Analiza și Prognoza Seriilor de Timp

Seminar 2: Modele ARMA

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

Cuprins Seminar

Test 1: Operatorul Lag

Întrebare

Care este rezultatul aplicării $(1 - L)^2$ lui X_t?

A. $X_t - X_{t-1}$

B. $X_t - 2X_{t-1} + X_{t-2}$

C. $X_t + X_{t-1} + X_{t-2}$

D. $X_t - X_{t-2}$

Test 1: Răspuns

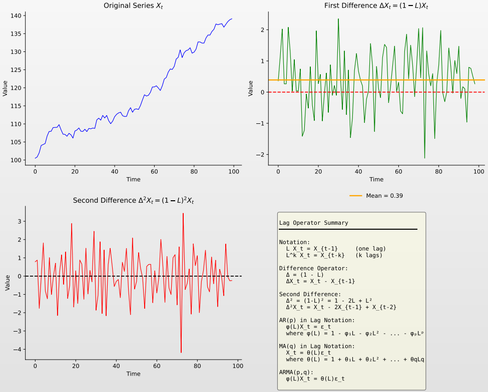
Răspuns: B

$X_t - 2X_{t-1} + X_{t-2}$

Explicație:

$$(1-L)^2 X_t = (1 - 2L + L^2) X_t$$
$$= X_t - 2X_{t-1} + X_{t-2}$$

Aceasta este **diferența de ordinul doi** a lui X_t.



Operatorul lag: $L^k X_t = X_{t-k}$

TSA_ch2_lag_operator

Test 2: Staționaritatea AR(1)

Întrebare

Pentru ce valoare a lui ϕ procesul AR(1) $X_t = 0.5 + \phi X_{t-1} + \varepsilon_t$ este staționar?

A. $\phi = 1.2$

B. $\phi = 1.0$

C. $\phi = -0.8$

D. $\phi = -1.5$

Test 2: Răspuns

Răspuns: C

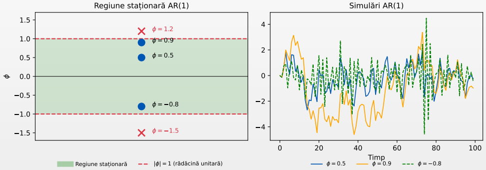
$\phi = -0.8$ (Staționar)

Condiția de staționaritate AR(1):

$$|\phi| < 1$$

Verificarea fiecărei opțiuni:

- A: $|1.2| = 1.2 > 1$ ×
- B: $|1.0| = 1$ (rădăcină unitară) ×
- C: $|-0.8| = 0.8 < 1$ ✓
- D: $|-1.5| = 1.5 > 1$ ×



AR(1): regiunea staționară $|\phi| < 1$

TSA_ch2_ar1

Test 3: Modelul ACF

Întrebare

Observați următorul model ACF: vârf semnificativ la lag 1, apoi toate lag-urile în benzile de încredere. PACF arată descreștere graduală. Ce model este sugerat?

A. AR(1)

B. MA(1)

C. ARMA(1,1)

D. Zgomot alb

Test 3: Răspuns

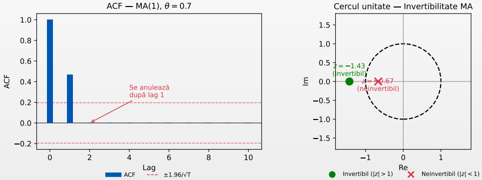
Răspuns: B

MA(1)

Regula cheie de identificare:

- ACF se anulează după lag $q \Rightarrow$ MA(q)
- PACF se anulează după lag $p \Rightarrow$ AR(p)

Aici: ACF se anulează la lag 1, PACF descrește $\Rightarrow$ **MA(1)**



MA(1): ACF se anulează după lag 1

TSA_ch2_ma1

Test 4: Invertibilitatea MA

Întrebare

Pentru procesul MA(1) $X_t = \varepsilon_t + 1.5\varepsilon_{t-1}$, este procesul invertibil?

A. Da, deoarece procesele MA sunt întotdeauna invertibile

B. Da, deoarece $1.5 > 0$

C. Nu, deoarece $|\theta| = 1.5 > 1$

D. Nu, deoarece procesele MA nu sunt niciodată invertibile

Test 4: Răspuns

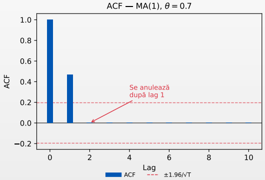
Răspuns: C

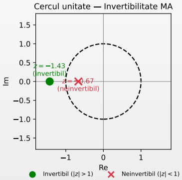
Nu este invertibil ($|\theta| = 1.5 > 1$)

Invertibilitatea MA(1):
Necesită $|\theta| < 1$

Echivalent: rădăcina lui $\theta(z) = 1 + \theta z = 0$ trebuie să fie în afara cercului unitate.
Aici: $z = -1/1.5 = -0.67$ este **în interior**.

$\Rightarrow$ **Nu este invertibil**





Invertibilitate: rădăcina în afara cercului unitate

TSA_ch2_ma1

Test 5: Reprezentarea ARMA

Întrebare

Forma compactă $\phi(L)X_t = \theta(L)\varepsilon_t$ reprezintă ce model?

A. Model AR pur
B. Model MA pur
C. Model ARMA
D. Niciunul dintre cele de mai sus

Test 5: Răspuns

Răspuns: C

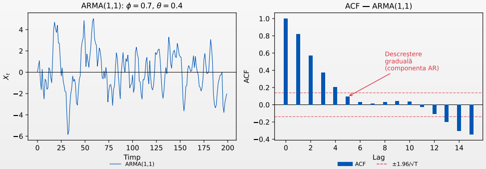
Model ARMA

Notația cu polinoame lag:

- $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$
- $\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$

Cazuri speciale:

- $\theta(L) = 1$: AR pur
- $\phi(L) = 1$: MA pur



ARMA(1,1): combină AR și MA

TSA_ch2_arma

Test 6: Criterii Informaționale

Întrebare

Când comparăm ARMA(1,1) și ARMA(2,1) folosind BIC, care afirmație este corectă?

A. BIC mai mic înseamnă întotdeauna prognoze mai bune
B. BIC penalizează complexitatea mai puțin decât AIC
C. Modelul cu BIC mai mic este preferat
D. BIC poate compara doar modele cu același număr de parametri

Test 6: Răspuns

Răspuns: C

Modelul cu BIC mai mic este preferat

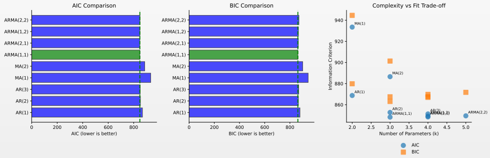
Criterii Informaționale:

$$\text{AIC} = -2\ln(\hat{L}) + 2k$$

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

BIC penalizează complexitatea **mai mult** decât AIC (pentru $n > 7$).

$\Rightarrow$ BIC favorizează modele mai simple.



Selecția modelului: AIC vs BIC

TSA_ch2_model_selection

Test 7: Testul Ljung-Box

Întrebare

După ajustarea unui model ARMA(2,1), rulați testul Ljung-Box pe reziduuri și obțineți valoare-p $= 0.02$. Ce concluzie trageți?

A. Modelul este adecvat
B. Reziduurile sunt zgomot alb
C. Există autocorelație semnificativă în reziduuri
D. Modelul are prea mulți parametri

Test 7: Răspuns

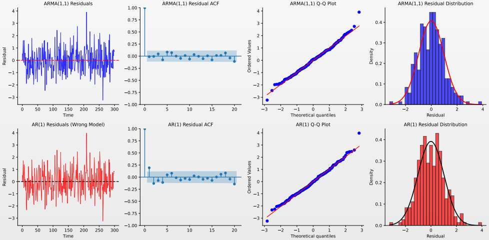
Răspuns: C

Autocorelație semnificativă în reziduuri

Testul Ljung-Box:

- H_0: Reziduurile sunt zgomot alb
- H_1: Autocorelație prezentă

valoare-p $= 0.02 < 0.05$
$\Rightarrow$ **Respingem H_0**
Modelul este **inadecvat** — încercați alte ordine.



Diagnostice: ACF trebuie să fie zgomot alb

TSA_ch2_diagnostics

Test 8: Prognoză

Întrebare

Pentru un model AR(1) cu $\phi = 0.6$ și medie $\mu = 10$, ce se întâmplă cu prognozele când orizontul $h \to \infty$?

A. Prognozele cresc fără limită

B. Prognozele converg la 0

C. Prognozele converg la $\mu = 10$

D. Prognozele oscilează pentru totdeauna

Test 8: Răspuns

Răspuns: C

Prognozele converg la $\mu = 10$

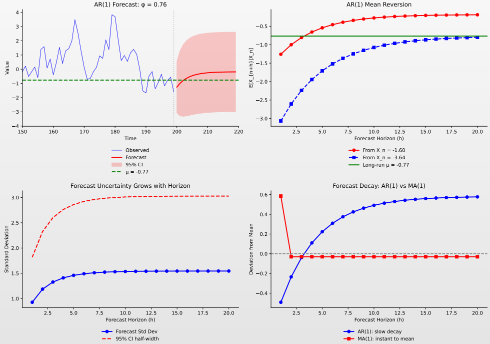
Formula de prognoză AR(1):

$$\hat{X}_{n+h|n} = \mu + \phi^h (X_n - \mu)$$

Deoarece $|\phi| = 0.6 < 1$:

$$\lim_{h \to \infty} \phi^h = 0$$

$\Rightarrow$ Prognozele converg la μ.
Revenire la medie.



Prognoze AR(1): revenire la medie

TSA_ch2_forecasting

Test 9: Rădăcinile AR(2)

Întrebare

Un proces AR(2) are rădăcinile caracteristice (ale $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 = 0$) $z_1 = 0.8$ și $z_2 = -0.5$. Este staționar?

A. Da, deoarece ambele rădăcini sunt în interiorul cercului unitate
B. Nu, deoarece o rădăcină este negativă
C. Nu, deoarece rădăcinile trebuie să fie în afara cercului unitate
D. Nu se poate determina fără mai multe informații

Test 9: Răspuns

Răspuns: C

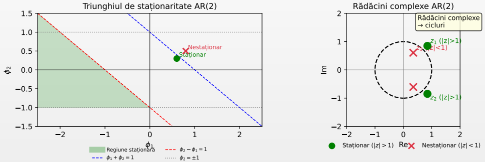
Rădăcinile trebuie să fie în afara cercului unitate

Condiția de staționaritate:
Rădăcinile lui $\phi(z) = 0$ trebuie să fie **în afara** cercului unitate ($|z| > 1$).

Aici:

- $|z_1| = 0.8 < 1$ ×
- $|z_2| = 0.5 < 1$ ×

Ambele în interior $\Rightarrow$ **Nestaționar**



AR(2): rădăcini și triunghiul de staționaritate

TSA_ch2_ar2

Test 10: Proprietățile MA(q)

Întrebare

Pentru un proces MA(2), ACF-ul:

A. Descrește exponențial
B. Se anulează după lag 2
C. Se anulează după lag 1
D. Nu se anulează niciodată

Test 10: Răspuns

Răspuns: B

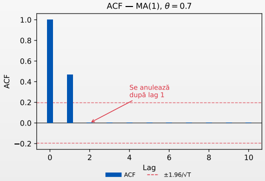
Se anulează după lag 2

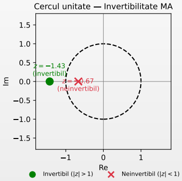
Proprietatea ACF pentru MA(q):

$$\rho(h) = 0 \text{ pentru } h > q$$

- MA(1): ACF se anulează după lag 1
- MA(2): ACF se anulează după lag 2
- MA(q): ACF se anulează după lag q

Caracteristica cheie de identificare.





MA: anularea ACF este semnătura

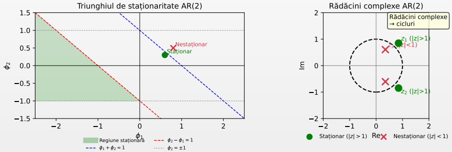
TSA_ch2_ma1

Adevărat sau Fals? — Întrebări

Afirmație	**A/F?**
1. Un proces AR(2) poate prezenta comportament pseudo-ciclic.	?
2. Procesele MA necesită o condiție de staționaritate.	?
3. PACF-ul unui proces AR(p) se anulează după lag p.	?
4. Dacă AIC selectează ARMA(2,1) și BIC selectează ARMA(1,1), nu pot fi ambele corecte.	?
5. Intervalele de încredere se îngustează pe măsură ce orizontul crește.	?
6. Ecuațiile Yule-Walker pot fi folosite pentru a estima parametrii MA.	?

Adevărat sau Fals? — Răspunsuri

1. **ADEVĂRAT**: AR(2) cu rădăcini complexe ⇒ oscilații amortizate
2. **FALS**: Procesele MA sunt întotdeauna staționare; au nevoie de *invertibilitate*
3. **ADEVĂRAT**: Caracteristica cheie de identificare a AR(p)
4. **FALS**: Ambele sunt „corecte" pentru criteriile lor (AIC: estimare, BIC: parcimonie)
5. **FALS**: IC se *lărgesc* cu orizontul (mai multă incertitudine)
6. **FALS**: Yule-Walker este pentru AR; MA folosește MLE



AR(2): rădăcini complexe ⇒ cicluri

TSA_ch2_ar2

Exercițiul 1: Proprietățile AR(1)

Problemă: Considerați procesul AR(1):

$$X_t = 2 + 0.7X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, 9)$$

Calculați:

1. Media μ
2. Varianța $\gamma(0)$
3. Autocovarianța $\gamma(1)$ și $\gamma(2)$
4. Autocorelația $\rho(1)$ și $\rho(2)$

Exercițiul 1: Soluție

Dat: $c = 2$, $\phi = 0.7$, $\sigma^2 = 9$

1. **Media:**

$$\mu = \frac{c}{1-\phi} = \frac{2}{1-0.7} = \frac{2}{0.3} = \mathbf{6.67}$$

2. **Varianța:**

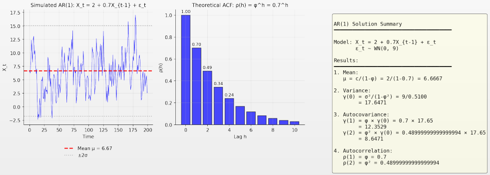
$$\gamma(0) = \frac{\sigma^2}{1-\phi^2} = \frac{9}{1-0.49} = \mathbf{17.65}$$

3. **Autocovarianța:**

$$\gamma(1) = \phi \cdot \gamma(0) = 0.7 \times 17.65 = \mathbf{12.35}$$
$$\gamma(2) = \phi^2 \cdot \gamma(0) = 0.49 \times 17.65 = \mathbf{8.65}$$

4. **Autocorelația:**

$$\rho(1) = \phi = \mathbf{0.7}, \quad \rho(2) = \phi^2 = \mathbf{0.49}$$



Simulare AR(1) și ACF

Exercițiul 2: Proprietățile MA(1)

Problemă: Considerați procesul MA(1):

$$X_t = 5 + \varepsilon_t - 0.4\varepsilon_{t-1}, \quad \varepsilon_t \sim WN(0, 4)$$

Calculați:

1. Media μ
2. Varianța $\gamma(0)$
3. Autocovarianța $\gamma(1)$
4. Autocorelația $\rho(1)$
5. Este acest proces invertibil?

Exercițiul 2: Soluție

Dat: $\mu = 5$, $\theta = -0.4$, $\sigma^2 = 4$

1. **Media:**

$$\mathbb{E}[X_t] = \mu = \mathbf{5}$$

2. **Varianța:**

$$\gamma(0) = \sigma^2(1+\theta^2) = 4(1.16) = \mathbf{4.64}$$

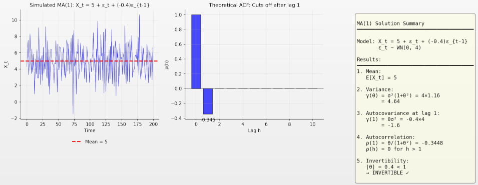
3. **Autocovarianța:**

$$\gamma(1) = \theta\sigma^2 = -0.4 \times 4 = \mathbf{-1.6}$$

4. **Autocorelația:**

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{-1.6}{4.64} = \mathbf{-0.345}$$

5. **Invertibilitate:** $|\theta| = 0.4 < 1 \Rightarrow$ Da



MA(1): ACF se anulează după lag 1

TSA_ch2_ex2_ma1

Exercițiul 3: Rădăcinile Caracteristice

Problemă: Considerați procesul AR(2):

$$X_t = 0.5X_{t-1} + 0.3X_{t-2} + \varepsilon_t$$

1. Scrieți ecuația caracteristică
2. Găsiți rădăcinile caracteristice
3. Este acest proces staționar?

Exercițiul 3: Soluție

1. Ecuația caracteristică:

$$\phi(z) = 1 - 0.5z - 0.3z^2 = 0$$

Sau: $0.3z^2 + 0.5z - 1 = 0$

2. Rădăcinile (formula cuadratică):

$$z = \frac{-0.5 \pm \sqrt{0.25 + 1.2}}{0.6}$$

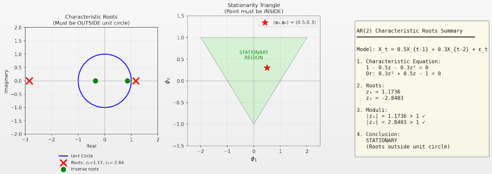
$$z_1 = \mathbf{1.17}, \quad z_2 = \mathbf{-2.84}$$

3. Verificarea staționarității:

$|z_1| = 1.17 > 1$ ✓

$|z_2| = 2.84 > 1$ ✓

Ambele în afara cercului unitate $\Rightarrow$ **Staționar**



Rădăcini în afara cercului $\Rightarrow$ staționar

TSA_ch2_ex3_roots

Exercițiul 4: Prognoză

Problemă: Ați ajustat un model AR(1):

$$X_t = 3 + 0.8X_{t-1} + \varepsilon_t, \quad \sigma^2 = 4$$

Dat $X_{100} = 20$, calculați:

1. Prognoza la 1 pas înainte $\hat{X}_{101|100}$
2. Prognoza la 2 pași înainte $\hat{X}_{102|100}$
3. Prognoza pe termen lung când $h \to \infty$
4. Intervalul de încredere de 95% pentru $\hat{X}_{101|100}$

Exercițiul 4: Soluție

Dat: $c = 3,\ \phi = 0.8,\ \sigma^2 = 4,\ X_{100} = 20$
Media: $\mu = \frac{3}{1-0.8} = \mathbf{15}$

1. **Prognoza la un pas:**

$$\hat{X}_{101|100} = 3 + 0.8 \times 20 = \mathbf{19}$$

2. **Prognoza la doi pași:**

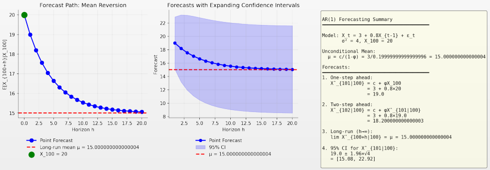
$$\hat{X}_{102|100} = 3 + 0.8 \times 19 = \mathbf{18.2}$$

3. **Prognoza pe termen lung:**

$$\lim_{h \to \infty} \hat{X}_{100+h|100} = \mu = \mathbf{15}$$

4. **IC 95%:**

$$19 \pm 1.96 \times 2 = \mathbf{[15.08,\ 22.92]}$$



Prognoza converge la medie

TSA_ch2_ex4_forecast

Exercițiu Python 1: Simulare și Ajustare AR(1)

Sarcină:

1. Simulați 300 de observații dintr-un AR(1) cu $\phi = 0.6$
2. Reprezentați grafic seria și ACF/PACF
3. Ajustați un model AR(1) și comparați $\hat{\phi}$ vs ϕ real
4. Examinați diagnosticele reziduurilor

Cod cheie:

```
from statsmodels.tsa.arima.model import ARIMA
model = ARIMA(x, order=(1, 0, 0)).fit()
print(model.summary())
```

TSA_ch2_python_simulate

Exercițiu Python 2: Selecția Modelului

Sarcină:

1. Încărcați o serie de timp și verificați staționaritatea (testul ADF)
2. Comparați AIC/BIC pentru AR(1), MA(1), ARMA(1,1), ARMA(2,1)
3. Selectați cel mai bun model
4. Generați prognoze cu intervale de încredere

Funcții cheie:

- `adfuller(x)` pentru testul de staționaritate
- `model.aic`, `model.bic` pentru criterii
- `model.get_forecast(h)` pentru predicții

TSA_ch2_python_selection

Exercițiu Python 3: Verificarea Diagnosticelor

Sarcină: După ajustarea unui model, efectuați diagnostice complete:

1. Reprezentați grafic reziduurile în timp
2. Reprezentați grafic ACF-ul reziduurilor
3. Creați graficul Q-Q pentru normalitate
4. Rulați testul Ljung-Box

Funcții cheie:

- `model.resid` pentru reziduuri
- `plot_acf(resid)` pentru graficul ACF
- `stats.probplot(resid)` pentru graficul Q-Q
- `acorr_ljungbox(resid, lags=[10])` pentru test

TSA_ch2_python_diagnostics

Discuție 1: Selecția Modelului

Scenariu: Modelați rate de inflație lunare. După verificarea staționarității:

- ACF: semnificativ la lag-urile 1, 2, 3, apoi descrește
- PACF: semnificativ la lag-urile 1, 2, apoi se anulează
- AIC selectează ARMA(2,3)
- BIC selectează AR(2)

Întrebări:

1. Ce sugerează modelul ACF/PACF?
2. De ce nu sunt de acord AIC și BIC?
3. Ce model ați alege și de ce?
4. Ce verificări suplimentare ați efectua?

Discuție 2: Evaluarea Prognozei

Scenariu: Ajustați un model ARMA(1,1) pe randamente zilnice de acțiuni. Ajustarea în eșantion arată bine (Ljung-Box $p = 0.45$), dar RMSE în afara eșantionului este mai rău decât mersul aleatoriu.

Întrebări:

1. Este aceasta surprinzător? De ce sau de ce nu?
2. Ce ne spune despre predictibilitatea randamentelor?
3. Ar trebui să concluzionați că modelul ARMA este inutil?
4. Ce alternative ați putea considera?

Indiciu: Gândiți-vă la Ipoteza Piețelor Eficiente și la ce captează ARMA vs. volatility clustering.

Exercițiu AI: Gândire critică

Prompt de testat în ChatGPT / Claude / Copilot

"Descarcă de pe FRED numărul lunar de cereri inițiale de șomaj din SUA (seria ICNSA, neajustat sezonier) din 2010-01 până în 2024-12 (180 observații). Calculează diferențele logaritmice pentru a obține rata de creștere. Testează staționaritatea cu ADF și KPSS. Estimează modele ARMA cu ordine (p, q) între $(1, 0)$ și $(3, 3)$, selectează cel mai bun model după AIC/BIC. Verifică reziduurile (Ljung-Box, normalitate) și prognozează 6 luni. Vreau cod Python complet cu grafice."

Exercițiu:

1. Rulați prompt-ul într-un LLM la alegere și analizați critic răspunsul.
2. Verifică staționaritatea *înainte* de a estima ARMA? Folosește ambele teste (ADF + KPSS)?
3. Cum gestionează vârful COVID-19 din date? Îl tratează ca outlier sau îl ignoră?
4. Selecția ordinelor (p, q) e justificată prin ACF/PACF sau doar prin AIC?
5. Reziduurile sunt testate complet? (Ljung-Box, Q-Q plot, heteroscedasticitate)

Atenție: Codul generat de AI poate rula fără erori și arăta profesional. *Asta nu înseamnă că e corect.*

Rezumat Formule Cheie

Concept	Formula
Media AR(1)	$\mu = c/(1-\phi)$
Varianța AR(1)	$\gamma(0) = \sigma^2/(1-\phi^2)$
ACF AR(1)	$\rho(h) = \phi^h$
Staționaritate AR(1)	$\lvert\phi\rvert < 1$
Varianța MA(1)	$\gamma(0) = \sigma^2(1+\theta^2)$
ACF MA(1)	$\rho(1) = \theta/(1+\theta^2)$, $\rho(h) = 0$ pentru $h > 1$
Invertibilitate MA(1)	$\lvert\theta\rvert < 1$
Prognoza AR(1)	$\hat{X}_{n+h\mid n} = \mu + \phi^h(X_n - \mu)$
IC Prognoză	$\hat{X} \pm z_{\alpha/2} \times \sqrt{\text{MSFE}(h)}$
AIC	$-2\ln(\hat{L}) + 2k$
BIC	$-2\ln(\hat{L}) + k\ln(n)$

Exercițiu de Replicare: Box & Jenkins (1970)

Obiectiv

Replicați analiza seriei "Airline Passengers" (seria G din Box & Jenkins, 1970) cu date actualizate.

1. Descărcați seria lunară de pasageri aerieni din România (INS sau Eurostat, 2010–2024)
2. Aplicați transformarea ln și diferențierea $(1 - B)(1 - B^{12})$
3. Identificați ordinul ARIMA folosind ACF/PACF (comparați cu Fig. 9.2 din Box & Jenkins)
4. Estimați modelul ARIMA$(0, 1, 1)(0, 1, 1)_{12}$ — "airline model"
5. Verificați reziduurile: Ljung-Box, normalitate, ACF reziduuri
6. **Întrebare**: Modelul airline este încă adecvat pentru date moderne? Ce s-a schimbat?

Ref: Box, G.E.P. & Jenkins, G.M. (1970). *Time Series Analysis: Forecasting and Control*, Holden-Day.

Întrebări?

Succes la exerciții.

Următorul Seminar: ARIMA și Modele Sezoniere