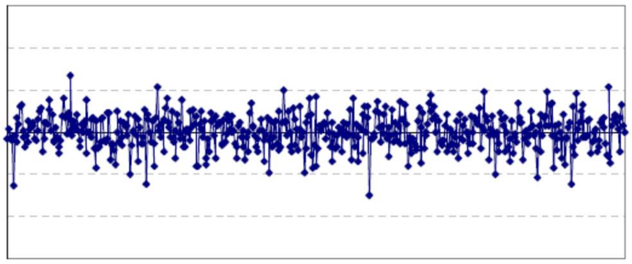
CURS 8 - SERII DE TIMP

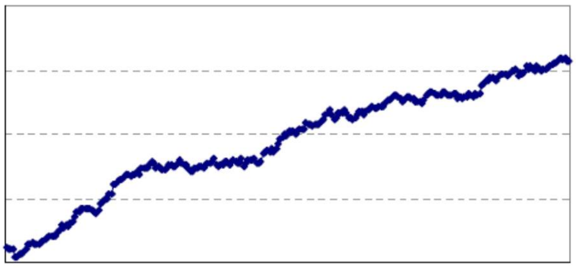
Valorile discrete înregistrate în timp, obţinute ca rezultat al observaţiilor făcute asupra diferitelor fenomene, sunt înregistrate sub forma seriilor cronologice, denumite si serii de timp sau serii dinamice. O definiţie foarte succintă a unei serii cronologice ar putea fi o colecţie de valori înregistrate secvenţial în timp.

Studiul seriilor de timp, în diferite domenii de activitate, precede înregistrările documentare ale istoriei. Economiştii secolelor trecute au făcut de multe ori referire la metode de analiză a seriilor de timp preluate din domeniul astronomiei și folosite ca sursa de inspiraţie pentru analiza datelor din domeniul economic. Exemple în acest sens sunt economistul francez Antoine Augustin Cournot (1801-1877) si economistul englez William Stanley Jevons (1835- 1882), care au făcut în mod explicit apel la metode de observare a seriilor de timp preluate din astronomie sau meteorologie. Mai târziu, la începutul sec. XX, analiza seriilor de timp s-a dezvoltat ca un domeniu de cercetare specific științelor economice. În prezent, metodele de previziune bazate pe analiza seriilor de timp sunt foarte frecvent folosite datorita simplităţii lor, fiind implementate în majoritatea pachetelor software de previziune.

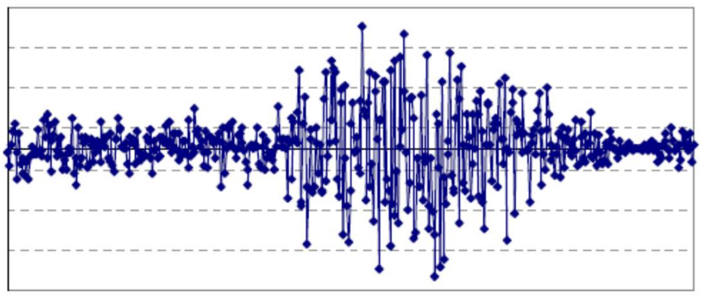
Seriile de timp pot fi *clasificate* după mai multe criterii, unul dintre cele mai relevante fiind stabilitatea evoluţiei în timp. După acest criteriu, putem avea serii de timp *staţionare* și *non-staționare*, așa cum se poate observa în Fig.1.



(a) Serie de timp staționară în medie și varianță



(b) Serie de timp non-staționară în medie



(c) Serie de timp staționară în medie dar non-staționară în varianță

Fig. 1. Staționaritatea seriilor de timp

Staționaritatea poate fi apreciată din mai multe puncte de vedere.

Stationaritatea în medie presupune ca fenomenul observat sa aibă aproximativ aceeaşi medie pe subintervale, ceea ce implică inexistenta unui trend crescător sau descrescător. Astfel de evoluţii se mai numesc și evoluţii orizontale.

Stationaritatea în varianță presupune ca seria de timp observată să aibă o deviaţie față de medie aproximativ constantă pe subintervale.

Înainte de aplicarea unei metode de modelare si ulterior de predictie, seriile de timp sunt supuse în general unei proceduri de *prelucrare preliminară* (*ajustare*). Rolul prelucrării preliminare este multiplu și vizează următoarele aspecte:

- În situaţia în care o serie de timp este nestationară, de multe ori se aplică proceduri pentru a o *staţionariza*. Printre aceste proceduri se numără: calcularea diferenţelor de ordinul întâi, calcularea indicelui de creștere, logaritmarea, transformari de tip Box-Cox ($y_t = \frac{x_t{}^{\alpha}-1}{\alpha}$, $\alpha \neq 0$).
- Atunci când datele prezintă fluctuaţii sezoniere, iar obiectivul urmărit este observarea tendinţei pe termen lung sau a fluctuaţiilor ciclice, se recurge deseori la înlăturarea fluctuaţiilor sezoniere. Aceasta procedură poartă denumirea de *desezonalizare*.
- Ajustarea seriilor de timp mai poate presupune *înlăturarea valorilor aberante* (valori foarte mari sau foarte mici datorate în special erorilor de înregistrare), *interpolarea* (estimarea valorilor istorice care lipsesc) etc.
- O practică des întâlnită în cazul datelor lunare este *înlăturarea efectului numărului de zile lucrătoare* dintr-o lună. De exemplu, dacă o lună are mai puţine zile lucrătoare decât luna anterioară nu înseamnă că nivelul vânzărilor pentru un anumit produs a scăzut, ci trebuie comparat nivelul vânzărilor pe cele două luni, după ce este înlăturat efectul diferenţei dintre numărul de zile lucrătoare pentru cele două luni.
- Atunci când datele sunt exprimate în prețuri curente, iar inflaţia este semnificativă, înainte de a aplica o metodă oarecare de extrapolare, se recurge la eliminarea inflaţiei. Acest lucru se realizează prin transformarea seriei de timp în prețuri constante (comparabile) prin *inflatare* sau *deflatare*. Inflatarea are avantajul utilizării prețurilor celor mai actuale ca prețuri de referinţă. De exemplu se ia ca referinţă ultimul an și toate valorile sunt transformate în prețurile ultimului an.

Procese de tip “mers la întâmplare”

În anul 1900, Louise Jean-Baptiste Alphonse Bachelier, în teza sa *Teoria Speculației,* a fost prima persoană consemnată din istorie care a aproximat evoluţia prețurilor pe bursele de valori cu o mişcare browniană sau, altfel spus, cu un “mers la întâmplare”.Matematicianul de origine franceză este considerat un pionier al matematicii financiare și al proceselor aleatoare.

Procesul de tip “mers la întâmplare” este un proces în care valoarea la momentul t +1 este egală cu valoarea la momentul anterior plus o eroare:

$$y_{t+1} = y_t + e_t,$$

unde: e_t este o variabilă aleatoare normal distribuită de medie 0 și varianță constantă.

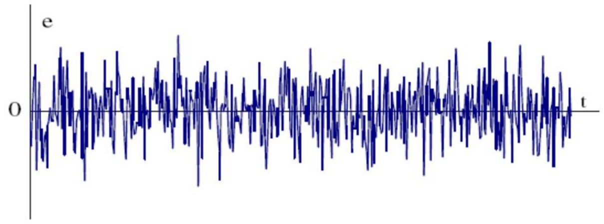


Fig. 2. Serie de timp reprezentând o variabilă aleatoare N(0,1)

Pentru că valorile au aceeași medie, 0, și aceeași varianță, 1, pe fiecare subinterval de timp, în graficul de mai sus avem un exemplu de serie staţionară în medie și varianță.

Seria din Fig. 3 reprezintă un “mers la întâmplare” pur, în sensul ca datele au fost create artificial, în conformitate cu formula prezentată, fără să existe alți factori sau alte procese care să influenţeze evoluţia datelor. Din grafic se poate observa cu ușurință că procesul este în medie nestaţionar.

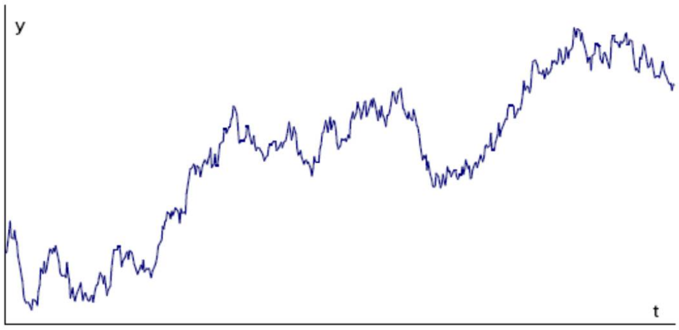


Fig. 3. Proces de tip mers la întâmplare

În realitatea economică nu exista însă date care să reprezinte astfel de procese pure, ci date care pot fi aproximate “grosso modo” cu astfel de procese. Să luăm ca exemplu evoluţia zilnică a cursului valutar Ron/Euro în perioada ian.-sept. 2008 (Fig. 4).

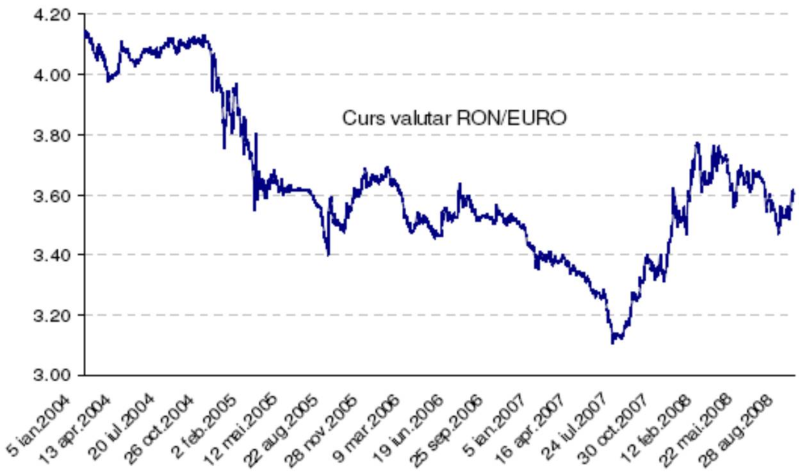


Fig. 4. Evoluția cursului valutar Ron/Euro în perioada ian.-sept. 2008

Evoluţia cursului valutar nu poate fi considerată un “mers la întâmplare” pur, aceasta fiind determinată de o serie de factori (factori care influenţează raportul cerere-ofertă pentru o anumită valută) a căror influență nu o vom discuta aici dar care elimină caracterul pur aleator al procesului.

În cazul în care o serie de valori ar urma un de proces de tip “mers la întâmplare” pur, previziunea cea mai potrivită este ultima valoare înregistrată. Adică previziunea pentru momentul următor, n+1, este chiar valoarea din momentul n**:**

$$\hat{y}_{n+1} = y_n ,$$

unde: $\hat{y}_{n+1}$ - valoarea previzionată pentru momentul t+1;

y_n - valoarea reală înregistrată în momentul t.

Dacă însă procesul care stă la baza seriei de date nu este pur aleator, cum de altfel se întâmplă, este foarte probabil ca o altă metodă să fie mai potrivită.

Metodele de descompunere a seriilor de timp

Metodele de descompunere se întemeiază pe ipoteza ca valorile reale a unei serii de date statistice pot fi descompuse în următoarele categorii de variaţii sau mişcări.

- *Variaţie de lungă durată*, stabilă, numită și tendinţa. Această variaţie determină sensul general al evoluţiei procesului sau fenomenului evidenţiat de seria de date statistice, marcând direcţia fundamentala a mişcării.
- *Variaţii periodice, sezoniere sau ritmice în jurul tendinţei*. Aceste variaţii au loc în anumite condiţii specifice, fiind rezultatul acţiunii unor factori naturali (încasările în turism, producţia în construcţii, exportul unor produse agricole etc.).
- *Variaţii întâmplătoare, accidentale, perturbatoare sau reziduale*. Aceste variaţii sunt determinate de abaterile de la tendinţa și variaţiile periodice, fiind rezultatul acţiunii unor factori imprevizibil, necuantificabili, de regulă de scurtă durată.

Nu orice serie statistică conţine toate aceste tipuri de variaţii. De exemplu, sunt numeroase cazuri în care variaţiile sezoniere pot sa lipsească din seria de date statistice.

Folosirea seriilor statistice în previziune, impune asocierea celor trei componente după unul din cele doua modele fundamentale: modelul aditiv sau modelul multiplicativ.

Modelul aditiv se utilizează în cazul unei relaţii de independență totală a fiecărei componente față de celelalte, rezultând din simpla însumare a acestora.

$$Y = T + S + E,$$

unde:

Y - variabila rezultativă;

T - tendinţa;

S - variaţii sezoniere;

E - variaţii întâmplătoare.

Modelul multiplicativ presupune existența unei relaţii de proporţionalitate a componentelor seriei de date, valorile observate fiind produsul acestor componente.

$$Y = T \times S \times E$$

Evoluţia unei serii statistice de timp cu sezonalitate aditivă, respectiv multiplicativă, se prezintă în Fig. 10.

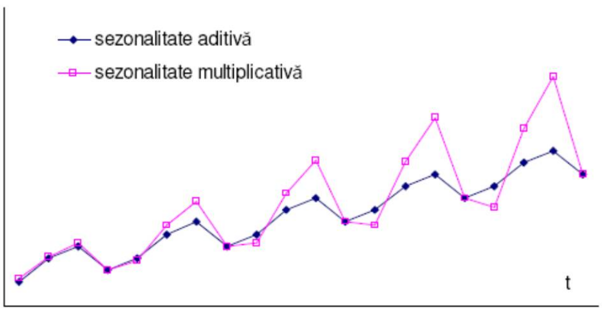


Fig.4. Tipuri de variaţie sezonieră

Etapele realizării unei previziuni prin metoda clasică de descompunere:

1. Se calculează o *medie mobilă de ordin k*. În general se utilizează ordinul 4 pentru date trimestriale, ordinul 12 pentru date lunare, 5 sau 7 pentru date zilnice etc. Dacă ordinul este par, se

aplică apoi o medie mobilă de ordin 2 pentru centrarea mediei mobile. Aceasta medie are rolul de a nivela fluctuaţiile sezoniere, reprezentând o estimare provizorie a componentei de trend.

2. Se face diferența/raportul dintre *valorile observate și valorile obţinute prin media mobilă centrată* obţinându-se coeficienţii de sezonalitate.

3. Se calculează *componenta sezonieră*, S_t, ca o medie aritmetică/geometrică a valorilor obţinute în etapa anterioară pentru fiecare unitate de timp care formează ciclul sezonier (daca datele sunt lunare este calculata o medie pentru fiecare luna). Cu alte cuvinte, componenta sezonieră este reprezentată de coeficienţii de sezonalitate medii.

4. Prin diferenţa/raportul dintre valorile observate si componenta sezoniera se obţine *seria de timp desezonalizată* (Y-S).

5. Se estimează *componenta de trend*, T_t, printr-o metodă specifică acestei componente (funcţii de tendință, spor mediu etc.) pornind de la *seria desezonalizată*.

6. Prin diferența/raportul dintre datele iniţiale observate și suma/produsul componentelor de sezonalitate și trend se obține *componenta aleatoare*, E_t.

7. Se extrapolează componenta de trend și cea sezonieră pe perioada aferentă orizontului de previziune.

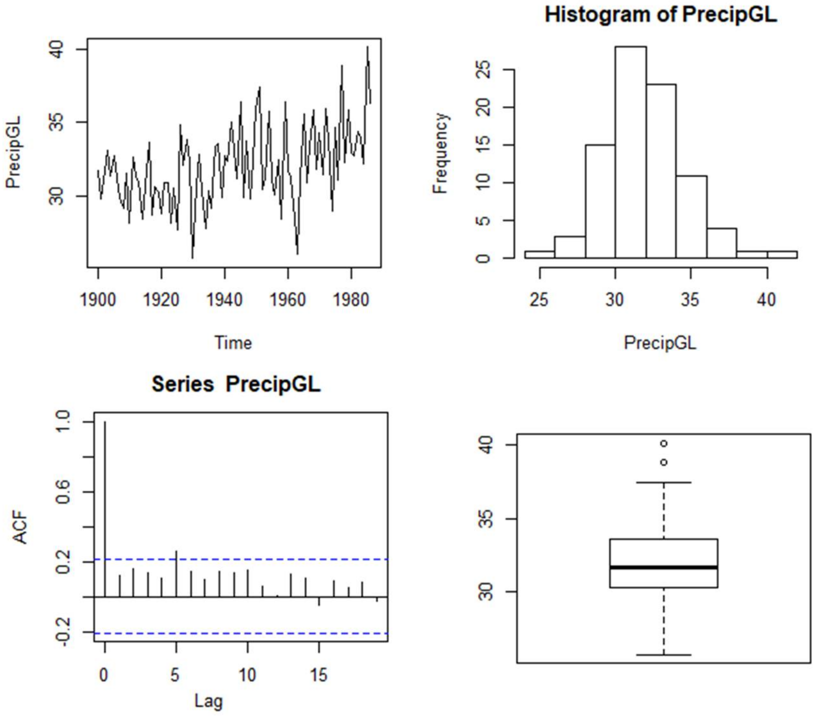
8. Se recompun valorile previzionate prin adunare/înmulțire.

Analiza seriilor de timp presupune

1. **Analiza grafica**
   - Plot - a se vedea evolutia
   - histograma – a se vedea asimetria
   - boxplot – a se determina valorile aberante si teste din package **outliers**
   - functia de autocorelatie

2. **Teste statistice**
   - Homoscedasticitatea: testul Bartlett – Bartlett.test (se construieste in indice CO care are cate 29 de 1, 2 si 3 si se aplicat testul pt x=PrecipGL si y=CO)
   - Testare ipotezelor legate de existent trendului – testul Mann-Kendall
     https://cran.r-project.org/web/packages/Kendall/Kendall.pdf
     https://cran.r-project.org/web/packages/trend/trend.pdf
     a. ip nula este ca nu exista trend monoton al seriei, iar cea alternative este ca exista un astfel de trend

```
library(Kendall)
data(PrecipGL)
par(mfrow=c(2,2))
plot(PrecipGL)
hist(PrecipGL)
acf(PrecipGL)
boxplot(PrecipGL)
```



MannKendall(PrecipGL)

tau = 0.265, 2-sided pvalue =0.00029206 # trend monoton

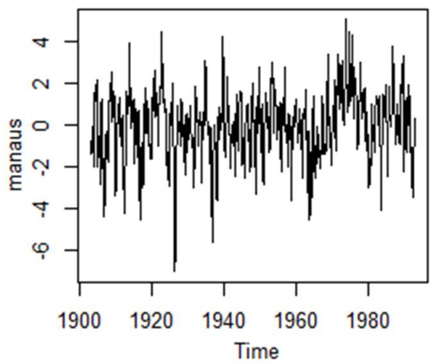
b. exista si varianta sezoniera a acest test:

SeasonalMannKendall(PrecipGL)

tau = 0.265, 2-sided pvalue =0.00028797 # trend sezonier monoton

```
library(boot)
data(manaus)
manaus
SeasonalMannKendall(manaus)
```

tau = 0.0877, 2-sided pvalue =2.2559e-05



Alternativ:

```
library(trend)
mk.test(manaus)
```

Mann-Kendall trend test

data: manaus
z = 4.3756, n = 1080, p-value = 1.211e-05
alternative hypothesis: true S is not equal to 0
sample estimates:

S	varS	tau
5.180400e+04	1.401620e+08	8.891681e-02

c. Determinarea tendintei liniare prin metoda neparametrica a lui **Sen** – **library trend**

This test computes both the slope (i.e. linear rate of change) and confidence levels according to Sen's method. First, a set of linear slopes is calculated as follows:

$$d_k = \frac{x_j - x_i}{j - i}$$

for $(1 \leq i < j \leq n)$, where d is the slope, x denotes the variable, n is the number of data, and i, j are indices.

Sen's slope is then calculated as the median from all slopes: $b_{Sen} = \text{median}(d_k)$.

This function also computes the upper and lower confidence limits for sens slope.

```
library(trend)

sens.slope(PrecipGL)
```

Sen's slope

data: PrecipGL
z = 3.6222, n = 87, p-value = 0.0002921
alternative hypothesis: true z is not equal to 0
95 percent confidence interval: [0.01952381, 0.06217391]
sample estimates: Sen's slope: 0.04

```
sens.slope(manaus)
```

Sen's slope
data: manaus
z = 4.3756, n = 1080, p-value = 1.211e-05
alternative hypothesis: true z is not equal to 0
95 percent confidence interval: [0.0003428378, 0.0009065682]
sample estimates: Sen's slope: 0.0006241918

- Testarea stationaritatii
  - in medie si tendinta – testul KPSS, in care ipoteza nula este este stationaritatea in nivel (sau trend), iar cea alternative este nestationaritatea in nivel (trend)
  https://cran.r-project.org/web/packages/tseries/tseries.pdf

```
library(tseries)
kpss.test(x, null = c("Level", "Trend"), lshort = TRUE)

kpss.test(PrecipGL)
```

KPSS Test for Level Stationarity

data: PrecipGL
KPSS Level = 1.1118, Truncation lag parameter = 3, p-value = 0.01

Warning message: In kpss.test(PrecipGL) : p-value smaller than printed p-value

kpss.test(PrecipGL, null = c("Trend"))

KPSS Test for Trend Stationarity

data: PrecipGL
KPSS Trend = 0.091423, Truncation lag parameter = 3, p-value = 0.1

Warning message:
In kpss.test(PrecipGL, null = c("Trend")) : p-value greater than printed p-value

stationaritate in trend si nestationaritate in nivel

- testul Dickey-Fuller si ADF (Augmented Dickey-Fuller), in care ipoteza nula este existent unei unitati, iar cea alternative este stationaritatea

- Determinarea punctelor de ruptura – Testele Pettitt, Buishand U Test, Hubert – in package **trend**.