 2 ori

(D) Nu poate fi făcută staționară prin diferențiere

Întrebarea 2: Răspuns

Răspuns Corect: (C) 2 ori

I(d) înseamnă “integrată de ordin d” — necesită d diferențe pentru staționaritate.



TSA_ch3_quiz2_differencing

Întrebarea 3

Întrebare

Rulați un test ADF și obțineți o statistică de test de -2.1 cu valori critice: -3.43 (1%), -2.86 (5%), -2.57 (10%). Ce concluzie trageți?

(A) Respingem H_0: seria este staționară la toate nivelurile

(B) Respingem H_0: seria este staționară doar la nivel de 10%

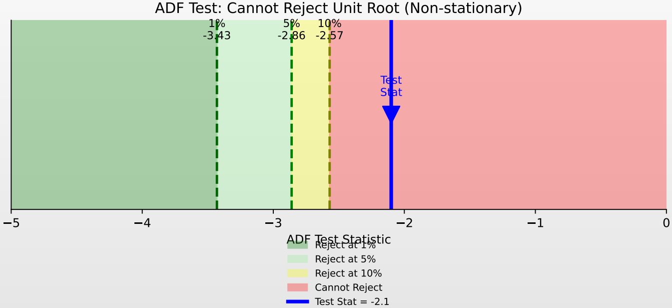
(C) Nu respingem H_0: seria probabil are rădăcină unitate

(D) Testul este neconcludent

Întrebarea 3: Răspuns

Răspuns Corect: (C) Nu respingem H_0: seria are rădăcină unitate

Statistică de test $-2.1 > -2.57$ (VC 10%) $\Rightarrow$ Nu putem respinge la niciun nivel. Luați în considerare diferențierea.



TSA_ch3_quiz3_adf_test

Întrebarea 4

Întrebare

Pentru un model ARIMA(1,1,0), care este tiparul ACF al seriei **diferențiate** ΔY_t?

(A) Se întrerupe după lag 1

(B) Scade exponențial

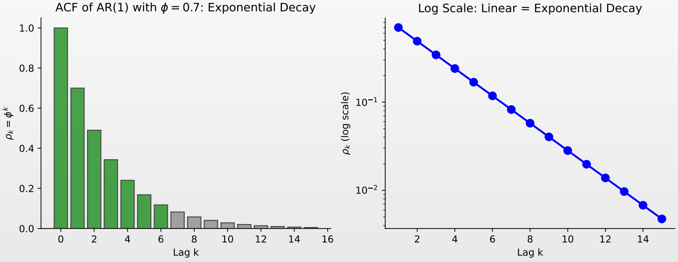
(C) Alternează în semn

(D) Este zero la toate lag-urile

Întrebarea 4: Răspuns

Răspuns Corect: (B) Scade exponențial

ARIMA(1,1,0) $\Rightarrow \Delta Y_t$ urmează AR(1) cu ACF $\rho_k = \phi_1^k$ (descreștere geometrică).



TSA_ch3_quiz4_acf_decay

Întrebarea 5

Întrebare

Ce se întâmplă cu intervalele de încredere ale prognozei ARIMA pe măsură ce orizontul h crește pentru o serie I(1)?

(A) Rămân constante

(B) Se îngustează (mai multă precizie)

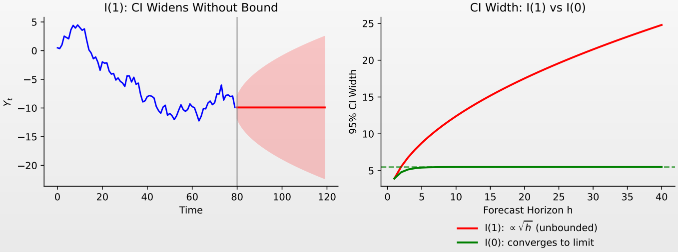
(C) Se lărgesc nelimitat

(D) Se lărgesc dar converg la o limită

Întrebarea 5: Răspuns

Răspuns Corect: (C) Se lărgesc nelimitat

Pentru I(1): lățimea IC $\propto \sqrt{h}$ (nelimitată). Pentru I(0): IC converg la o limită.



TSA_ch3_quiz5_forecast_ci

Bibliografie I

Teste de rădăcină unitate

- Dickey, D.A., & Fuller, W.A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *JASA*, 74(366), 427–431.
- Phillips, P.C.B., & Perron, P. (1988). Testing for a Unit Root in Time Series Regression, *Biometrika*, 75(2), 335–346.
- Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., & Shin, Y. (1992). Testing the Null Hypothesis of Stationarity, *Journal of Econometrics*, 54(1-3), 159–178.

Modele ARIMA și selecție automată

- Box, G.E.P., & Jenkins, G.M. (1970). *Time Series Analysis: Forecasting and Control*, Holden-Day.
- Hyndman, R.J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The `forecast` Package for R, *Journal of Statistical Software*, 27(3), 1–22.

Bibliografie II

Manuale și referințe suplimentare

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

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