Analiza și Prognoza Seriilor de Timp

Seminar 1: Introducere în Seriile de Timp

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

Notă: Unele întrebări introductive din Seminarul 0 sunt reluate aici pentru consolidare.

Formule Cheie de Rețínut

Descompunere:

- Aditivă: $X_t = T_t + S_t + \varepsilon_t$
- Multiplicativă: $X_t = T_t \times S_t \times \varepsilon_t$

Netezire Exponențială:

- SES: $\hat{X}_{t+1} = \alpha X_t + (1-\alpha)\hat{X}_t$
- Holt: adaugă trend b_t
- HW: adaugă sezonalitate S_t

Staționaritate:

- $\mathbb{E}[X_t] = \mu$ (constantă)
- $\text{Var}(X_t) = \sigma^2$ (constantă)
- $\text{Cov}(X_t, X_{t+h}) = \gamma(h)$

Mers Aleatoriu:

- $X_t = X_{t-1} + \varepsilon_t$
- $\text{Var}(X_t) = t\sigma^2$ (crește)

Rezumatul Conceptelor Cheie

Concept	Punct Cheie	Când se Folosește
Descompunere aditivă	Amplitudine sezonieră constantă	Varianță stabilă
Descompunere multiplicativă	Sezonalitatea crește cu nivelul	Varianță în creștere
SES	Doar nivel (α)	Fără trend, fără sezonalitate
Holt	Nivel + Trend (α, β)	Trend, fără sezonalitate
Holt-Winters	Nivel + Trend + Sezonalitate	Trend și sezonalitate
Testul ADF	H_0: rădăcină unitară	Test pentru nestaționaritate
Testul KPSS	H_0: staționară	Confirmă staționaritatea
Diferențiere	Elimină trendul stochastic	Mers aleatoriu, rădăcină unitară
Regresie	Elimină trendul determinist	Trend liniar/polinomial

Test 1: Bazele Seriilor de Timp

Întrebare

Care dintre următoarele *nu* este o caracteristică a datelor de tip serie de timp?

A. Observațiile sunt ordonate în timp

B. Observațiile consecutive sunt de obicei corelate

C. Observațiile sunt independente și identic distribuite

D. Datele au o ordonare temporală naturală

Răspunsul pe slide-ul următor...

Test 1: Răspuns

Răspuns: C – Observațiile sunt independente și identic distribuite

Întrebare: Care *nu* este o caracteristică a datelor de tip serie de timp?

A. Observațiile sunt ordonate în timp ×

B. Observațiile consecutive sunt de obicei corelate ×

C. **Observațiile sunt independente și identic distribuite ✓**

D. Datele au o ordonare temporală naturală ×

Observațiile seriilor de timp sunt de obicei **dependente** (autocorelate), nu independente. Ipoteza observațiilor i.i.d. este fundamentală pentru analiza transversală, dar este încălcată în seriile de timp. Această dependență temporală este ceea ce face analiza seriilor de timp unică și necesită metode specializate.

Test 2: Descompunere

Întrebare

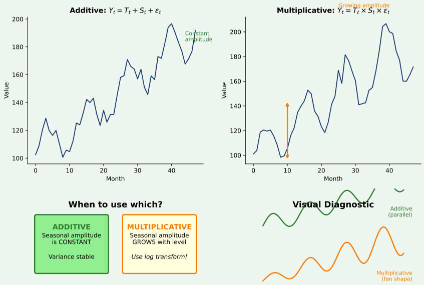
Când ar trebui să folosiți descompunerea multiplicativă în loc de cea aditivă?

A. Când tiparul sezonier are amplitudine constantă
B. Când varianța seriei este stabilă în timp
C. Când fluctuațiile sezoniere cresc proporțional cu nivelul
D. Când seria de timp nu are componentă de trend

Răspunsul pe slide-ul următor...

Test 2: Răspuns

Răspuns: C – Când fluctuațiile sezoniere cresc proporțional cu nivelul



Multiplicativă: $X_t = T_t \times S_t \times \varepsilon_t$, amplitudinea sezonieră **scalează cu nivelul** (model în evantai)

TSA_ch1_ts_basics

Test 3: Netezire Exponențială

Întrebare

În Netezirea Exponențială Simplă cu $\alpha = 0.9$, ce se întâmplă?

A. Prognozele sunt foarte netede și stabile
B. Observațiile recente au foarte puțină pondere
C. Prognozele reacționează rapid la schimbările recente
D. Prognoza este în esență o medie pe termen lung

Răspunsul pe slide-ul următor...

Test 3: Răspuns

Răspuns: C – Prognozele reacționează rapid la schimbările recente

Cu $\alpha = 0.9$: $\hat{X}_{t+1} = 0.9X_t + 0.1\hat{X}_t$

Aceasta înseamnă 90% pondere pe cea mai recentă observație. Valorile mari ale lui α fac prognozele foarte receptive la date noi. Valorile mici ale lui α (de exemplu, 0.1) produc prognoze mai netede, mai stabile, care mediază peste mai mult istoric.

Test 4: Staționaritate

Întrebare

Un proces de mers aleatoriu $X_t = X_{t-1} + \varepsilon_t$ este:

A. Strict staționar
B. Slab staționar
C. Nestaționar deoarece varianța crește cu timpul
D. Staționar după adăugarea unei constante

Răspunsul pe slide-ul următor...

Test 4: Răspuns

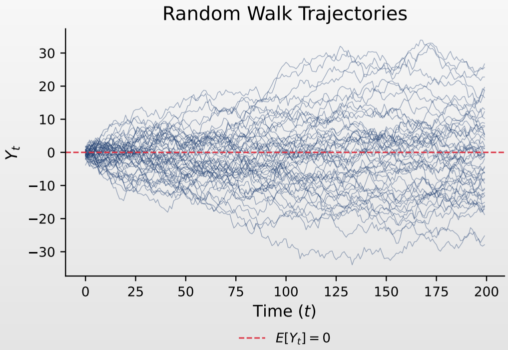
Răspuns: C – Nestaționar deoarece varianța crește cu timpul

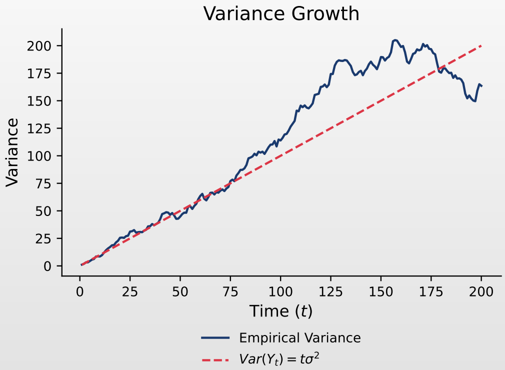
Pentru mersul aleatoriu: $X_t = \sum_{i=1}^{t} \varepsilon_i$

- $\mathbb{E}[X_t] = 0$ (medie constantă – OK)
- $\text{Var}(X_t) = t\sigma^2$ (varianța depinde de t – *nu* e acceptabil)

Deoarece varianța nu este constantă, procesul încalcă condiția de staționaritate. Soluție: **diferențierea** dă $\Delta X_t = \varepsilon_t$ care *este* staționară.

Vizual: Mers Aleatoriu vs Staționar





Traiectoriile mersului aleatoriu evoluează fără direcție determinată; varianța crește liniar cu timpul $\Rightarrow$ nestaționar.

TSA_ch1_random_walk

Test 5: Teste pentru Rădăcină Unitară

Întrebare

Rulați testele ADF și KPSS. ADF nu reușește să respingă H_0, iar KPSS respinge H_0. Ce concluzie trageți?

A. Seria este staționară
B. Seria are o rădăcină unitară (nestaționară)
C. Rezultatele sunt neconcludente
D. Trebuie să rulați mai multe teste

Răspunsul pe slide-ul următor...

Test 5: Răspuns

Răspuns: B – Seria are o rădăcină unitară (nestaționară)

- ADF: H_0 = rădăcină unitară. Nu respingem ⇒ evidență PENTRU rădăcină unitară
- KPSS: H_0 = staționară. Respingem ⇒ evidență ÎMPOTRIVA staționarității

Ambele teste sunt de acord: seria este **nestaționară**. Ar trebui să diferențiați seria înainte de a modela cu ARMA.

Test 6: Evaluarea Prognozei

Întrebare

Care metrică este cea mai potrivită pentru compararea acurateții prognozei între diferite serii de timp cu scale diferite?

A. Eroarea Absolută Medie (MAE)
B. Rădăcina Erorii Medii Pătratice (RMSE)
C. Eroarea Absolută Medie Procentuală (MAPE)
D. Eroarea Medie Pătratică (MSE)

Răspunsul pe slide-ul următor...

Test 6: Răspuns

Răspuns: C – Eroarea Absolută Medie Procentuală (MAPE)

$\text{MAPE} = \frac{100}{n} \sum \left| \frac{e_t}{X_t} \right|$ exprimă erorile ca **procente**.

- MAE, RMSE, MSE sunt **dependente de scală** (unități ale lui X_t)
- MAPE este **independentă de scală** (întotdeauna în %)
- Avertisment: MAPE devine instabilă când X_t este aproape de zero

Test 7: Tipuri de Trend

Întrebare

Un trend determinist poate fi eliminat prin:

A. Diferențiere
B. Regresie pe timp
C. Ajustare sezonieră
D. Netezire cu medie mobilă

Răspunsul pe slide-ul următor...

Test 7: Răspuns

Răspuns: B – Regresie pe timp

Trend determinist: $Y_t = \alpha + \beta t + \varepsilon_t$ unde β este fix.
Metoda de eliminare: Regresați Y_t pe t, apoi analizați reziduurile $\hat{\varepsilon}_t = Y_t - \hat{\alpha} - \hat{\beta} t$
De ce nu diferențiere? Diferențierea unui trend determinist dă: $\Delta Y_t = \beta + \Delta\varepsilon_t$, care elimină trendul dar lasă o constantă. Pentru trenduri *stochastice* (rădăcini unitare), diferențierea este corectă.

Test 8: Interpretarea ACF

Întrebare

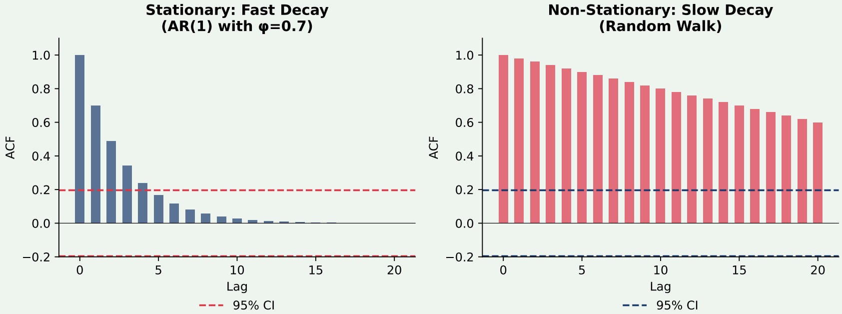
Dacă ACF-ul unei serii de timp descrește foarte lent (rămâne semnificativ pentru multe lag-uri), aceasta sugerează:

A. Seria este zgomot alb
B. Seria este probabil nestaționară
C. Seria nu are autocorelație
D. Seria este perfect predictibilă

Răspunsul pe slide-ul următor...

Test 8: Răspuns

Răspuns: B – Seria este probabil nestaționară



Staționară: ACF descrește rapid ($\rho_k = \phi^k \to 0$) **Nestaționară**: ACF rămâne aproape de 1 $\Rightarrow$ diferențiere necesară

TSA_ch1_acf_patterns

Test 9: Metoda Holt

Întrebare

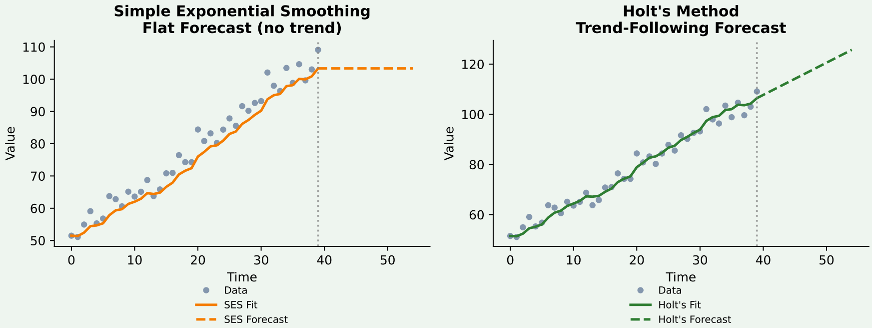
Netezirea exponențială Holt diferă de SES prin adăugarea:

A. O componentă sezonieră
B. O componentă de trend
C. O componentă ciclică
D. O componentă neregulată

Răspunsul pe slide-ul următor...

Test 9: Răspuns

Răspuns: B – O componentă de trend



Holt: $L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + b_{t-1})$; $b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$ **Prognoză**: $\hat{Y}_{t+h} = L_t + h \cdot b_t$

TSA_ch1_ts_basics

Test 10: Zgomot Alb

Întrebare

Care proprietate *nu* este necesară pentru ca un proces să fie zgomot alb?

A. $\mathbb{E}[\varepsilon_t] = 0$

B. $\text{Var}(\varepsilon_t) = \sigma^2$ (constantă)

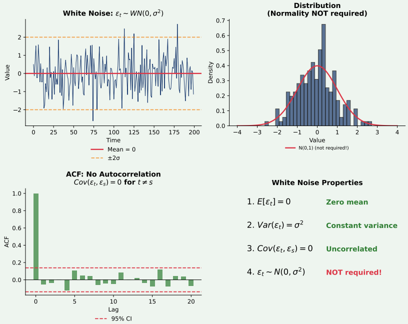
C. $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$ pentru $t \neq s$

D. $\varepsilon_t \sim N(0, \sigma^2)$

Răspunsul pe slide-ul următor...

Test 10: Răspuns

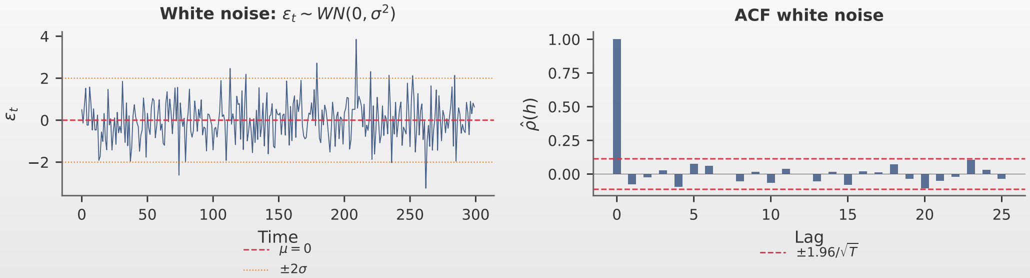
Răspuns: D – Normalitatea *nu* este necesară



Zgomot alb: Medie zero, varianță constantă, necorelat. **Zgomot alb Gaussian**: Adaugă normalitate $\Rightarrow$ de asemenea independent (nu doar necorelat)

TSA_ch1_white_noise

Vizual: Proprietățile Zgomotului Alb



Stânga: zgomotul alb fluctuează în jurul lui zero. Dreapta: ACF nu arată autocorelație (toate valorile aproape de zero după lag 0).

TSA_ch1_white_noise

Test 11: Orizont de Prognoză

Întrebare

Pe măsură ce orizontul de prognoză *h* crește, ce se întâmplă de obicei cu intervalele de prognoză?

A. Devin mai înguste

B. Rămân la aceeași lățime

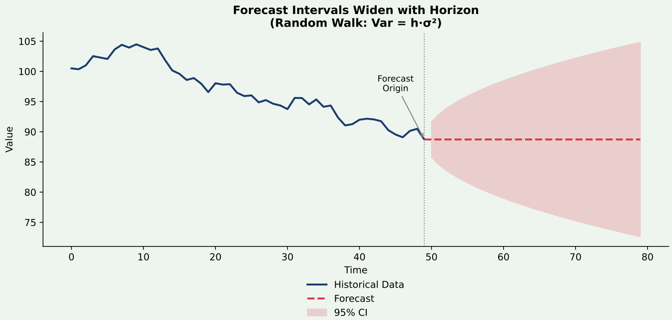
C. Devin mai largi

D. Dispar

Răspunsul pe slide-ul următor...

Test 11: Răspuns

Răspuns: C – Devin mai largi



Mers aleatoriu: Var $= h\sigma^2$ (crește liniar) **IC 95%**: $\hat{Y}_{t+h} \pm 1.96\sqrt{h}\sigma$ (se lărgește cu $\sqrt{h}$)

TSA_ch1_random_walk

Test 12: Detectarea Sezonalității

Întrebare

ACF-ul arată vârfuri semnificative la lag-urile 12, 24 și 36 pentru date lunare. Aceasta sugerează:

A. Fără sezonalitate
B. Sezonalitate anuală
C. Sezonalitate săptămânală
D. Zgomot aleatoriu

Răspunsul pe slide-ul următor...

Test 12: Răspuns

Răspuns: B – Sezonalitate anuală

Recunoașterea modelului:

- Lag 12: corelație cu aceeași lună de anul trecut
- Lag 24: corelație cu aceeași lună de acum doi ani
- Lag 36: corelație cu aceeași lună de acum trei ani

Perioada sezonieră: $s = 12$ (date lunare cu ciclu anual)
Modele comune: Vânzări retail (vârfuri în decembrie), consum de energie (vară/iarnă), date turistice

Test 13: Validare Încrucișată în Seriile de Timp

Întrebare

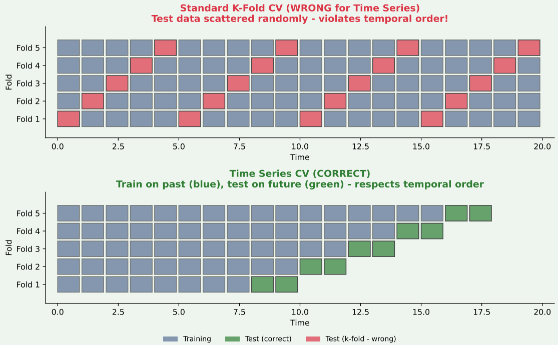
De ce nu putem folosi validarea încrucișată standard k-fold pentru seriile de timp?

A. Datele seriilor de timp sunt prea mici
B. Ar încălca ordonarea temporală (viitorul prezicând trecutul)
C. Validarea încrucișată este întotdeauna invalidă
D. Seriile de timp nu au nevoie de validare

Răspunsul pe slide-ul următor...

Test 13: Răspuns

Răspuns: B – Ar încălca ordonarea temporală



Regulă: Nu folosiți niciodată date viitoare pentru a prezice trecutul. Folosiți CV cu fereastră mobilă/în expansiune.

TSA_ch1_ts_basics

Test 14: Limitarea MAPE

Întrebare

MAPE (Eroarea Absolută Medie Procentuală) *nu* ar trebui folosită când:

A. Comparați modele pe același set de date
B. Valorile reale pot fi zero sau aproape de zero
C. Prognozați prețuri de acțiuni
D. Datele au un trend

Răspunsul pe slide-ul următor...

Test 14: Răspuns

Răspuns: B – Când valorile reale pot fi zero sau aproape de zero

Formula MAPE: $\text{MAPE} = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|$

Problema: Când $Y_t \approx 0$, împărțirea face MAPE $\to \infty$

Alternative:

- **SMAPE**: $\frac{200\%}{n} \sum \frac{|Y_t - \hat{Y}_t|}{|Y_t| + |\hat{Y}_t|}$ (mărginită 0–200%)
- **MASE**: $\frac{1}{n} \sum \frac{|e_t|}{\frac{1}{n-1} \sum |Y_t - Y_{t-1}|}$ (fără scală)

Adevărat sau Fals? (Setul 1)

Întrebare

Marcați fiecare afirmație ca Adevărat (A) sau Fals (F):

1. O serie de timp cu medie constantă este întotdeauna staționară. _______
2. Varianța unui mers aleatoriu crește liniar cu timpul. _______
3. Prognozele SES sunt întotdeauna plate (constante pentru toate orizonturile). _______
4. Testele ADF și KPSS au aceeași ipoteză nulă. _______
5. RMSE mai mic înseamnă întotdeauna prognoze mai bune. _______
6. Autocorelația la lag 0 este întotdeauna egală cu 1. _______

Răspunsul pe slide-ul următor...

Adevărat sau Fals: Răspunsuri (Setul 1)

Răspunsuri

1. O serie de timp cu medie constantă este întotdeauna staționară. **FALS**
   De asemenea, este nevoie de varianță constantă și covarianță care depinde doar de lag.
2. Varianța unui mers aleatoriu crește liniar cu timpul. **ADEVĂRAT**
   $\text{Var}(X_t) = t\sigma^2$ pentru mersul aleatoriu pornind de la X_0.
3. Prognozele SES sunt întotdeauna plate (constante pentru toate orizonturile). **ADEVĂRAT**
   SES nu are componentă de trend, deci $\hat{X}_{t+h} = L_t$ pentru orice h.
4. Testele ADF și KPSS au aceeași ipoteză nulă. **FALS**
   ADF: H_0 = rădăcină unitară. KPSS: H_0 = staționară. Ipoteze opuse.
5. RMSE mai mic înseamnă întotdeauna prognoze mai bune. **FALS**
   Depinde de context. RMSE este dependent de scală; poate supraajusta la valori extreme.
6. Autocorelația la lag 0 este întotdeauna egală cu 1. **ADEVĂRAT**
   $\rho(0) = \gamma(0)/\gamma(0) = 1$ prin definiție.

Adevărat sau Fals? (Setul 2)

Întrebare

Marcați fiecare afirmație ca Adevărat (A) sau Fals (F):

1. ACF-ul unui proces AR(1) staționar descrește exponențial. ______
2. Zgomotul alb este întotdeauna distribuit normal. ______
3. Diferențierea poate face o serie nestaționară să devină staționară. ______
4. PACF-ul unui proces MA(1) se anulează după lag 1. ______
5. Ar trebui să folosiți întotdeauna setul de test pentru ajustarea hiperparametrilor. ______
6. Holt-Winters este potrivit pentru date fără sezonalitate. ______

Răspunsul pe slide-ul următor...

Adevărat sau Fals: Răspunsuri (Setul 2)

Răspunsuri

1. ACF-ul unui AR(1) staționar descrește exponențial. **ADEVĂRAT**
   Pentru AR(1): $\rho(h) = \phi^h$, care descrește exponențial.
2. Zgomotul alb este întotdeauna distribuit normal. **FALS**
   Zgomotul alb necesită doar medie zero, varianță constantă, fără autocorelație. Zgomotul alb Gaussian este un caz special.
3. Diferențierea poate face o serie nestaționară să devină staționară. **ADEVĂRAT**
   Diferențierea elimină trendurile stochastice (rădăcinile unitare).
4. PACF-ul unui MA(1) se anulează după lag 1. **FALS**
   ACF-ul se anulează pentru MA. PACF descrește pentru procesele MA.
5. Ar trebui să folosiți întotdeauna setul de test pentru ajustarea hiperparametrilor. **FALS**
   Folosiți setul de validare pentru ajustare. Setul de test este doar pentru evaluarea finală.
6. Holt-Winters este potrivit pentru date fără sezonalitate. **FALS**
   Folosiți metoda Holt (fără componentă sezonieră) sau SES pentru date nesezoniere.

Exercițiu 1: Netezire Exponențială Simplă

Problemă: Date fiind următoarele date și $\alpha = 0.3$:

t	1	2	3	4	5
X_t	10	12	11	14	13

Începând cu $\hat{X}_1 = X_1 = 10$, calculați:

a) Prognozele $\hat{X}_2, \hat{X}_3, \hat{X}_4, \hat{X}_5$

b) Prognoza pentru $t = 6$: $\hat{X}_6$

c) Erorile de prognoză $e_t = X_t - \hat{X}_t$ pentru $t = 2, 3, 4, 5$

d) MAE și RMSE

Formula: $\hat{X}_{t+1} = \alpha X_t + (1 - \alpha)\hat{X}_t$

Exercițiu 1: Soluție

Folosind $\hat{X}_{t+1} = 0.3X_t + 0.7\hat{X}_t$:

t	1	2	3	4	5	6
X_t	10	12	11	14	13	?
$\hat{X}_t$	10	10	10.6	10.72	11.70	**12.09**
e_t	–	2	0.4	3.28	1.30	–

Calcule:

- $\hat{X}_2 = 0.3(10) + 0.7(10) = 10$
- $\hat{X}_3 = 0.3(12) + 0.7(10) = 10.6$
- $\hat{X}_4 = 0.3(11) + 0.7(10.6) = 10.72$
- $\hat{X}_5 = 0.3(14) + 0.7(10.72) = 11.70$
- $\hat{X}_6 = 0.3(13) + 0.7(11.70) = \mathbf{12.09}$

$\textbf{MAE} = \frac{|2|+|0.4|+|3.28|+|1.30|}{4} = 1.745$ $\quad$ $\textbf{RMSE} = \sqrt{\frac{4+0.16+10.76+1.69}{4}} = 2.04$

Exercițiu 2: Autocovarianță

Problemă: Pentru un proces staționar cu:

- $\mathbb{E}[X_t] = 5$
- $\gamma(0) = 4$ (varianță)
- $\gamma(1) = 2$
- $\gamma(2) = 1$

Calculați:

a) Funcția de autocorelație $\rho(0), \rho(1), \rho(2)$

b) $\mathrm{Cov}(X_t, X_{t-1})$

c) $\mathrm{Corr}(X_5, X_7)$

d) Dacă $X_t = 6$, care este $\mathbb{E}[X_{t+1}|X_t = 6]$ presupunând AR(1)?

Exercițiu 2: Soluție

a) Autocorelații:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

- $\rho(0) = \gamma(0)/\gamma(0) = 1$
- $\rho(1) = \gamma(1)/\gamma(0) = 2/4 = 0.5$
- $\rho(2) = \gamma(2)/\gamma(0) = 1/4 = 0.25$

b) $\text{Cov}(X_t, X_{t-1}) = \gamma(1) = 2$ (prin staționaritate, covarianța la lag 1)

c) $\text{Corr}(X_5, X_7) = \rho(|7-5|) = \rho(2) = 0.25$

d) Pentru AR(1) cu $\phi = \rho(1) = 0.5$:

$$\mathbb{E}[X_{t+1}|X_t] = \mu + \phi(X_t - \mu) = 5 + 0.5(6-5) = 5.5$$

TSA_ch1_ex1_autocovariance

Exercițiu 3: Proprietățile Mersului Aleatoriu

Problemă: Considerați un mers aleatoriu $X_t = X_{t-1} + \varepsilon_t$ unde $\varepsilon_t \sim WN(0, 4)$ și $X_0 = 100$.

Calculați:

a) $\mathbb{E}[X_{10}]$

b) $\mathrm{Var}(X_{10})$

c) $\mathrm{Cov}(X_5, X_{10})$

d) Intervalul de încredere de 95% pentru X_{100}

e) După observarea $X_5 = 108$, care este cea mai bună prognoză pentru X_6?

Exercițiu 3: Soluție

Mers aleatoriu: $X_t = X_0 + \sum_{i=1}^{t} \varepsilon_i$ cu $\sigma^2 = 4$

a) $\mathbb{E}[X_{10}] = X_0 = 100$ (media rămâne la valoarea de pornire)

b) $\text{Var}(X_{10}) = 10 \times \sigma^2 = 10 \times 4 = 40$

c) $\text{Cov}(X_5, X_{10}) = \min(5, 10) \times \sigma^2 = 5 \times 4 = 20$

d) Pentru X_{100}:

- $\mathbb{E}[X_{100}] = 100$, $\text{Var}(X_{100}) = 400$, $SD = 20$
- IC 95%: $100 \pm 1.96 \times 20 = [60.8, 139.2]$

e) Cea mai bună prognoză: $\hat{X}_6 = X_5 = 108$
(Mers aleatoriu: cea mai bună prognoză este ultima valoare observată)

TSA_ch1_ex2_random_walk

Exercițiu Python 1: Încărcare și Reprezentare Grafică

Sarcină: Încărcați datele S&P 500 și creați un grafic de bază pentru seria de timp.

Cod de Pornire

```
import yfinance as yf
import matplotlib.pyplot as plt

# Descărcați datele S&P 500
sp500 = yf.download('^GSPC', start='2020-01-01', end='2025-01-01')

# TODO: Reprezentați grafic prețurile de închidere
# TODO: Adăugați titlu, etichete și grilă
# TODO: Calculați și afișați statistici de bază
```

Întrebări:

1. Care este media și deviația standard a randamentelor?
2. Seria pare staționară? De ce sau de ce nu?

Exercițiu Python 2: Descompunere

Sarcină: Efectuați descompunerea STL pe datele pasagerilor aerieni.

Cod de Pornire

```
from statsmodels.tsa.seasonal import STL
import pandas as pd

# Încărcați pasagerii aerieni
url = 'https://raw.githubusercontent.com/..../airline.csv'
airline = pd.read_csv(url, parse_dates=['Month'],
                      index_col='Month')

# TODO: Aplicați descompunerea STL cu period=12
# TODO: Reprezentați grafic toate componentele
# TODO: Ce procent din varianță este explicat de trend?
```

Indiciu: Folosiți `STL(data, period=12).fit()`

Exercițiu Python 3: Netezire Exponențială

Sarcină: Comparați SES, Holt și Holt-Winters pe date reale.

Cod de Pornire

```
from statsmodels.tsa.holtwinters import (SimpleExpSmoothing,
    ExponentialSmoothing)

# Împărțiți datele: 80% antrenare, 20% test
train = airline[:'1958']
test = airline['1959':]

# TODO: Ajustați SES, Holt și Holt-Winters
# TODO: Generați prognoze pentru perioada de test
# TODO: Calculați RMSE pentru fiecare metodă
# TODO: Care metodă are cele mai bune performanțe? De ce?
```

Exercițiu Python 4: Testarea Staționarității

Sarcină: Testați staționaritatea folosind testele ADF și KPSS.

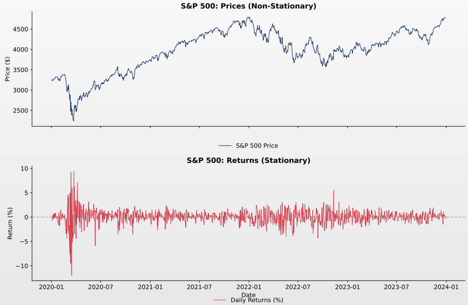
Cod de Pornire

```
from statsmodels.tsa.stattools import adfuller, kpss

# Testați prețurile S&P 500
prices = sp500['Close']
returns = prices.pct_change().dropna()

# TODO: Rulați testul ADF pe prețuri și randamente
# TODO: Rulați testul KPSS pe prețuri și randamente
# TODO: Interpretați rezultatele

# ADF: adfuller(series)
# KPSS: kpss(series, regression='c')
```

Întrebări:

1. Prețurile sunt staționare? Randamentele sunt staționare?
2. ADF și KPSS sunt de acord?

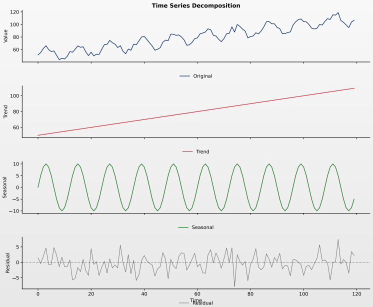
Studiu de Caz: Indicele S&P 500



- **Sus**: Nivelul prețului S&P 500 – trend ascendent clar (nestaționar)
- **Jos**: Randamente zilnice $r_t = \log(P_t/P_{t-1})$ – staționare
- Randamentele fluctuează în jurul mediei zero fără trend
- Gruparea volatilității (*volatility clustering*) este vizibilă – perioade de volatilitate ridicată/scăzută

TSA_ch1_ts_basics

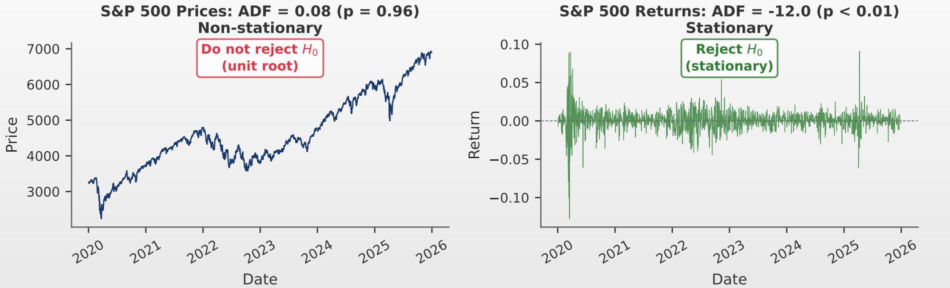
Descompunerea Seriilor de Timp: Exemplu Real



- **Trend**: Direcția pe termen lung a seriei
- **Sezonalitate**: Modele periodice regulate
- **Reziduu**: Ce rămâne după eliminarea trendului și sezonalității
- Descompunerea ajută la înțelegerea structurii datelor înainte de modelare

TSA_ch1_ts_basics

Testarea Staționarității: Rezultate ADF



- Testul ADF compară statistica de test cu valorile critice
- Dacă statistica de test $<$ valoarea critică $\Rightarrow$ respingem rădăcina unitară (seria este staționară)
- Prețuri: Statistica ADF $> -2.86 \Rightarrow$ nestaționară
- Randamente: Statistica ADF $< -2.86 \Rightarrow$ staționară

TSA_ch1_unit_root_tests

Comparație Staționaritate: Prețuri vs Randamente

Rezultate Test ADF

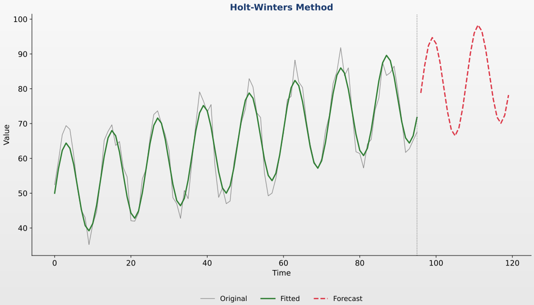
Serie	Statistică ADF	valoare-p	Concluzie
Prețuri S&P 500	-0.82	0.812	Nestaționară
Randamente S&P 500	-45.3	< 0.001	Staționară

Observație Cheie

Prețurile financiare sunt de obicei $I(1)$ – integrate de ordinul 1.
Luând diferențe de ordinul întâi (randamente) se obține staționaritatea.
De aceea modelăm **randamentele**, nu prețurile.

TSA_ch1_unit_root_tests

Prognoză cu Netezire Exponențială



- Metoda Holt-Winters pentru date cu trend și sezonalitate
- Parametrii de netezire α, β, γ controlează adaptabilitatea
- Prognozele captează atât continuarea trendului cât și tiparul sezonier
- Simplă dar eficientă pentru multe aplicații de afaceri

TSA_ch1_ts_basics

Întrebare de Discuție 1

Scenariu

Analizați date lunare de vânzări pentru o companie de retail. Datele arată sezonalitate clară (vânzări ridicate în decembrie) și un trend ascendent. Vârfurile sezoniere au devenit mai mari în timp.

Discutați:

1. Ar trebui să folosiți descompunere aditivă sau multiplicativă? De ce?
2. Ce metodă de netezire exponențială ați recomanda?
3. Cum ați evalua modelul de prognoză?
4. Ce ar putea merge prost dacă ați folosi descompunerea greșită?

Întrebare de Discuție 2

Scenariu

Un coleg afirmă: „Am rulat testul ADF pe datele mele de prețuri de acțiuni și am obținut o valoare-p de 0.65, deci datele mele sunt staționare și pot ajusta direct un model ARMA."

Discutați:

1. Ce este greșit în această interpretare?
2. Ce testează de fapt ipotezele ADF?
3. Ce ar trebui să facă colegul înainte de a ajusta un model ARMA?
4. Cum ar putea ajuta testul KPSS să clarifice situația?

Întrebare de Discuție 3

Scenariu

Construiți un model de prognoză și obțineți rezultate excelente: MAPE de 2% pe setul dumneavoastră de date. Managerul este impresionat și vrea să implementeze modelul imediat.

Discutați:

1. Ce întrebări ar trebui să puneți înainte de implementare?
2. Ați folosit împărțiri corecte antrenare/validare/test?
3. Ar putea exista scurgere de date în evaluare?
4. Ce diagnostice suplimentare ați rula?
5. Cum ați monitoriza modelul în producție?

Întrebare de Discuție 4

Scenariu

Trebuie să prognozați cererea zilnică de electricitate pentru săptămâna următoare. Datele arată: (1) modele zilnice puternice (vârfuri la ora 18), (2) modele săptămânale (mai scăzut în weekend), și (3) modele anuale (mai ridicat vara/iarna).

Discutați:

1. Cum ați gestiona modelele sezoniere multiple?
2. Ar funcționa Holt-Winters aici? De ce sau de ce nu?
3. Care este avantajul termenilor Fourier în acest caz?
4. Cum ați configura împărțirea antrenare/validare/test?

Exercițiu AI: Gândire critică

Prompt de testat în ChatGPT / Claude / Copilot

"Descarcă de pe FRED indicele lunar al prețurilor de consum din SUA (seria CPIAUCSL) din 2010-01 până în 2024-12 (180 observații). Calculează rata inflației an-la-an (%). Reprezintă grafic seria originală și rata inflației. Calculează statisticile descriptive (medie, deviație standard, asimetrie, kurtoză). Aplică descompunerea aditivă și identifică componenta de trend și sezonalitatea. Construiește o prognoză pe 12 luni folosind media mobilă simplă (SMA cu fereastră de 12 luni). Vreau cod Python complet cu grafice."

Exercițiu:

1. Rulați prompt-ul într-un LLM la alegere și analizați critic răspunsul.
2. Calculează corect rata inflației an-la-an? (diferență procentuală la lag 12, nu la lag 1)
3. Statisticile descriptive includ și autocorelația? Seria e tratată ca i.i.d.?
4. Descompunerea aleasă (aditivă) e potrivită, sau multiplicativă ar fi mai bună?
5. Prognoza SMA e realistă? Captează punctele de cotitură din inflație?

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

Concluzii Cheie de Astăzi

1. **Seriile de timp sunt dependente** – nu i.i.d. ca datele transversale
2. **Alegeți descompunerea cu înțelepciune** – multiplicativă când amplitudinea sezonieră crește
3. **Înțelegeți parametrii de netezire** – α mare = reactiv, α mic = neted
4. **Testați staționaritatea** – folosiți atât ADF cât și KPSS împreună
5. **Evaluare corectă** – nu ajustați niciodată pe setul de test.
6. **Mersul aleatoriu este nestaționar** – varianța crește cu timpul

Următorul Seminar

Identificarea, estimarea și prognoza modelelor ARMA/ARIMA

Referințe

- Hyndman, R.J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. 3rd ed., OTexts. https://otexts.com/fpp3/
- Box, G.E.P., Jenkins, G.M., Reinsel, G.C., & Ljung, G.M. (2015). *Time Series Analysis: Forecasting and Control*. 5th ed., Wiley.
- Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.
- Brockwell, P.J., & Davis, R.A. (2016). *Introduction to Time Series and Forecasting*. 3rd ed., Springer.

Surse de Date și Software

Instrumente Software:

- `statsmodels` – Modele statistice pentru Python
- `pandas` – Manipulare date și serii de timp
- `matplotlib`, `seaborn` – Vizualizare
- `scipy` – Funcții statistice

Date și Exemple:

- Serii de timp simulate pentru ilustrații
- Exemple bazate pe Hyndman & Athanasopoulos (2021)

Exercițiu Monte Carlo: Puterea Testului ADF

Obiectiv

Evaluați prin simulare Monte Carlo puterea testului ADF pentru diferite valori ale lui ϕ și T.

1. Generați $B = 5000$ serii AR(1): $y_t = \phi y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$
2. Variați: $\phi \in \{0{,}90;\ 0{,}95;\ 0{,}97;\ 0{,}99;\ 1{,}00\}$ și $T \in \{50, 100, 200, 500\}$
3. Pentru fiecare combinație, aplicați testul ADF și înregistrați rata de respingere la $\alpha = 5\%$
4. Construiți un tabel 5×4 cu ratele de respingere empirice
5. **Analizați**: (a) Pentru $\phi = 1$: rata $\approx 5\%$? (dimensiune corectă?)
   (b) Pentru $\phi = 0{,}95$, $T = 100$: cât este puterea? E suficientă?
   (c) Cum crește puterea cu T? Dar cu distanța $|1 - \phi|$?

Cod: `from statsmodels.tsa.stattools import adfuller; np.random.seed(42)`

TSA_ch1_unit_root_tests

Vă mulțumesc!

Întrebări?

Succes la exerciții!