Modele Autoregresive Liniare

2.1. Noțiuni specifice Analizei Seriilor de Timp Univariate

O caracteristică intrinsecă a unei serii de timp este aceea că **observațiile adiacente sunt dependente**.
Analiza Seriilor de Timp constă în tehnici de analiză a acestei dependențe.
Principalul instrument în analiza seriilor de timp este modelul stochastic. Se caută şi se foloseşte un model stochastic de un anumit tip, capabil să reproducă comportamentul trecut al unei serii, pe baza structurii sale de autocorelație.

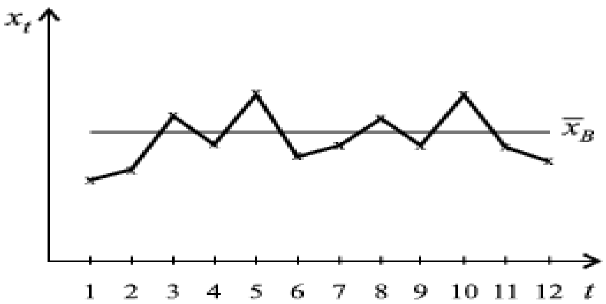
În modelarea seriilor de timp economice trebuie să se țină cont de următoarele caracteristici:
♦ **Existența unei tendințe generale**, care are un rol important în efectuarea prognozelor.
♦ **Prezența autodeterminării**. Autodeterminarea trebuie înțelesă astfel:
  - *Există un anumit grad de dependență a valorii curente, de valorile anterioare.*
  - *Există acțiuni compensatorii în vederea contracarării unor factori din afara sistemului*

♦ **Posibilitatea de descriere a comportamentului** seriei prin funcția sa de autocorelație (ACF).

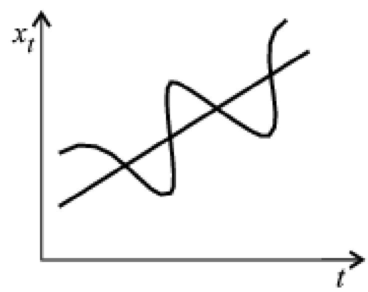
Serii de timp staționare și nestaționare

Este necesar ca ceea ce observăm cu privire la o serie de date să fie stabil, aşa încât să putem face afirmații despre viitor.
Seriile staționare sunt caracterizate prin oscilații relativ regulate în jurul unui nivel de referință. Seriile staționare nu prezintă trenduri vizibile. Prin contrast, seriile nestaționare prezintă trenduri vizibile.
O serie de timp este staționară dacă:
- Are media constantă de-a lungul timpului (nu prezintă trend)
- Are dispersie constantă de-a lungul timpului
- Corelația de-a lungul timpului depinde de legătura dintre y_t şi y_{t-k} (y la lagul k) şi nu depinde de alte variabile.

Cum recunoaştem o serie staționară?
- Analiza grafică
- Evoluția funcției de autocorelație (ACF) şi a funcției de autocorelație parțială (PACF)
- Testul Dickey-Fuller



Serie staționară



Serie nestaționară cu trend liniar

Un proces stochastic poate fi descris sau caracterizat prin momentele sale de ordinele 1 şi 2:
Media: $E(y_t) = \mu(t)$; Varianța: $Var(y_t) = E(y_t - \mu)^2 = \sigma^2(t)$; Autocovarianța: $\text{cov}(y_{t_1}, y_{t_2}) = \gamma(t_1, t_2)$.

Principalele concepte pe care se fundamentează metodologia Box-Jenkins

a) Staționaritatea strictă
Un proces stochastic este **strict staționar** dacă procesul are un comportament care nu se modifică în timp.
Prin **Lag** se înțelege **decalajul** sau **întârzierea**, exprimate în unități de timp, între modificarea variabilei cauzale şi manifestarea efectului.
Observație: Condiția de staționaritate strictă este o condiție greu de verificat empiric.
În locul staționarității stricte este utilizată mai frecvent o definiție mai puțin restrictivă a staționarității.
b) Staționaritatea slabă (Staționaritatea în covarianță)

Un **proces stochastic este slab staționar** dacă are media şi dispersia constante în timp iar covarianța sa între două perioade de timp depinde numai de decalajul dintre cele două perioade şi nu de momentul de timp la care este calculată. Aceasta înseamnă că procesul rămâne în echilibru în jurul unui nivel mediu constant.

Seria de timp y_t este slab staționară (staționară în covarianță, staționară de ordinul 2) dacă:

Media este constantă: $E(y_t) = \mu$;

Varianța este constantă: $Var(y_t) = E(y_t - \mu)^2 = \sigma^2$;

Covarianța: $\text{cov}(y_t, y_{t+k}) = E[(y_t - \mu)(y_{t+k} - \mu)] = \gamma_k$. $Cov(y_s, y_t)$ depinde numai de decalajul $t - s$.

O serie de timp care nu este staționară în sensul definiției de mai sus, se numeşte serie **nestaționară**.

Analiza de regresie bazată pe serii cronologice presupune că seriile de timp sunt staționare. Testele clasice ***t*** şi **F** sunt bazate pe această presupunere.

Seriile nestaționare sunt caracterizate de faptul că prezintă o tendință de creştere sau de scădere, odată cu trecerea timpului. Avem:

• **Serie nestaționară omogenă** sau **serie nestaționară în medie**. Media seriei nu este constantă în timp ci are o traiectorie liniară cu panta pozitivă sau negativă.

• **Serie nestaționară neomogenă** sau **serie nestaționară în varianță**. Media şi varianța seriei nu sunt constante în timp.

Nestaționaritate deterministă şi nestaționaritate stochastică

În practică, cele mai multe serii de timp economice sunt nestaționare.

a) Serie nestaționară de tip TS (Trend Stationarity). Reprezintă o serie nestaționară de tip determinist.

O serie nestaționară este de tip TS dacă Trendul este de tip determinist şi se elimină prin scădere $(y_t - y_{Tt})$ sau prin împărțire (y_t / y_{Tt}) .

Notă: *Efectul produs de un şoc aleator este tranzitoriu*, deci nu se acumulează în timp. *Staționarizarea unei serii de tip TS se realizează prin estimarea trendului şi eliminarea lui din seria inițială.*

b) Serie nestaționară de tip DS (Differency Stationarity).

Trendul este de tip stochastic şi se elimină prin calculul diferențelor de ordinul 1 sau de ordin ≥ 2 .

$\Delta y_t = y_t - y_{t-1}$, $\Delta^2 y_t = \Delta(y_t - y_{t-1}) = \Delta y_t - \Delta y_{t-1}) = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$.

O serie de tip DS se prezintă sub forma $y_t = \rho\, y_{t-1} + \beta + \varepsilon_t$.

Dacă $\rho = 1$ spunem că seria y_t are **o rădăcină unitate**.

Atunci: $y_t = y_{t-1} + \beta + \varepsilon_t = \beta + (\beta + y_{t-1} + \varepsilon_{t-1}) + \varepsilon_t = \cdots = y_0 + \beta t + \sum_{j=1}^{t} \varepsilon_j$.

Proprietăți: 1) Media seriei este variabilă în timp, deoarece $E(y_t) = y_0 + \beta t$.

2) Dispersia depinde de timp: $Var(y_t) = t\,\sigma_\varepsilon^2$.

3) Covarianța nu este nulă: $\text{cov}(y_t, y_{t+h}) = h\,\sigma_\varepsilon^2$.

Notă: Un şoc la un moment dat nu are efect tranzitoriu, ci influențează la infinit valorile seriei dinamice. Staționarizarea unei serii de tip DS se realizează prin aplicarea operatorului de diferență.

Exemplu de proces stochastic nestaționar. Mersul aleator sau **la întâmplare (Random Walk)**

O serie de timp y_t este un mers la întâmplare dacă satisface relația: $y_t = y_{t-1} + \varepsilon_t$, unde y_0 este un număr real care indică valoarea de start a procesului şi ε_t este zgomot alb. Procesul Random Walk este o serie staționară în prima diferență deoarece prima diferență a procesului $y_t - y_{t-1} = \varepsilon_t$ este un proces staționar. Este o serie I(1).

<u>Funcția de autocovarianță.</u>

Pentru un proces staționar avem $E(y_t) = \mu$ pentru orice t.

Definim $\text{cov}(y_t, y_{t+k}) = E(y_t y_{t+k}) - E(y_t)E(y_{t+k}) = \gamma_k$, **autocovarianța la decalajul sau lag-ul k**.

Dacă $k = 0$ se obține $\gamma_0 = Var(y_t)$, deci varianța lui Y, iar dacă $k = 1$ se obține $\gamma_1 = \text{cov}(y_t, y_{t+1}) = \text{cov}(y_t, y_{t-1})$, autocovarianța dintre două valori adiacente ale lui Y, adică la distanță de un lag.

Funcția de autocovarianță a unui proces staționar este şirul autocovarianțelor de lag *k*, $\{\gamma_k\}$, unde $k = 0, \pm1, \pm2, \pm3, \ldots$.

$\gamma_k = \gamma_{-k}$ pentru toate valorile lui *k*. O serie de timp este numită *ergodică* dacă $\gamma_k \to 0$ când $k \to \infty$.

<u>Funcția de autocorelație (ACF)</u>

Considerăm o serie de timp staționară, y_t. Atunci când dorim să determinăm corelația dintre y_t și valorile sale din trecut y_{t-k}, vorbim despre coeficientul de autocorelație.

Pentru o serie staționară, definim **autocorelația la lag-ul k**: $\rho_k = \gamma_k / \gamma_0$.

Funcția de autocorelație (ACF) a unui proces y_t, $t \in Z$, este un șir $\rho_k = \rho(y_t, y_{t-k})$, $k = 0, \pm1, \pm2, \pm3, \dots$ care verifică relațiile: $\rho_0 = 1$ și $-1 \le \rho_k \le 1$. Autocorelațiile sunt funcții pare, deci $\rho_k = \rho_{-k}$.

O serie staționară nu prezintă autocorelații dacă $\rho_k = 0$ pentru toți $k > 0$.

Pentru un proces pur nedeterminist ne așteptăm ca $\rho_k \to 0$ atunci când $k \to \infty$.

Coeficienții de autocorelație măsoară corelația dintre observații aflate la o anumită distanță unele de altele. Ei pot descrie proprietățile dinamice ale unui proces stochastic staționar.

Corelograma este graficul funcției de autocorelație (ACF) în raport cu decalajul *k*, și oferă numeroase informații privind comportamentul seriei de timp. Proprietățile dinamice ale fiecărui model staționar determină o anumită formă a corelogramei (un anumit tipar).

Interpretarea corelogramei Analiza coeficienților de autocorelație arată că o serie de timp ar putea fi:

a) **Serie aleatoare**.

b) **Serie ce prezintă corelație pe termen scurt**.

c) **Serie alternantă**.

d) **Serie nestaționară**.

e) **Serie cu variații de sezonalitate**. Într-un astfel de caz, corelograma are aceeași formă sinusoidală ca și seria analizată. Mai întâi, trebuie eliminată sezonalitatea din seria de date inițiale.

f) **Serie ce conține** una sau mai multe **valori extreme (outliers)**. Se recomandă ca, chiar de la început, aceste observații extreme să fie ajustate.

Exemplu de proces staționar.

Procesul Zgomot Alb (White Noise) este un proces ce descrie o variabilă aleatoare cu media zero și varianța constantă, dar fără o structură anume. Dacă ε_t este un șir de variabile aleatoare necorelate (independente), fiecare cu media zero și varianța constantă finită, atunci ε_t este staționar, fiind satisfăcute condițiile de staționaritate, adică:

$E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma_\varepsilon^2$, $\mathrm{cov}(\varepsilon_t, \varepsilon_s) = 0$, $t \ne s$.

Funcția de autocovarianță și Funcția de autocorelație sunt:

$$\gamma_k = \begin{cases} \sigma_\varepsilon^2, & k = 0 \\ 0, & k \ne 0 \end{cases} \quad \text{și} \quad \rho_k = \begin{cases} 1, & k = 0 \\ 0, & k \ne 0 \end{cases}$$

Un astfel de șir de variabile aleatoare necorelate este referit ca proces de zgomot alb sau **proces aleator complet** și pentru el se folosește notația $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$.

<u>Funcția de autocorelație de selecție (SACF)</u> **Estimarea ACF.**

Deoarece în aplicațiile practice dispunem doar de o realizare a unui proces stochastic, se pot calcula doar coeficienții de autocorelație **de selecție** de lag *k*, adică $r_k = \hat{\rho}_k = \hat{\gamma}_k / \hat{\gamma}_0$.

Funcția de autocorelație de selecție (**SACF**) este mulțimea $\{r_k = \hat{\rho}_k = \hat{\gamma}_k / \hat{\gamma}_0 \mid k \in Z\}$.

Graficul coeficienților r_k în raport cu k este numit **corelogramă**.

Autocorelația de selecție la decalajul 1 (lag-ul 1) se calculează cu ajutorul relației:

$$r_1 = \hat{\rho}_1 = \frac{\sum_{t=2}^{n} (y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=1}^{n} (y_t - \bar{y})^2}.$$

Definim **autocorelația de selecție** la lag-ul *k*, a lui y_t, prin relația:

$$r_k = \hat{\rho}_k = \frac{\sum_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n} (y_t - \bar{y})^2}, \text{ unde } k = 0, \dots, n-1.$$

Observație: Pentru a identifica dacă o serie este staționară, se calculează coeficienții de autocorelație. Dacă în cadrul seriei coeficienților de autocorelație estimați există o secvență de valori ce diferă semnificativ de zero sau **valorile sunt sensibil egale**, atunci **seria de date este nestaționară**.

Exemplu: Pentru a ilustra calcularea unei funcții SACF, considerăm 10 valori ale unei serii de timp y_t:

13, 8, 15, 4, 4, 12, 11, 7, 14, 12. Să se calculeze autocorelațiile de selecție pentru lag-urile 1, 2 și 3.

Calculăm media de selecție și obținem $\bar{y} = 10$. Folosim formulele

$$r_1 = \hat{\rho}_1 = \frac{\sum_{t=2}^{n}(y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \quad , \quad r_k = \hat{\rho}_k = \frac{\sum_{t=k+1}^{n}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$$

$$\hat{\rho}_1 = \frac{(13-10)(8-10)+(8-10)(15-10)+\cdots+(7-10)(14-10)+(14-10)(12-10)}{(13-10)^2+(8-10)^2+\cdots+(12-10)^2} = \frac{-27}{144} = -0{,}188$$

$$\hat{\rho}_2 = \frac{(13-10)(15-10)+(8-10)(4-10)+\cdots+(11-10)(14-10)+(7-10)(12-10)}{(13-10)^2+(8-10)^2+\cdots+(12-10)^2} = \frac{-29}{144} = -0{,}201$$

$$\hat{\rho}_3 = \frac{26}{144} = 0{,}181$$

Funcția de autocorelație parțială (PACF)

Corelația parțială dintre y_t și y_{t-k} măsoară corelația dintre y_t și y_{t-k} fără a ține seama de corelațiile dintre y_t și $y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}$. Pentru a măsura efectul direct al lui y_{t-k} asupra lui y_t se folosește coeficientul de corelație parțială.

Coeficientul de autocorelație parțială de ordinul k este coeficientul ϕ_{kk}, din modelul de regresie $y_t = \phi_{k1} y_{t-1} + \phi_{k2} y_{t-2} + \cdots + \phi_{kk} y_{t-k} + \varepsilon_t$.

Autocorelația parțială de lag k, între două variabile separate de ***k*** unități de timp, este coeficientul ϕ_{kk}, al lui y_{t-k}, din modelul de regresie $y_t = \phi_{k1} y_{t-1} + \phi_{k2} y_{t-2} + \cdots + \phi_{kk} y_{t-k} + \varepsilon_t$.

Interpretarea lui ϕ_{kk}: Măsoară corelația dintre y_t și y_{t-k} în condițiile în care variabilele $y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}$ sunt menținute constante (se izolează influența variabilei y_{t-k}). Arată cu câte unități se modifică y_t, în medie, atunci când y_{t-k} crește cu o unitate iar celelalte variabile $y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}$ rămân nemodificate.

Funcția de autocorelație parțială (PACF) este mulțimea coeficienților $\{\phi_{11}, \phi_{22}, \ldots\} = \{\phi_{kk}, k \geq 1\}$. *Observație*: În practică (inclusiv în softurile de statistică) coeficienții de autocorelație parțială nu se determină prin modele de regresie ci se folosesc ecuațiile Yule-Walker, ce redau relațiile dintre coeficienții de autocorelație și coeficienții de autocorelație parțială.

Ecuațiile Yule-Walker exprimă coeficienții de autocorelație ρ_k sub forma unor funcții de coeficienții autoregresivi ϕ_i. Definim matricea coeficienților liniari de autocorelație prin $R(p)$, vectorul parametrilor modelului prin ϕ și vectorul coeficienților de autocorelație prin ρ.

Sistemul de ecuații Yule-Walker poate fi scris în formă matriceală ca $R(p) \cdot \phi = \rho$

În general, ϕ_{kk} este obținut ca un raport de determinanți care implică $\rho_1, \rho_2 \ldots, \rho_k$.

Astfel, ϕ_{22} este obținut ca un raport de determinanți care implică ρ_1, ρ_2.

$$\begin{cases} \rho_1 = \rho_0 \phi_1 + \rho_1 \phi_2 \\ \rho_2 = \rho_1 \phi_1 + \rho_0 \phi_2 \end{cases} \Rightarrow \begin{cases} \phi_1 + \rho_1 \phi_2 = \rho_1 \\ \rho_1 \phi_1 + \phi_2 = \rho_2 \end{cases} \Rightarrow \phi_{22} = \begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix} \Big/ \begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \neq 0.$$

Coeficientul ϕ_{33} este obținut ca un raport de determinanți care implică ρ_1, ρ_2, ρ_3.

$$\begin{cases} \rho_1 = \rho_0 \phi_1 + \rho_1 \phi_2 + \rho_2 \phi_3 \\ \rho_2 = \rho_1 \phi_1 + \rho_0 \phi_2 + \rho_1 \phi_3 \\ \rho_3 = \rho_2 \phi_1 + \rho_1 \phi_2 + \rho_0 \phi_3 \end{cases} \Rightarrow \begin{cases} \phi_1 + \rho_1 \phi_2 + \rho_2 \phi_3 = \rho_1 \\ \rho_1 \phi_1 + \phi_2 + \rho_1 \phi_3 = \rho_2 \\ \rho_2 \phi_1 + \rho_1 \phi_2 + \phi_3 = \rho_3 \end{cases} \Rightarrow \phi_{33} = \begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix} \Bigg/ \begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}.$$

Estimarea PACF. Funcția de autocorelație parțială de selecție (SPACF)

Funcția de autocorelație parțială de selecție (SPACF) este mulțimea coeficienților $\{\hat{\phi}_{11}, \hat{\phi}_{22}, \ldots\} = \{\hat{\phi}_{kk}, k \geq 1\}$. SPACF are un rol foarte important în identificarea ordinului unui proces autoregresiv.

Coeficienții de autocorelație parțială de selecție sunt obținuți prin aceste formule, dar cu mențiunea că $\rho_1, \rho_2 \ldots, \rho_k$ sunt înlocuiți prin estimațiile lor $r_1, r_2 \ldots, r_k$.

$$\hat{\phi}_{11} = r_1 = \hat{\rho}_1, \ \hat{\phi}_{22} = \frac{r_2 - r_1^2}{1 - r_1^2} = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2}, \ \ldots$$

Testarea semnificației coeficienților de autocorelație

Bartlett a determinat următoarea expresie pentru varianța estimatorului coeficientului de autocorelație:

$\widehat{Var}(\widehat{\rho}_k) = \dfrac{1}{n}[1 + 2(\widehat{\rho}_1^2 + \widehat{\rho}_2^2 + \cdots + \widehat{\rho}_{k-1}^2)]$

Propoziție: Dacă seria $(y_1, y_2, ..., y_n)$ provine dintr-un proces zgomot alb gaussian, atunci $\widehat{\rho}_k$ are o distribuție normală cu media 0 şi varianța $Var(\widehat{\rho}_k) = 1/n$, adică $\widehat{\rho}_k \sim N(0, 1/n)$. Avem $se(\widehat{\rho}_k) = 1/\sqrt{n}$.

Consecințe:

– Un interval de încredere 95% pentru ρ_k are forma $(-1{,}96/\sqrt{n}\ ;\ +1{,}96/\sqrt{n})$ sau $(-2/\sqrt{n}\ ;\ +2/\sqrt{n})$.

– 95% din valorile coeficienților de autocorelație se află in intervalul $(-2/\sqrt{n}\ ;\ +2/\sqrt{n})$.

Testarea semnificației unui singur coeficient de autocorelație ρ_k **:**

$H_0 : \rho_k = 0$ (ρ_k nu este semnificativ statistic)

$H_1 : \rho_k \neq 0$ (ρ_k este semnificativ statistic).

În corelograme, intervalele de încredere 95% (ale coeficienților de autocorelație individuali) sunt reprezentate ca două benzi, bazate pe abaterea standard aproximată $\sigma_{\hat{\rho}_k} \approx 1/\sqrt{n}$.

O regiune de acceptare 95% pentru ipoteza nulă va fi dată prin intervalul $(-1{,}96/\sqrt{n}\ ;\ +1{,}96/\sqrt{n})$, pentru $k \neq 0$. Dacă valoarea coeficientului de autocorelație cade în afara acestui interval, pentru o valoare dată a lui *k*, atunci ipoteza nulă că $\rho_k = 0$ este respinsă. Se acceptă ipoteza alternativă: $\rho_k \neq 0$.

În selecțiile finite, $\hat{\rho}_k$ este un estimator deplasat al lui ρ_k. Deplasarea este mare în cazul selecțiilor de volum redus şi este fără importanță în cazul selecțiilor de volum mare.

Problema1: Considerăm o selecție de volum n=200 de observații asupra unui proces stochastic pentru care se cunosc $\rho_1 = -0{,}4$ şi $\rho_k = 0, k \geq 2$. Autocorelațiile estimate sunt:

k	1	2	3	4	5	6	7	8	9	10
r_k	-0,38	-0,08	0,11	0,08	0,02	0,00	0,00	0,00	0,07	-0,08

Să se testeze şi să se indice care coeficient de autocorelație este semnificativ diferit de zero.

Deoarece varianța estimată a lui r_k este $Var(r_k) = Var(\widehat{\rho}_k) = 1/n = 1/200 = 0{,}005$ rezultă că abaterea standard va fi $se(\widehat{\rho}_k) = 1/\sqrt{n} = 1/\sqrt{200} = \sqrt{0{,}005} = 0{,}07$. Un coeficient ρ_k este semnificativ diferit de zero dacă se află în afara intervalului $(-2/\sqrt{n}\ ;\ +2/\sqrt{n})$, în acest caz $(-0{,}14\ ;\ +0{,}14)$.

Dacă $\widehat{\rho}_k \in (-0{,}14\ ;\ +0{,}14)$ acceptăm ipoteza că ρ_k **nu este semnificativ** diferit de zero.

Deoarece avem $r_1 = \widehat{\rho}_1 = -0{,}38$, (în valoare absolută mai mare decât cinci abateri standard, nu două) înseamnă că $\widehat{\rho}_1 \notin (-0{,}14\ ;\ +0{,}14)$, deci acceptăm ipoteza că ρ_k **este semnificativ** diferit de zero ($\rho_1 \neq 0$). Pentru restul coeficienților acceptăm ipoteza că nu sunt semnificativi, $\rho_k = 0,\ \ k \geq 2$.

Problema2: Considerăm o selecție de volum n=50 de observații asupra unui proces stochastic. Autocorelațiile estimate sunt:

k	1	2	3	4	5	6	7	8	9	10	11	12
r_k	0,84	0,73	0,61	0,54	0,47	0,46	0,38	0,29	0,17	0,05	0,01	-0,01

a) Să se calculeze coeficienții de autocorelație parțială de selecție pentru lag-urile 1, 2 şi 3.

b) Să se testeze şi să se indice care coeficient de autocorelație parțială de selecție este semnificativ diferit de zero.

a) Folosim ecuațiile Yule-Walker

$$\widehat{\phi}_{11} = r_{11} = r_1 = \widehat{\rho}_1 = 0{,}084\ ;\ \widehat{\phi}_{22} = \frac{\begin{vmatrix} 1 & 0{,}84 \\ 0{,}84 & 0{,}73 \end{vmatrix}}{\begin{vmatrix} 1 & 0{,}84 \\ 0{,}84 & 1 \end{vmatrix}} = 0{,}08; \qquad \widehat{\phi}_{33} = \frac{\begin{vmatrix} 1 & 0{,}84 & 0{,}84 \\ 0{,}84 & 1 & 0{,}73 \\ 0{,}73 & 0{,}84 & 0{,}61 \end{vmatrix}}{\begin{vmatrix} 1 & 0{,}84 & 0{,}73 \\ 0{,}84 & 1 & 0{,}84 \\ 0{,}73 & 0{,}84 & 1 \end{vmatrix}} = -0{,}075$$

b) Deoarece varianța estimată a lui $\widehat{\phi}_{kk}$ este $Var(r_{kk}) = Var(\widehat{\phi}_{kk}) = 1/n =$ 1/50=0,02 rezultă că abaterea standard va fi $se(\widehat{\phi}_{kk}) = 1/\sqrt{n} = 1/\sqrt{50} = \sqrt{0{,}02} = 0{,}1414$. Un coeficient $\widehat{\phi}_{kk}$ este semnificativ diferit de zero dacă se află în afara intervalului $(-2/\sqrt{n}\ ;\ +2/\sqrt{n})$, în acest caz $(-0{,}2828\ ;\ +0{,}2828)$.

Dacă $\hat{\phi}_{kk} \in (-0{,}2828\ ;\ +0{,}2828)$ acceptăm ipoteza că ϕ_{kk} **nu este semnificativ** diferit de zero.

Deoarece avem $r_{11} = 0{,}84$, înseamnă că $r_{11} \notin (-0{,}2828\ ;\ +0{,}2828)$, deci acceptăm ipoteza că ϕ_{11} **este semnificativ** diferit de zero. Pentru coeficienții ϕ_{22} şi ϕ_{33} acceptăm ipoteza că nu sunt semnificativi statistic, deoarece $r_{22} = 0{,}08$ şi $r_{33} = -0{,}075$ aparțin intervalului $(-0{,}2828\ ;\ +0{,}2828)$.

<u>Testarea ipotezei că mai mulți coeficienți de autocorelație sunt zero</u>

În multe aplicații cu date financiare se testează dacă mai mulți coeficienți de autocorelație sunt simultan egali cu zero. Acesta este un test mai general pentru dependența liniară din seriile de timp.

Pentru a testa ipoteza că toți coeficienții de autocorelație sunt simultan nuli, se foloseşte statistica Q_{BP} (construită de Box şi Pierce), care are o distribuție χ^2 cu *m* grade de libertate, *m* fiind lungimea decalajului.

H_0 :toti $\rho_k = 0$ sau $\rho_1 = \rho_2 = \cdots = \rho_m = 0$ **(seria este staționară)**

H_1 :exista $\rho_k \neq 0$ **(seria este nestaționară)**

Se poate folosi statistica Ljung-Box $Q(m) = Q_{LB} = n(n+2)\sum_{k=1}^{m}\left(\frac{\widehat{\rho}_k^2}{n-k}\right) \sim \chi_m^2$ sau statistica $Q_{BP}(m) = n\sum_{k=1}^{m}\widehat{\rho}_k^2$

Dacă $Q_{LB} < \chi_{crt}^2 \Rightarrow$ acceptăm $H_0 \Rightarrow$ seria este staționară.

Dacă $Q_{LB} > \chi_{crt}^2 \Rightarrow$ respingem $H_0 \Rightarrow$ seria este nestaționară.

<u>Operatorul de decalaj (operatorul lag)</u>

Definim **operatorul de întârziere** sau **de decalaj**, sau **operatorul lag**, notat *L*, prin relația $Ly_t = y_{t-1}$.

Definim puterile operatorului lag prin $L^2 y_t = LLy_t = Ly_{t-1} = y_{t-2}$ şi, în general $L^k y_t = y_{t-k}$. Avem $L^0 \equiv 1$ şi $L^{-1} y_t = y_{t+1}$. Cu ajutorul acestor operatori putem scrie polinoame lag de diferite grade.

$\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ este numit polinomul autoregresiv de ordinul *p*.

$\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$ este numit polinomul de medie mobilă de ordin *q*.

<u>Serii integrate şi serii cointegrate</u>

Spunem că **seria** y_t **este integrată de ordinul 1**, sau **conține o rădăcină egală cu 1**, dacă y_t este nestaționară dar seria diferențelor Δy_t este staționară. O astfel de serie se notează prin $y_t \sim I(1)$. În general, pentru a deveni staționară, o serie nestaționară ar putea necesita diferențierea de mai multe ori.

O serie **este integrată de ordinul 2** dacă ea conține două rădăcini egale cu 1 şi astfel, este necesară diferențierea de două ori pentru a induce staționaritatea. Se notează prin $y_t \sim I(2)$.

O serie y_t **este integrată de ordinul d**, sau conține *d* rădăcini egale cu 1, dacă y_t este nestaționară dar seria $\Delta^d y_t$ este staționară. O astfel de serie se notează prin $y_t \sim I(d)$. Spunem că **seria y_t este integrată de ordinul 1**, sau **conține o rădăcină egală cu 1**, dacă $\mathbf{y_t}$ este nestaționară dar seria diferențelor $\mathbf{\Delta y_t}$ este staționară. O astfel de serie se notează prin $\mathbf{y_t} \sim$ I(1). În general, pentru a deveni staționară, o serie nestaționară ar putea necesita diferențierea de mai multe ori.

O serie **este integrată de ordinul 2** dacă ea conține două rădăcini egale cu 1 şi astfel, este necesară diferențierea de două ori pentru a induce staționaritatea. Se notează prin $\mathbf{y_t} \sim$ I(2).

O serie y_t este integrată de ordinul d, sau conține *d* rădăcini egale cu 1 şi se notează prin $\mathbf{y_t} \sim$ I(d).

Ordinul de integrare al unei serii este egal cu numărul care arată de câte ori trebuie diferențiată seria pentru a deveni staționară şi este egal cu numărul de rădăcini egale cu 1. Majoritatea seriilor de date economice conțin o singură rădăcină egală cu 1. Cele mai multe serii cronologice macroeconomice au tendință şi, în mod specific, ele conțin trenduri stochastice.

Numim **Serii cointegrate** seriile integrate de acelaşi ordin, ce admit o combinație liniară care este I(0) sau integrată de un ordin mai mic decât ordinul de integrare al seriilor inițiale.

Definiție: Dacă există o combinație liniară staționară între variabile aleatoare nestaționare, atunci variabilele combinate sunt cointegrate (există o dinamică non-staționară comună).

Testul Dickey-Fuller (Unit Root Test). Testul pentru staționaritate (pentru o rădăcină unitară).

1. Se testează seria y_t pentru a vedea dacă este staționară. Dacă răspunsul este da, atunci $y_t \sim I(0)$. Dacă răspunsul este nu, atunci $y_t \sim I(d)$, unde $d > 0$.

2. Mai întâi se consideră diferențele lui y_t ca $\Delta y_t = y_t - y_{t-1}$ și se testează dacă Δy_t este staționară. Dacă da, atunci $\Delta y_t \sim I(0)$ și $y_t \sim I(1)$. Dacă nu, atunci $y_t \sim I(d)$, unde $d > 0$.

Testarea existenței unei rădăcini egale cu 1

Considerăm un proces de mers la întâmplare: $y_t = y_{t-1} + \varepsilon_t$ unde ε_t este un proces zgomot alb. Acesta este un model AR(1). Deoarece coeficientul lui y_{t-1} este 1 avem o rădăcină egală cu 1, adică serie nestaționară.

Dacă în modelul de regresie $y_t = \rho\, y_{t-1} + \varepsilon_t$ găsim că $\rho = 1$, spunem că variabila y_t are o rădăcină unitară. Scriem ecuația $y_t = \rho\, y_{t-1} + \varepsilon_t$ sub forma: $\Delta y_t = (\rho - 1)\, y_{t-1} + \varepsilon_t = \delta\, y_{t-1} + \varepsilon_t$.

Dacă $\delta = 0$ putem scrie $\Delta y_t = y_t - y_{t-1} = \varepsilon_t$. Seria diferențelor de ordinul întâi este o serie staționară.

Testul Dickey-Fuller (Unit Root Test)

H_0: seria are rădăcină unitară și este nestaționară ($\rho = 1$ sau $\delta = 0$)

H_1: seria este staționară ($\rho < 1$ sau $\delta < 0$)

Dacă $\rho = 1$ sau $\delta = 0$, atunci seria nu este staționară.

Dacă $\rho < 1$ sau $\delta < 0$, atunci seria este staționară. Avem un test unilateral.

Sub ipoteza nulă că $\rho = 1$ sau $\delta = 0$ statistica folosită este cunoscută ca **statistica τ (tau)**. Aceasta este definită ca fiind $\tau = DF = \hat{\delta} / se(\hat{\delta})$. Această statistică nu urmează distribuția *t* uzuală, sub ipoteza nulă. Valorile critice sunt obținute prin experimente Monte Carlo și se găsesc în tabele construite de Dickey și Fuller. Dacă respingem ipoteza nulă că $\delta = 0$ sau $\rho = 1$, rezultă că seria este staționară și putem folosi testul *t*.

Dickey și Fuller au propus trei ecuații de regresie diferite, care pot fi folosite pentru a testa prezența unei rădăcini egale cu 1. În fiecare caz se formulează ipoteza nulă și ipoteza alternativă.

i) $H_0 : y_t = y_{t-1} + \varepsilon_t$ iar $H_1 : y_t = \rho y_{t-1} + \varepsilon_t, \rho < 1$. Nu există termen de interceptare și nu există trend.

ii) $H_0 : y_t = y_{t-1} + \varepsilon_t$ iar $H_1 : y_t = \rho y_{t-1} + \mu + \varepsilon_t, \rho < 1$. Există termen de interceptare și nu există trend.

iii) $H_0 : y_t = y_{t-1} + \varepsilon_t$ iar $H_1 : y_t = \rho y_{t-1} + \mu + \lambda t + \varepsilon_t, \rho < 1$. Există termen de interceptare și există trend.

Cele trei modele alternative pot fi rescrise sub forme echivalente. Vom avea modelele:

$y_t - y_{t-1} = (\rho - 1) y_{t-1} + \varepsilon_t \Rightarrow \Delta y_t = \delta\, y_{t-1} + \varepsilon_t$

$y_t - y_{t-1} = (\rho - 1) y_{t-1} + \mu + \varepsilon_t \Rightarrow \Delta y_t = \delta\, y_{t-1} + \mu + \varepsilon_t$

$y_t - y_{t-1} = (\rho - 1) y_{t-1} + \mu + \lambda t + \varepsilon_t \Rightarrow \Delta y_t = \delta\, y_{t-1} + \mu + \lambda t + \varepsilon_t$

Parametrul de interes în toate cele trei ecuații de regresie este δ. Dacă o serie este staționară, condiția $-1 < \rho < 1$ este echivalentă cu condiția de staționaritate: $-2 < \delta < 0$. Dacă $\delta = 0$, seria conține o rădăcină egală cu 1.

Testul ADF – Augmented Dickey Fuller Test ține seama și de posibilitatea ca erorile ε_t să nu fie zgomot alb.

Deoarece termenul eroare este puțin probabil să fie un proces de zgomot alb, Dickey și Fuller au extins procedura, sugerând o versiune îmbogățită, care folosește *p* decalaje ale variabilei dependente.

Testul *ADF* include termeni AR(*p*) ai termenului Δy_t în cele trei modele alternative. Dacă termenul eroare este autocorelat, ultimul din cele trei modele va fi: $\Delta y_t = \delta\, y_{t-1} + \mu + \lambda t + + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + \eta_t$.

O problemă care apare este **determinarea numărului optim de decalaje** ale variabilei dependente. Pentru aceasta se folosesc criteriile AIC și SIC din pachetul Eviews.

Valori critice pentru testele ADF

	10%	5%	1%
există intercept	–2,57	–2,86	–3,43
există intercept și trend	–3,12	–3,41	–3,96

Alte teste de rădăcină unitară:

Testul Dickey-Fuller with GLS Detrending(DFGLS) – înainte de aplicarea testului este eliminată atât tendința din date, cât și influența variabilelor exogene, printr-un proces de cvasi-diferențiere.

Testul Phillips-Perron (PP) – o metodă (neparametrică) de testare a rădăcinii unitate în condiții de autocorelare a erorilor.

Testul Kwiatkowski, Phillips, Schmidt, and Shin (KPSS)– analizează proprietățile reziduurilor din ecuația de regresie a lui y_t în funcție de exogenele x.