Serii temporale

Marina Gorunescu
mgorun@inf.ucv.ro

În cursul observării fenomenelor sau situațiilor din lumea înconjurătoare, se constată că datele care ne interesează și pe care le strângem au o cronologie a lor, care ne permite analizarea evoluției lor pe măsura scurgerii timpului.

O asemenea secvență temporală de date (observații secvențiale), notată $(\xi_t\,,\ t \in T)$, unde $T \subseteq \mathbf{R}$, se referă la *timp*, se numește ***serie temporală***, sau *serie dinamică*.

Datele ξ_t se pot referi la observații în timp discret, adică zile, săptămâni, trimestre, ani etc., sau pot fi date obținute în timp continuu.

O serie temporală este o mulțime de observații măsurate în perioade succesive de timp.

Datele reprezentate de seriile temporale sunt secvențe de date rezultate din măsurători care urmează o ordine deterministă.

In cazul celor mai multor analize statistice sunt studiate eşantioane aleatoare de observații.

Seriile temporale se vizualizează prin tabele sau prin grafice

- tabelul sau graficul fluxului de călători într-un aeroport pentru un anumit interval de timp;
- graficul temperaturilor într-o anumită lună a anului, înregistrate în ultimul secol;
- graficul frecvenței apariției unei anumite boli pe o anumită perioadă de timp;
- graficul valorilor unor analize medicale ce se efectuează periodic;
- graficul evoluției ratei de schimb valutar.

Exemplu

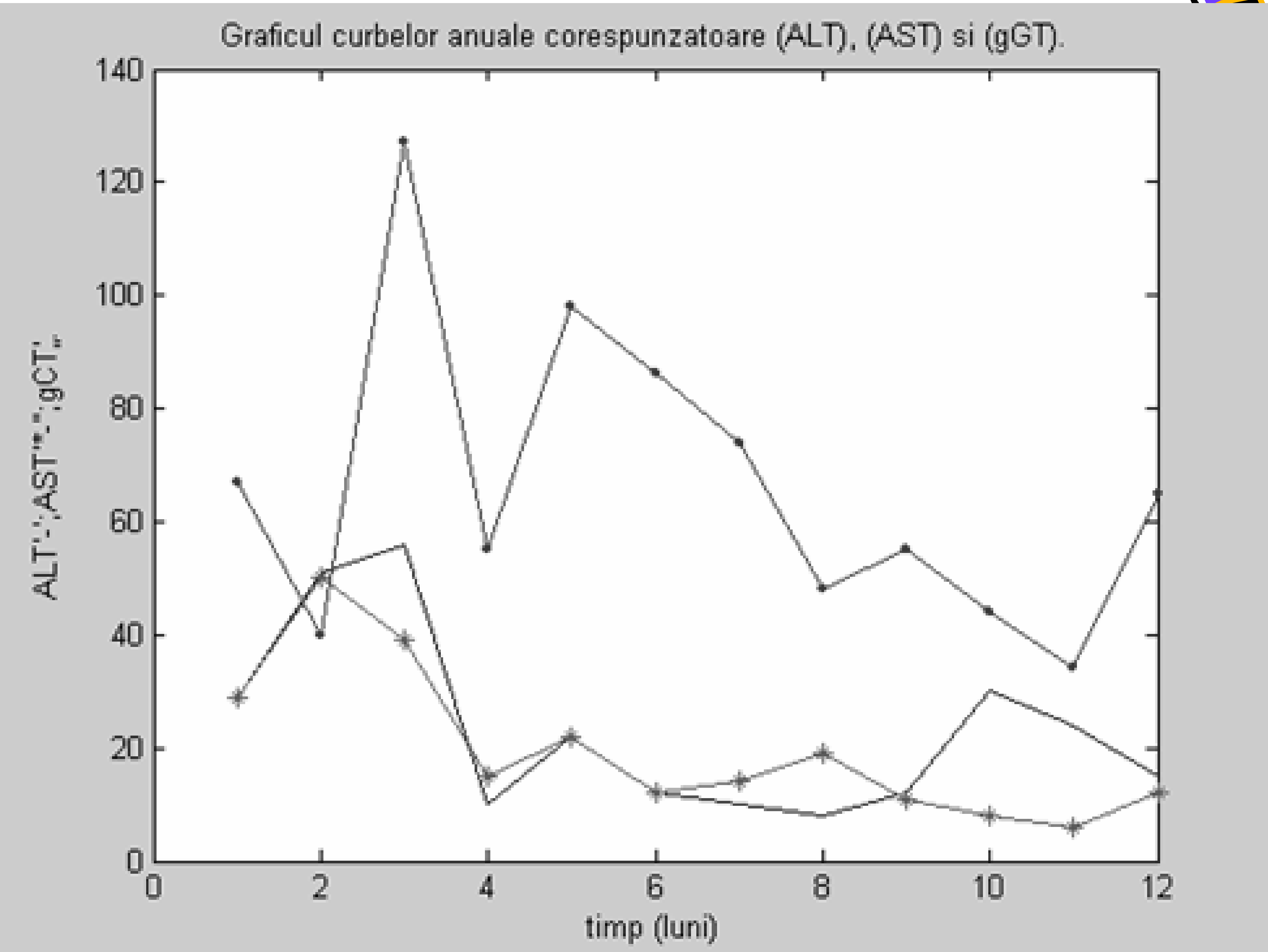
Reprezentare grafică următoare se referă la domeniul medical.
În tabelul de mai jos sunt valorile pentru principalele enzime serice ce indică leziunile colestatice şi hepatocelulare în măsură să indice diagnosticul de hepatită cronică de tip *C*.
Concret, analizele prelevate timp de 12 luni consecutiv, de la pacienți având hepatită cronică *C*, aceste analize lunare referindu-se la următoarele enzime serice semnificative din punct de vedere medical:

- *alanina transaminasă (ALT);*
- *aspartate transaminase (AST);*
- *gamma glutamyl transpepdiase (gGT).*

Tabelul următor prezintă analiza lunară a principalelor enzime serice în hepatita cronică.

	I	F	M	A	M	I	I	A	S	O	N	D
ALT	29	51	56	10	22	12	10	8	12	30	24	15
AST	29	50	39	15	22	12	14	19	11	8	6	12
gGT	67	40	127	55	98	86	74	48	55	44	34	65

```
» x=1:12;y1=[29 51 56 10 22 12 10 8 12 30 24 15];
» y2=[29 50 39 15 22 12 14 19 11 8 6 12];
»  y3=[67 40 127 55 98 86 74 48 55 44 34 65];
» plot(x,y1,x,y2,'*-',x,y3,'.-')
```



Graficul curbelor anuale corespunzatoare (ALT), (AST) si (gGT).
ALT'-';AST'*-';gCT',,
timp (luni)

Scopurile principale ale analizei seriilor temporale:

- Identificarea naturii fenomenului reprezentat de secvența observațiilor care sunt analizate;
- Prognozarea valorilor viitoare ale variabilei ce reprezintă seria temporală pe baza datelor cunoscute.

Pentru a înțelege pattern-ul sau comportarea datelor unei serii temporale este important să înțelegem că o serie temporală are patru componente:

- trend-ul;
- componenta ciclică;
- componenta sezonieră;
- componenta incidentală

Trendul seriei temporale

- *Trendul* (tendința) se referă la o componentă sistematică, generală, liniară sau, de cele mai multe ori, neliniară, care se schimbă cu trecerea timpului şi nu se repetă.

Deplasarea graduală a seriei temporale este percepută ca o direcție

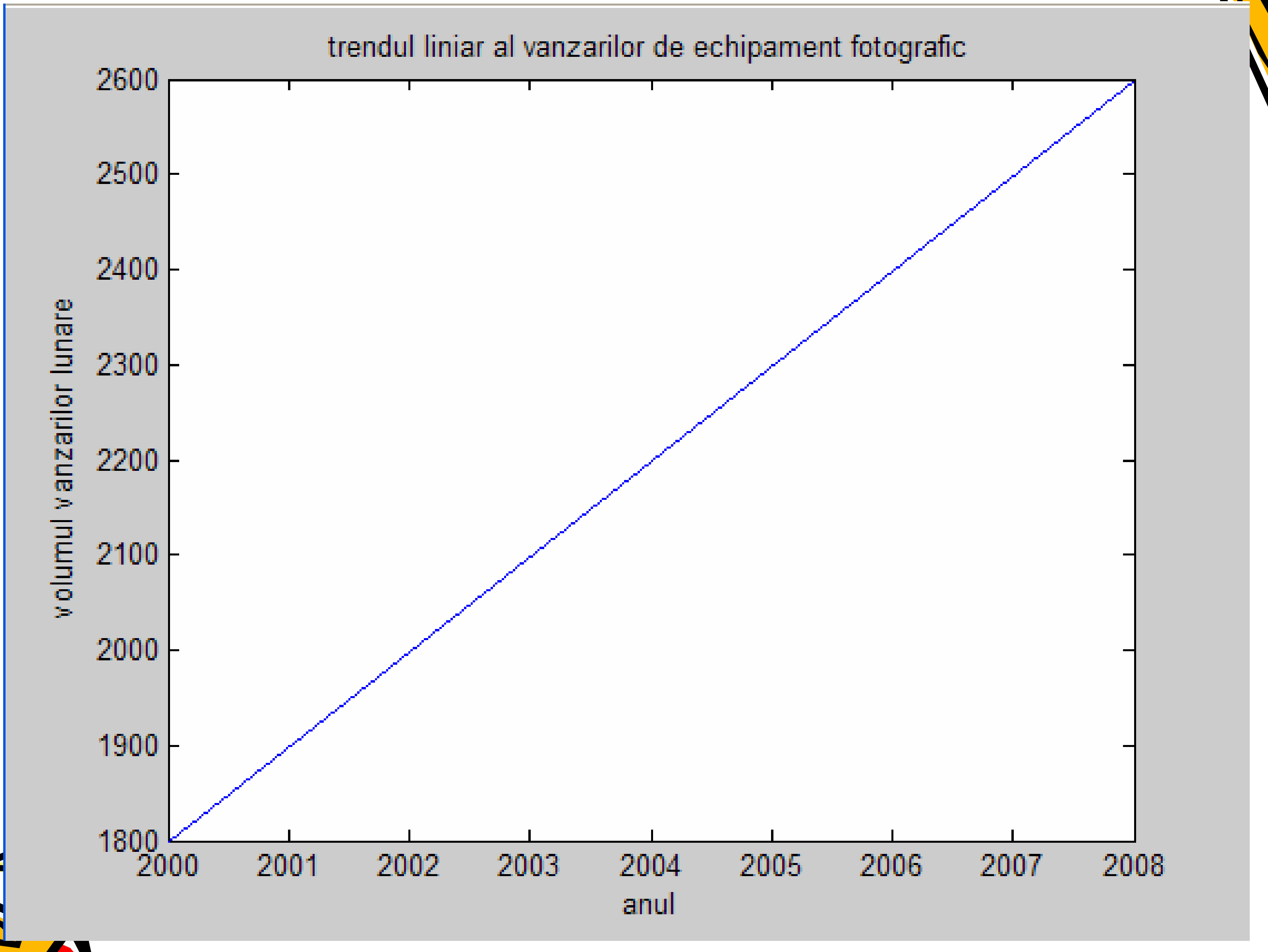
- **un trend**.

Se datorează de exemplu schimbărilor în populație, caracteristicilor demografice ale populației, schimbărilor în tehnologie, schimbărilor în preferințele consumatorilor.

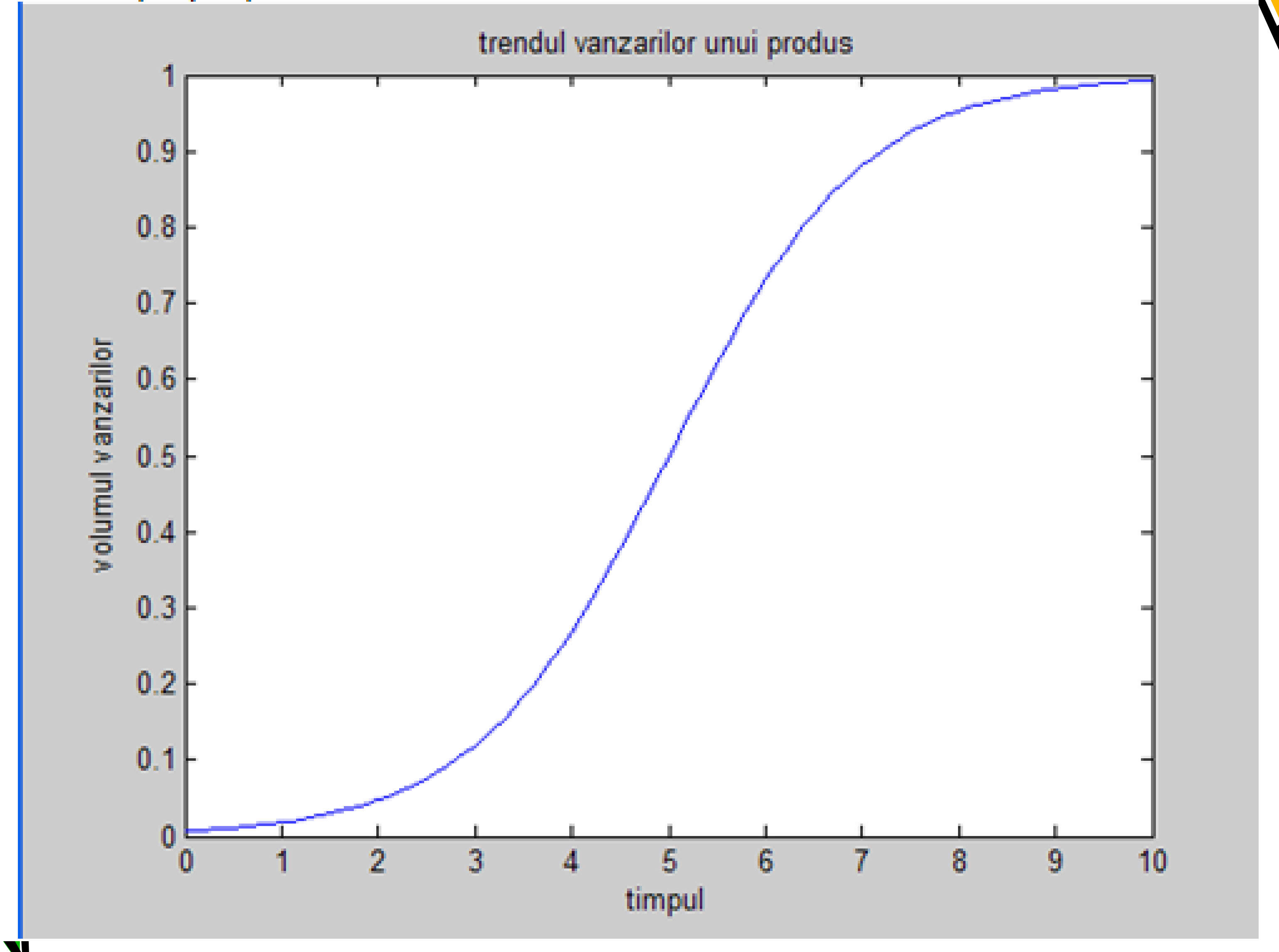
Exemplu

O firmă ce produce echipamente fotografice prezintă un **trend** ascendent al vânzărilor, după cum arată datele următoare:

- 2000: 1800 aparate/lună
- 2004: 2200 aparate/lună
- 2008: 2600 aparate/lună


trendul liniar al vanzarilor de echipament fotografic
volumul vanzarilor lunare
2600
2500
2400
2300
2200
2100
2000
1900
1800
2000
2001
2002
2003
2004
2005
2006
2007
2008
anul

Prezentăm simularea vânzărilor unui produs, de la introducerea lui pe piață, urmată de perioada de creștere a vânzărilor și în final de saturația pieței.



***Analiza trendului* (tendinței)**

În principiu, nu există tehnici de identificare automată a trendului în seriile temporale.

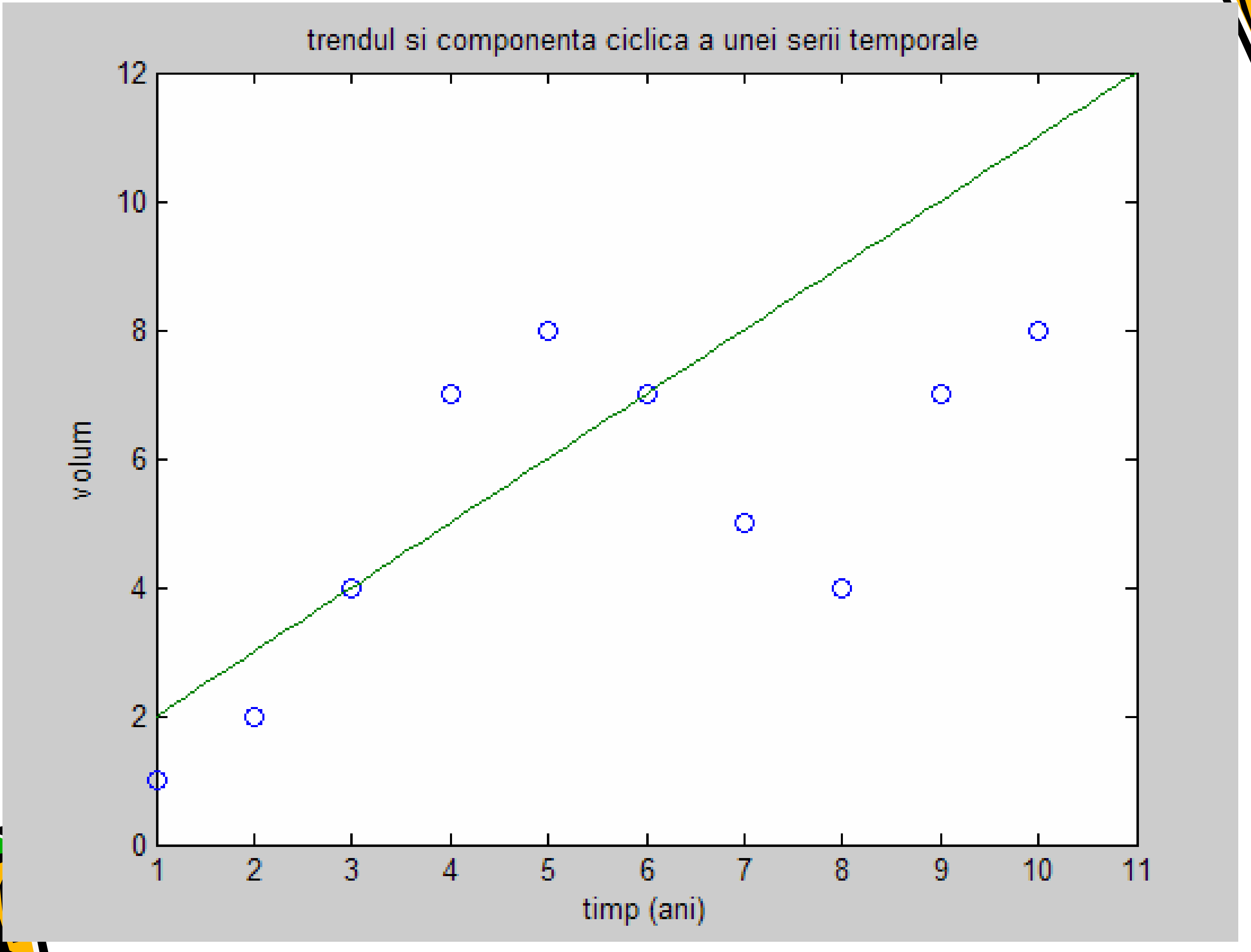
Totuşi, atâta timp cât tendința este monotonă (crescătoare sau descrescătoare), nu sunt probleme deosebite, de cele mai multe ori vizualizarea datelor este suficientă pentru identificarea acesteia.

Analiza trendului este necesară şi atunci când studiem mai multe serii temporale corespunzătoare aceleaşi perioade de timp.: analizăm motivele pentru care două sau mai multe asemenea serii temporale au aceeaşi direcţie (sunt puternic corelate), chiar dacă, la prima vedere, acest fapt pare inexplicabil.

Componenta ciclică

Nu toate valorile seriei temporale vor fi pe linia trend-ului. Orice secvență periodică de puncte, a cărei durată este cel puțin o unitate de măsură a timpului (de exemplu un an) se numește **componenta ciclică** a seriei temporale.

Asemenea componente apar în economie: perioade de inflație mai modeste urmate de perioade de inflație rapidă.


trendul si componenta ciclica a unei serii temporale
volum
timp (ani)
0
2
4
6
8
10
12
1
2
3
4
5
6
7
8
9
10
11

Componenta sezonieră

O serie temporală poate prezenta o formă regulară într-o perioadă de timp mai mică de un an.
De exemplu un producător de piscine ştie că volumul vânzărilor este mic în perioada de toamnă-iarnă în timp ce primăvara şi vara atinge maximum posibil. Producătorii de echipament de schi au un pattern al vânzărilor opus celui prezentat

Componenta seriei temporale ce prezintă variații ale datelor influențate de anotimpuri este **componenta sezonieră**.

Exemple

- Volumul traficului prezintă variații „sezoniere" în timpul unei zile cu valori maxime în timpul orelor de vârf, corespunzătoare orelor de începere respectiv încheiere a programului de lucru, minime în timpul nopții si moderate în rest.
- Domeniul vânzărilor în preajma Sărbătorilor de Crăciun şi Paşte.
- Domeniul internărilor în spital după Sărbătorile de Crăciun şi Paşte.

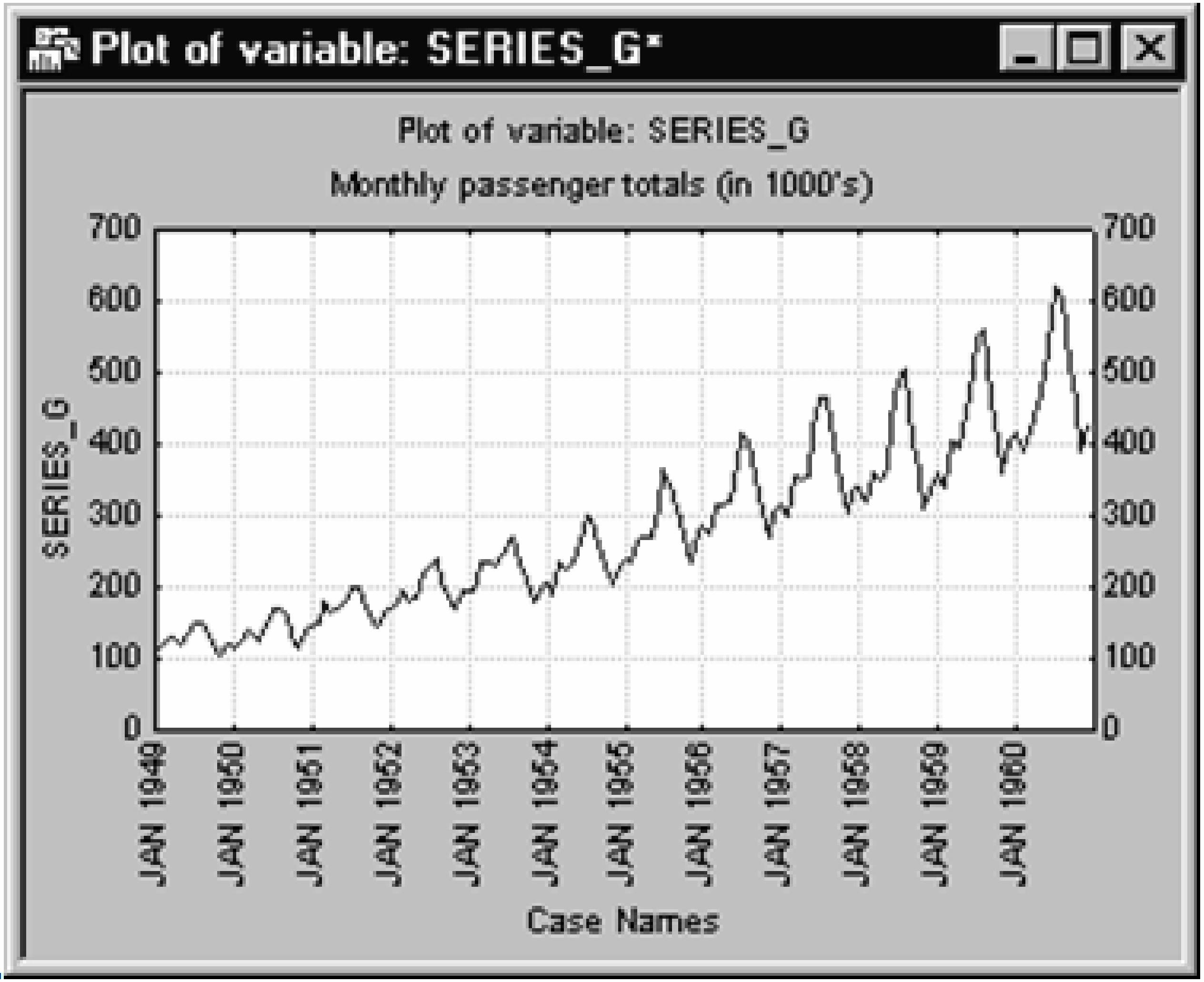
- (Box and Jenkins, 1976). *Acest exemplu clasic se referă* la numărul de pasageri pentru zborurile internaționale, înregistrat lunar din anul 1946 până în anul 1960. Baza de date corespunzătoare poate fi disponibilă sub forma următorului fişier (*Statistica*):

Data: Series_G.sta (1v by 144c)

Monthly passenger totals (in 1000's) 1949-1960; Box & Jenkins, 1976; series G.

	1 SERIES_G
JAN 1949	112
FEB 1949	118
MAR 1949	132
APR 1949	129
MAY 1949	121
JUN 1949	135
JUL 1949	148
AUG 1949	148
SEP 1949	136
OCT 1949	119
NOV 1949	104
DEC 1949	118
JAN 1950	115
FEB 1950	126
MAR 1950	141
APR 1950	135
MAY 1950	125
JUN 1950	149
JUL 1950	170
AUG 1950	170
SEP 1950	158
OCT 1950	133
NOV 1950	114
DEC 1950	140
JAN 1951	145
FEB 1951	150
MAR 1951	178
APR 1951	163

Graficul valorilor de trafic aerian privind aceste zboruri internaționale este ilustrat în figura de mai jos



Ce se observă din acest grafic?

- tendința crescătoare, liniară, a numărului de pasageri (plecând de la circa 100.000 în Ianuarie 1949 şi ajungând la aproximativ 400.000 în Ianuarie 1960).
- indiferent de an, în perioada lunilor de vară (la mijlocul anului) apare o tendință de creştere a numărului pasagerilor (datorată concediilor), după care urmează o descreştere –tendința sezonieră (periodică, repetitivă).

- *Ajustarea sezonieră* se referă la obținerea unei noi serii temporale, numită *serie ajustată*, după efectuarea unor anumite corecții asupra seriei inițiale în vederea îndepărtării factorilor perturbanți ce țin de anumite momente temporale (sezoniere).

Componenta incidentală

Componenta incidentală este factorul rezidual care apare din datele actuale ale seriei vs datele prognozate conform trendului, componentei ciclice şi componentei sezoniere. Această componentă este datorată factorilor de scurtă durată, pe care nu-i putem anticipa.

- *Prognoza* se referă la evaluarea unor valori viitoare ξ_{T+h}, $h > 1$, pe baza datelor cunoscute ξ_1, ξ_2,..., ξ_T.

De multe ori, în loc să se indice o anumită valoare prognozată, se indică un interval de prognoză.

Prognoza utilizând metode de netezire (smoothing methods)

Tehnicile de prognoză ce le vom prezenta se aplică seriilor temporale „aproape stabile", ceea ce înseamnă că nu prezintă modificări semnificative ale trend-ului ale componentei ciclice sau ale celei sezoniere.
În condiții reale, în care existența „zgomotului" este perturbatoare pentru identificarea trendului sau periodicității, este nevoie de un proces de ajustare/netezire a datelor.
Se urmărește netezirea/ ajustarea componentelor incidentale ale seriei prin folosirea procesului de mediere locală a datelor,

În acest mod, pot fi anihilate fluctuațiile locale și astfel se poate pune în evidență adevăratul trend al seriei.

În practică, această tehnică este cel mai adesea utilizată în domeniul finanțelor, de exemplu calculul prețului acțiunilor la Bursă, dar și alte domenii se bucură de avantajele sale privind „netezirea" datelor.

Media mobilă

În mediul de producție sunt necesare prognoze pentru mii de produse cunoscând date dintr-o săptămână sau o lună, caz în care este necesară o tehnică simplă, uşor de înțeles şi de folosit.

Media mobilă este media aritmetică a ultimelor n valori ale datelor

$$media\, mobila = \frac{p_M + p_{M-1} + \ldots + p_{M-n+1}}{n}$$

unde p_{M-i}, $i = 0, \ldots, n-1$ sunt cele n valori precedente.

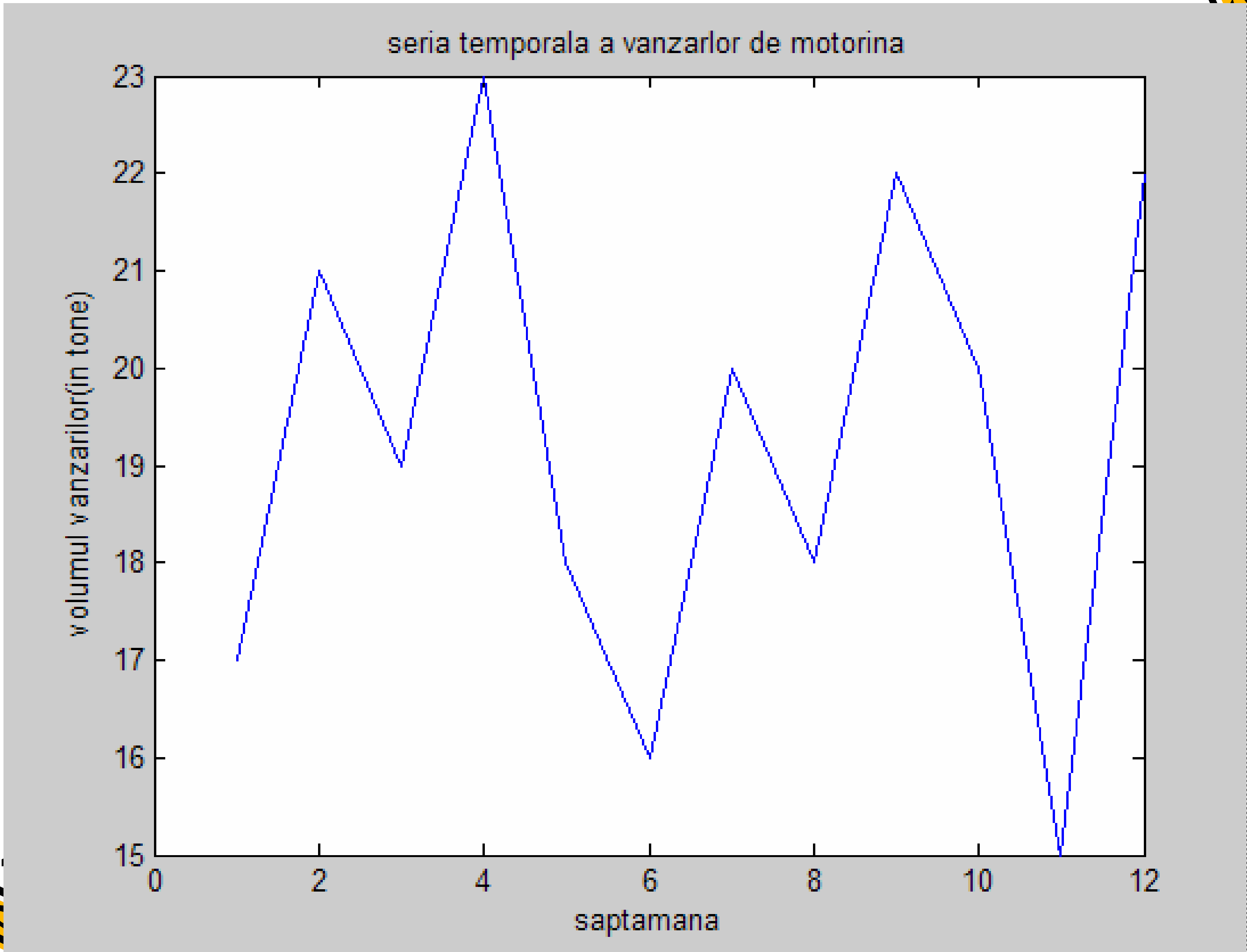
Orice nouă dată va înlocui cea mai veche dată din media mobilă, calculându-se astfel o nouă medie.

Astfel media se va schimba o dată cu apariția unei noi date, de unde şi termenul de medie mobilă.

Exemplu

Tabelul următor ilustrează vânzările de motorină ale unei stații X, pe parcursul a 12 săptămâni.

săptămâna	vanzări (tone)		săptămâna	vanzări (tone)
1	17		7	20
2	21		8	18
3	19		9	22
4	23		10	20
5	18		11	15
6	16		12	22


seria temporala a vanzarlor de motorina
volumul vanzarilor(in tone)
saptamana
23
22
21
20
19
18
17
16
15
0
2
4
6
8
10
12

Să calculăm media mobilă pentru 3 săptămâni, în cazul vânzărilor de motorină.
Media mobilă pentru primele trei luni este:

$$media\ mobila\ (sapt.1-3) = \frac{17+21+19}{3} = 19$$

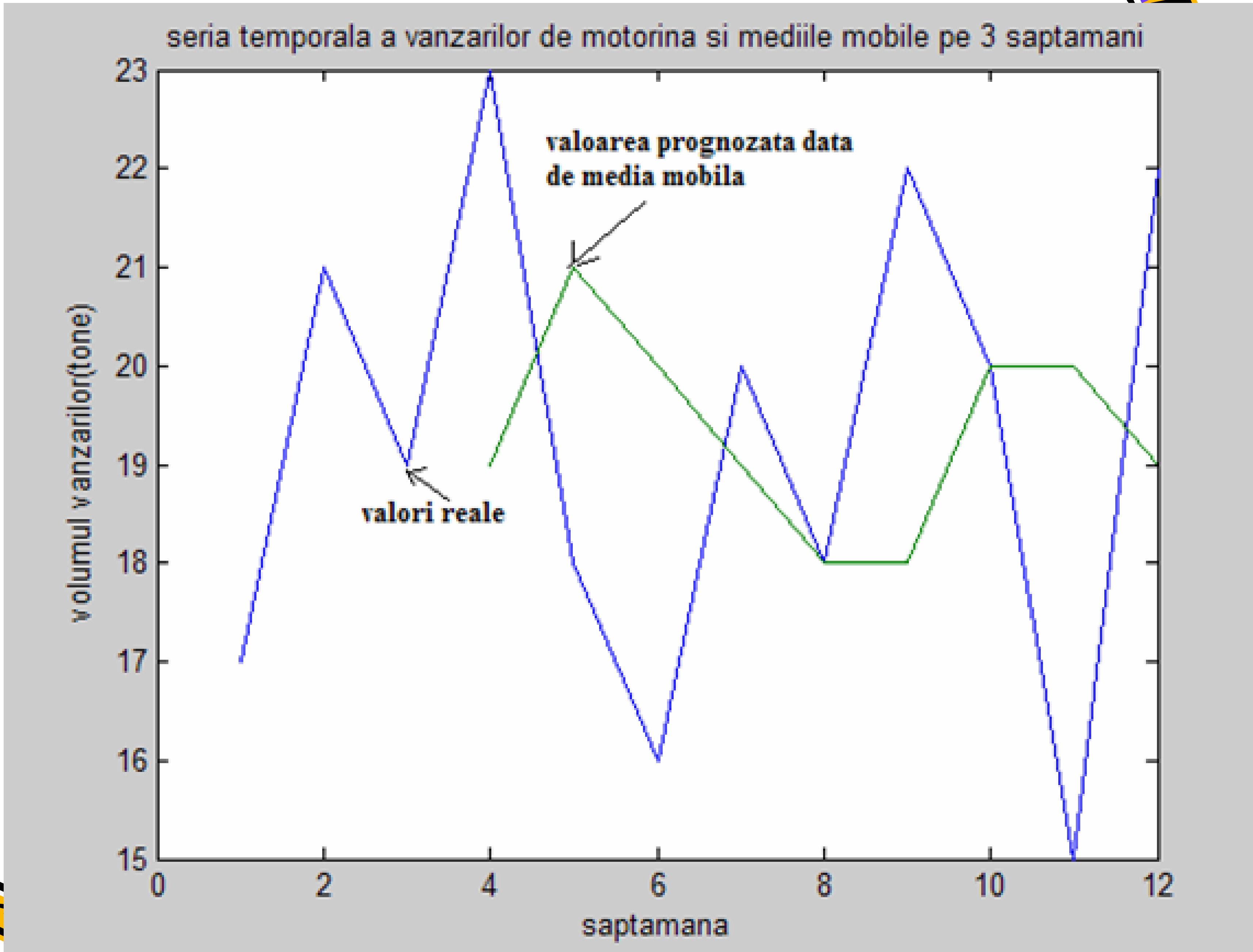
Această medie mobilă dă o prognoză pentru săptămâna a 4-a.

Eroarea în prognoza săptămânii a 4-a este $23-19=4$, eroarea fiind definită a fi diferența dintre valoarea reală din seria temporală şi valoarea prognozată.

Media mobilă pentru următoarele trei săptămâni este

$$media\,mobila\,(sapt.2-4)=\frac{21+19+23}{3}=21$$

Astfel prognoza pentru săptămâna a 5-a este 21, iar eroarea asociată este $18-21=-3$.


seria temporala a vanzarilor de motorina si mediile mobile pe 3 saptamani
valoarea prognozata data
de media mobila
valori reale
volumul vanzarilor(tone)
saptamana
23
22
21
20
19
18
17
16
15
0
2
4
6
8
10
12

săptămâna	valoarea din seria temporală	media mobilă	eroarea de prognozare	pătratul erorii de prognozare
1	17			
2	21			
3	19			
4	23	19	4	16
5	18	21	-3	9
6	16	20	-4	16
7	20	19	1	1
8	18	18	0	0
9	22	18	4	16
10	20	20	0	0
11	15	20	-5	25
12	22	19	3	9
		TOTAL	0	92

Pentru a aprecia acuratețea prognozei se ia în considerare media sumei pătratelor erorii (*mean squared error_ MSE*):

$$MSE = \frac{92}{9} = 10,22$$

În general problema constă în a găsi numărul optim de date ce vor fi introduse în media mobilă.
O abordare posibilă constă în a încerca diferite variante şi a o alege pe aceea ce minimizează *MSE*.
Dacă stabilim că acest număr *n* este optim pentru datele cunoscute (trecut), presupunem că va fi optim şi pentru viitor, şi vom face prognoza utilizând media mobilă pentru *n* date.

Media mobilă ponderată

În cazul mediei mobile fiecare dată luată în considerare are aceeaşi pondere.

O formă mai sofisticată de medie mobilă, care ține cont de importanța valorilor mediate este aşa-numita *medie mobilă ponderată* (*weighted moving average*-WMA), dată de formula:

$$WMA_M = \frac{np_M + (n-1)p_{M-1} + \ldots + 2p_{M-n+2} + p_{M-n+1}}{n + (n-1) + \ldots + 2 + 1}$$

Media mobilă ponderată implică acordarea unor ponderi specifice fiecărei date utilizate.

De obicei datelor cele mai recente li se atribuie ponderile cele mai mari.

În cazul datelor privind vânzările de motorină, media mobilă ponderată pentru trei săptămâni va fi:

$$Media\ mobila\ ponderata\,(sapt\ 4) = \frac{3 \cdot 19 + 2 \cdot 21 + 1 \cdot 17}{3 + 2 + 1} = 19.33$$

O altă variantă de medie mobilă este *media mobilă exponențială* (*exponential moving average* -EMA), sau *media mobilă exponențială ponderată* (*exponentially weighted moving average* -EWMA), în care ponderile pentru fiecare unitate de timp scad exponențial şi nu liniar (astfel, datele cele mai recente sunt şi celemai importante). Graficul unei asemenea descreşteri exponențiale a ponderilor este ilustrat mai jos.

Netezirea exponențială

Este o metodă simplă, deseori folosită de firme pentru prognoze.

Prezentăm modelul matematic:

$P_{t+1} = \alpha \cdot \xi_t + (1-\alpha) \cdot P_t$, unde:

P_{t+1} este prognoza pentru perioadă de timp $t+1$

ξ_t este valoarea seriei pentru perioadă de timp t

P_t este prognoza pentru perioadă de timp t

$\alpha \in [0,1]$ este constanta de netezire

Prognoza pentru orice perioadă de timp t este media ponderată a valorilor anterioare ale seriei temporale.

Considerăm valorile ξ_1, ξ_2, ξ_3 şi că $P_1 = \xi_1$.

Atunci:

$$P_2 = \alpha \cdot \xi_1 + (1-\alpha) \cdot P_1 = \alpha \cdot \xi_1 + (1-\alpha) \cdot \xi_1 = \xi_1$$

$$P_3 = \alpha \cdot \xi_2 + (1-\alpha) \cdot P_2 = \alpha \cdot \xi_2 + (1-\alpha) \cdot \xi_1$$

$$P_4 = \alpha \cdot \xi_3 + (1-\alpha) \cdot P_3 =$$

$$= \alpha \cdot \xi_3 + (1-\alpha) \cdot (\alpha \cdot \xi_2 + (1-\alpha) \cdot \xi_1) =$$

$$= \alpha \cdot \xi_3 + \alpha \cdot (1-\alpha) \cdot \xi_2 + (1-\alpha)^2 \cdot \xi_1$$

Dacă am determinat parametrul α pentru prognoza P_{t+1} avem nevoie de valoarea reală ξ_t şi de valoarea prognozată P_t.

Reluăm exemplul cu vânzările de motorină. Avem:

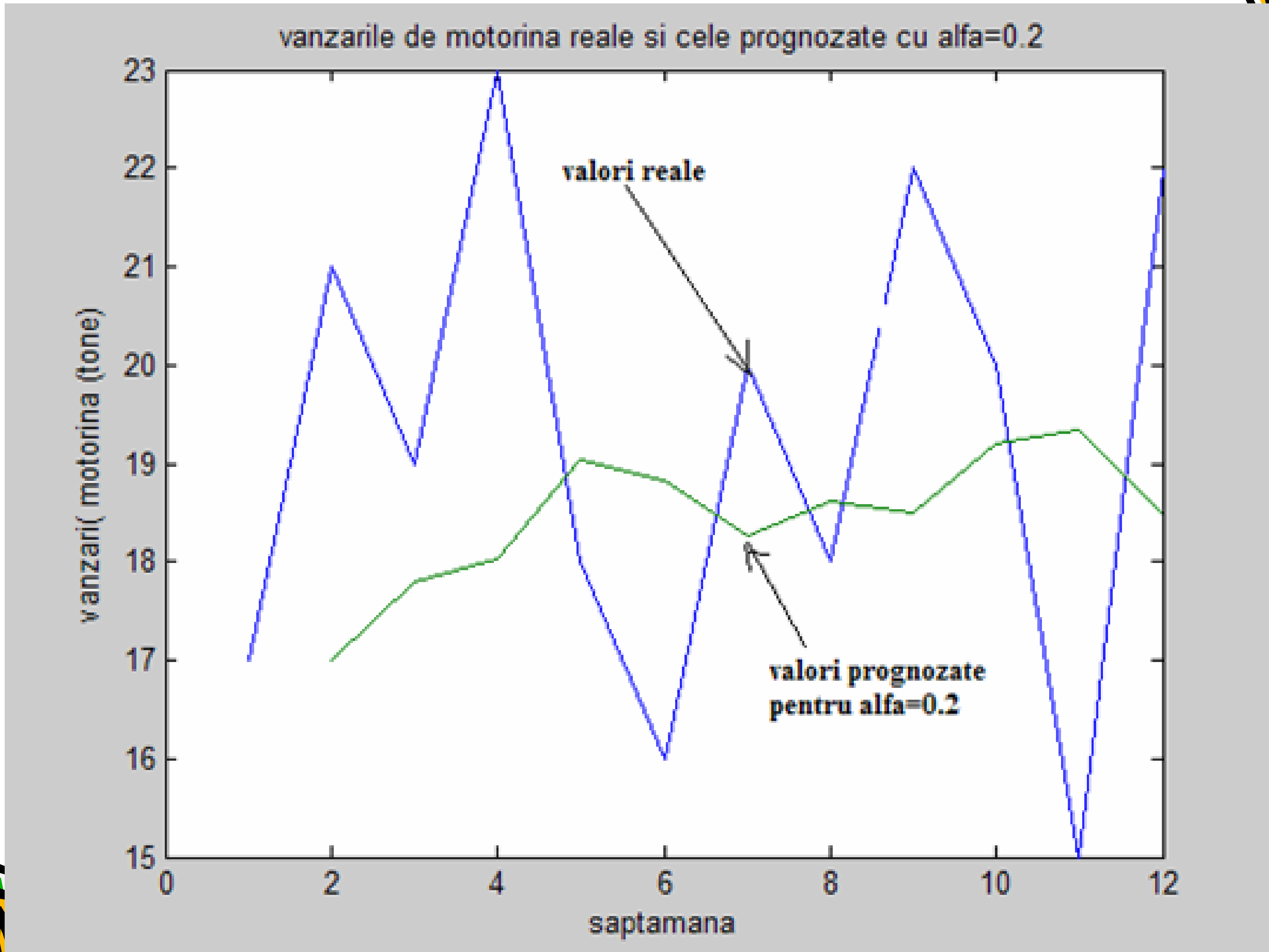
$\xi_1 = 17$, $P_2 = 17$ (conform argumentelor anterioare)

Deoarece $\xi_2 = 21$ eroarea va fi 4.

Pentru început considerăm $\alpha = 0.2$

$$P_3 = 0.2 \cdot \xi_2 + 0.8 \cdot P_2 = 0.2 \cdot 21 + 0.8 \cdot 17 = 17.8$$

$$P_4 = 0.2 \cdot \xi_3 + 0.8 \cdot P_3 = 0.2 \cdot 19 + 0.8 \cdot 17.8 = 18.04$$


vanzarile de motorina reale si cele prognozate cu alfa=0.2
vanzari(motorina (tone)
saptamana
valori reale
valori prognozate
pentru alfa=0.2
23
22
21
20
19
18
17
16
15
0
2
4
6
8
10
12

săptămâna	valoarea din seria temporală	valoarea prognozată	eroarea	pătratul erorii de prognozare
1	17			
2	21	17	4	16
3	19	17.8	1.2	1.44
4	23	18.04	4.96	24.6
5	18	19.03	-1.03	1.06
6	16	18.83	-2.83	8.01
7	20	18.26	1.74	3.03
8	18	18.61	-0.61	0.37
9	22	18.49	3.51	12.32
10	20	19.19	0.81	0.66
11	15	19.35	-4.35	18.92
12	22	18.48	3.52	12.39
		Total		98.8

$$MSE = \frac{98.8}{11} = 8.98$$

Pentru a alege cea mai bună valoare a parametrului α vom ține seama că valoarea prognozată P_{t+1} depinde de prognoza P_t şi de eroarea prognozei anterioare $\xi_t - P_t$:

$$P_{t+1} = \alpha \cdot \xi_t + (1-\alpha) \cdot P_t = P_t + \alpha \cdot (\xi_t - P_t)$$

Aşadar prognoza pentru perioada de timp $t+1$ se obține adăugând la prognoza pentru perioada de timp t o parte din eroarea prognozei pentru perioada de timp t.

Dacă seria temporală variază aleator este preferabilă o valoare cât mai mică a lui α .

In general alegem valoarea lui α care minimizează media sumei pătratelor erorii (*mean squared error_ MSE*).

Prezentăm în următorul tabel valorile prognozate, erorile şi pătratele acestora, in cazul parametrului $\alpha = 0.3$

săptămâna	valoarea din seria temporală	valoarea prognozată	eroarea	pătratul erorii de prognozare
1	17			
2	21	17	4	16
3	19	18.2	0.8	0.64
4	23	18.44	4.56	20.79
5	18	19.81	-1.81	3.28
6	16	19.27	-1.27	10.09
7	20	18.29	1.71	2.92
8	18	18.8	-0.8	0.64
9	22	18.56	3.44	11.83
10	20	19.59	0.41	0.17
11	15	19.71	-4.71	22.18
12	22	18.30	3.7	13.69
		Total		102.83

$$MSE = \frac{102.83}{11} = 9.35$$

Observăm că este preferabilă alegerea valorii parametrului $\alpha = 0.2$

Putem folosi metoda incrementală:
- dăm parametrului valorile
$\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$,
- calculăm media sumei pătratelor erorii corespunzătoare, *MSE*
- alegem acel α pentru care obținem cel mai mic *MSE*.

Cunoscând valoarea parametrului α , putem prognoza cantitățile de motorină ce vor fi vândute în a 13-a saptămână, a 14-a, etc.

O altă măsură a acurateții erorii este media valorilor absolute ale erorii (mean absolute deviation_*MAD*).

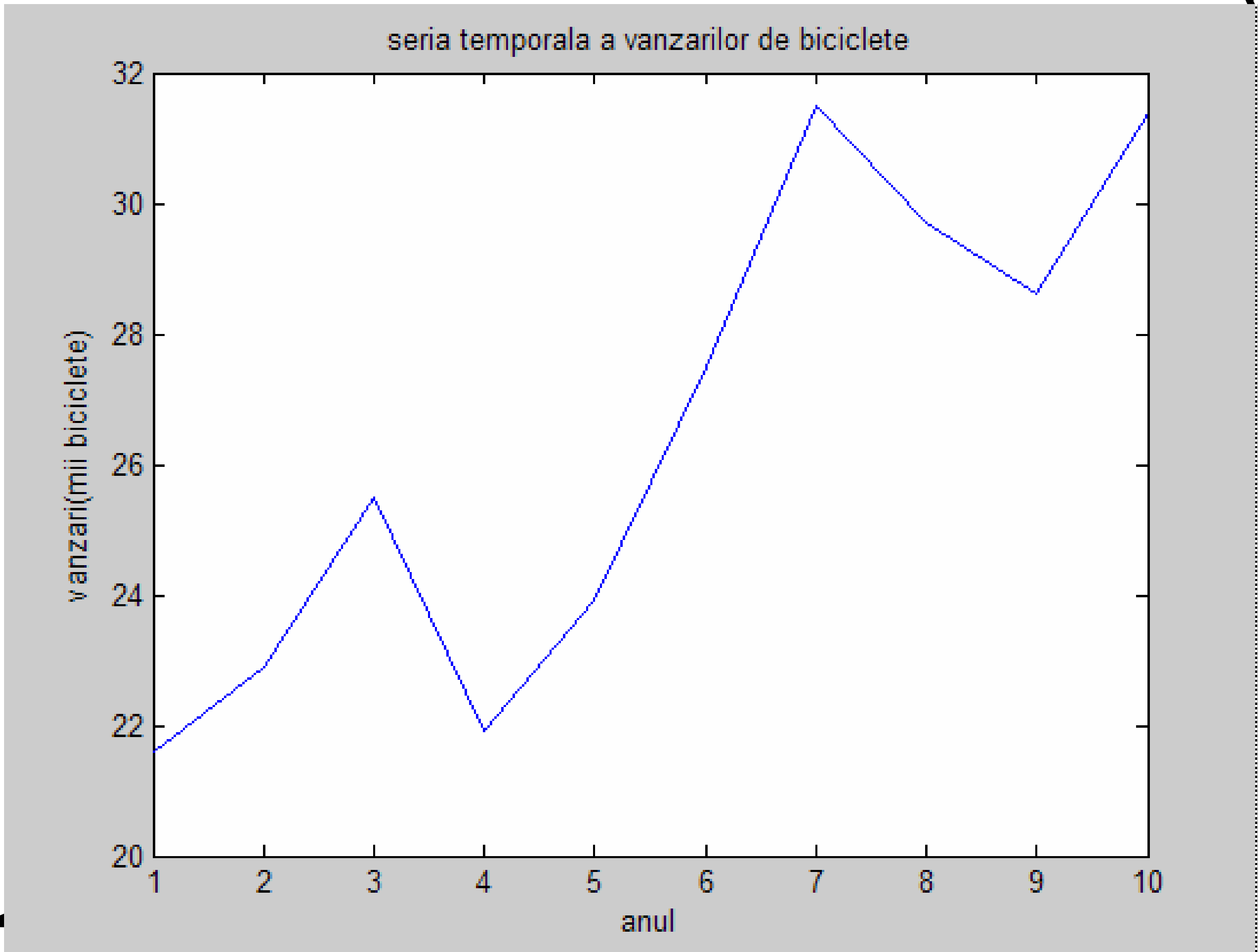
Diferența între *MSE* şi *MAD* constă că datorită faptului că *MSE* este suma pătratelor erorilor, aceasta este influențată de erorile mari de prognoză. Specialiştii au păreri contradictorii asupra alegerii uneia sau alteia dintre cele două metode.
Preferăm utilizarea *MSE*.

Prognoza utilizând trendul- exemplu

Vom prognoza valorile unei serii temporale care prezintă timp îndelungat un trend liniar.

Vânzările de biciclete a unei fabrici în ultimii 10 ani sunt:

anul	1	2	3	4	5	6	7	8	9	10
vânzări (mii)	21.6	22.9	25.5	21.9	23.9	27.5	31.5	29.7	28.6	31.4


seria temporala a vanzarilor de biciclete
vanzari(mii biciclete)
anul
32
30
28
26
24
22
20
1
2
3
4
5
6
7
8
9
10

Utilizarea regresiei liniare pentru identificarea trend-ului liniar

Pentru un trend liniar volumul vânzărilor exprimate ca funcție de timp este:

$$T_t = b_0 + b_1 t$$

unde:

T_t este valoarea trend-ului pentru vânzările de biciclete în perioada de timp t

b_0 este interceptorul liniei trend-ului

b_1 este panta liniei trend-ului

Conform formulelor pentru calculul dreptei de regresie, avem:

$$b_1 = \frac{\sum\limits_{t\in T} t\cdot\xi_t - \frac{1}{n}\sum\limits_{t\in T} t\cdot\sum\limits_{t\in T}\xi_t}{\sum\limits_{t\in T} t^2 - \frac{1}{n^2}\left(\sum\limits_{t\in T} t\right)^2}\ ;\ b_0 = \bar{\xi} - b_1\cdot\bar{t}$$

unde:

ξ_t valoarea actuală a seriei temporale in perioada de timp t

n numărul perioadelor de timp

$\bar{\xi}$ media valorilor seriei temporale, $\bar{\xi} = \frac{1}{n}\sum\limits_{t\in T}\xi_t$

$\bar{t}$ media valorilor lui t, adică $\bar{t} = \frac{1}{n}\sum\limits_{t\in T} t$

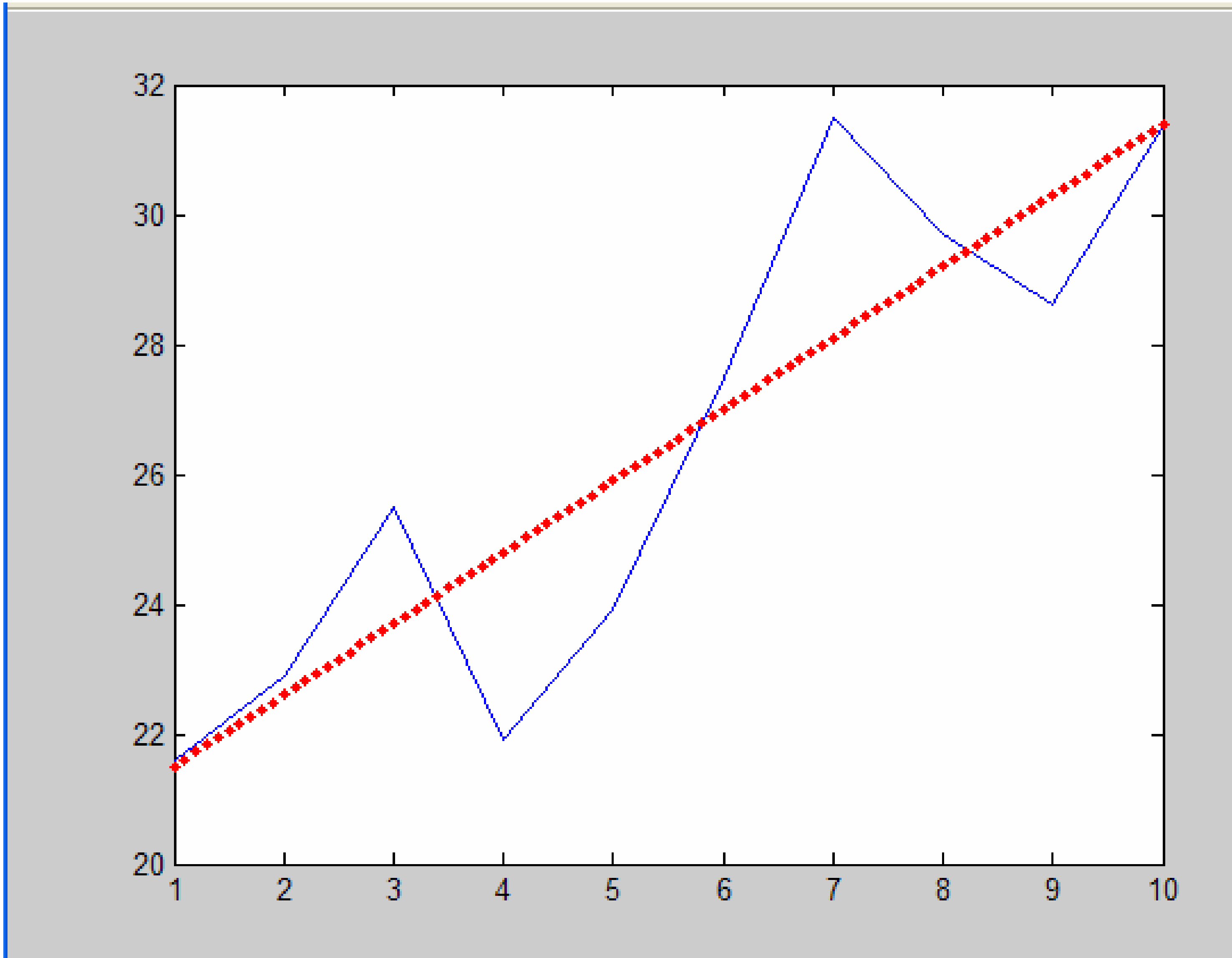
	t	ξ_t	$t \cdot \xi_t$	t^2
	1	21.6	21.6	1
	2	22.9	45.8	4
	3	25.5	76.5	9
	4	21.9	87.6	16
	5	23.9	119.5	25
	6	27.5	165	36
	7	31.5	220.5	49
	8	29.7	237.6	64
	9	28.6	257.4	81
	10	31.4	314	100
TOTAL	55	264.5	1545.5	385

$$\bar{t} = \frac{55}{10} = 5.5; \quad \bar{\xi} = \frac{264.5}{10} = 26.45; b_1 = \frac{1545.5 - \frac{1}{10} \cdot 55 \cdot 264.5}{385 - \frac{(55)^2}{10}} = 1.1;$$

$$b_0 = 26.45 - 1.1 \cdot 5.5 = 20.4$$

ecuația trend-ului liniar pentru vânzările de biciclete:

$$T_t = 20.4 + 1.1 \cdot t$$



Considerând că vânzările din ultimii 10 ani sunt un bun indicator pentru vânzările viitoare, putem spune că ecuația

$$T_t = 20.4 + 1.1 \cdot t$$

poate fi utilizată pentru a prognoza valori viitoare.
Astfel în al 11-lea an proiecția trend-ului ne spune că

$$T_{11} = 20.4 + 1.1 \cdot 11 = 32.5$$

adică putem prognoza că s-ar vinde 32 500 biciclete.

Anumite serii au trend-ul neliniar, cazuri de care nu ne ocupăm în acest context.

Prognoza utilizând trendul şi componenta sezonieră

Primul pas constă în eliminarea componentei sezoniere, după care vom face prognoza utilizând trend-ul.
Pasul următor constă în reintroducerea componentei sezoniere în scopul ajustării prognozei.

Seria temporala are un trend *T*, o componenta sezonieră *S* şi o componentă incidentală *I*.
Componentă incidentală este cea care dă efectele aleatoare din seria temporală, efecte care nu pot fi explicate de trend sau componenta sezonieră.

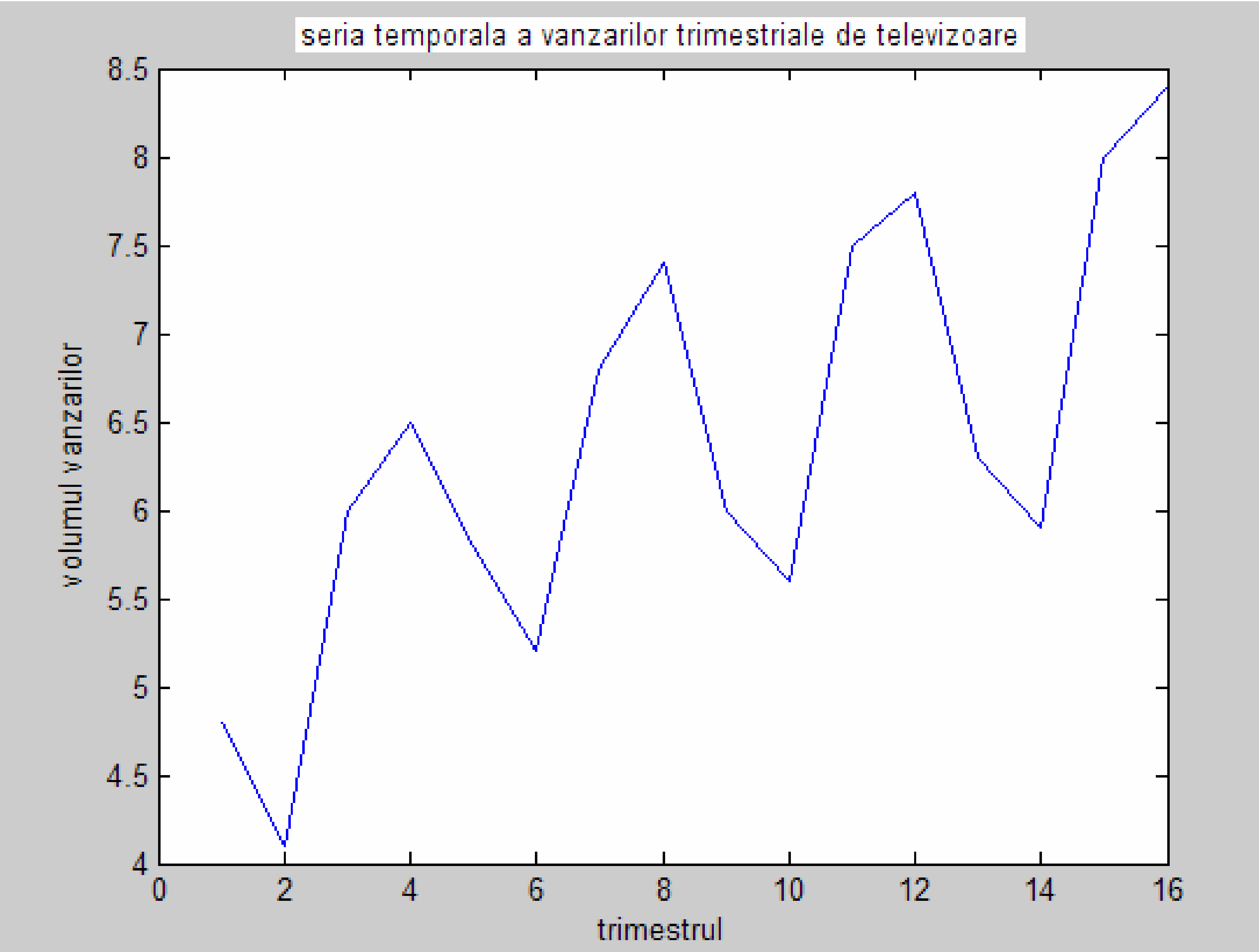
Notând T_t, S_t, I_t trend-ul, componenta sezonieră şi cea incidentală corespunzătoare perioadei de timp *t*, presupunem că valoarea reală ξ_t poate fi descrisă prin *modelul multiplicativ* de serie temporală:

$$\xi_t = T_t \cdot S_t \cdot I_t$$

exemplu

Vânzările (în mii de unități) de televizoare ale unei firme, din ultimii 4 ani, detaliate pe trimestre.

anul	trimestrul	vânzări	anul	trimestrul	vânzări
1	1 (1)	4.8	3	1 (9)	6
	2 (2)	4.1		2 (10)	5.6
	3 (3)	6		3 (11)	7.5
	4 (4)	6.5		4 (12)	7.8
2	1 (5)	5.8	4	1 (13)	6.3
	2 (6)	5.2		2 (14)	5.9
	3 (7)	6.8		3 (15)	8
	4 (8)	7.4		4 (16)	8.4


seria temporala a vanzarilor trimestriale de televizoare
volumul vanzarilor
trimestrul
8.5
8
7.5
7
6.5
6
5.5
5
4.5
4
0
2
4
6
8
10
12
14
16

Vom calcula indicii sezonieri:

Se observă că în trimestrul al doilea al fiecărui an (adică al 2-lea, al 6-lea, al 10-lea, al 14-lea) volumul vânzărilor este minim în timp ce în trimestrele 3 şi 4 al fiecărui an volumul creşte, ceea ce indică existența unui pattern sezonier.

Vom identifica computațional fiecare influență trimestrială sezonieră:

- Calculăm o medie mobilă pentru a izola componenta sezonieră S şi componen
  incidentală I

  Folosim datele din fiecare an şi anume cele 4 trimestre:

$$media\ mobila(1) = \frac{4.8 + 4.1 + 6 + 6.5}{4} = 5.35$$

valoarea 5.35 corespunde ultimei jumătăți a trimestrului al 2-lea şi primei jumătăți a trimestrului al 3-lea.

$$media\ mobila(2) = \frac{4.1 + 6 + 6.5 + 5.8}{4} = 5.6$$

valoarea 5.6 corespunde ultimei jumătăți a trimestrului al 3-lea şi primei jumătăți a trimestrului al 4-lea.

Cu ajutorul mediilor mobile dorim să izolăm componenta sezonieră şi cea încidentală. Dar valorile calculate nu corespund trimestrelor din seria temporală.

De exemplu 5.35 corespunde primei jumătăți a trimestrului al 3-lea în timp ce 5,6 corespunde ultimei jumătăți a trimestrului al 3-lea.
In aceste condiții vom considera media mobilă corespunzătoare trimestrului al 3-lea ca fiind $\frac{5.35+5.6}{2}=5.475$

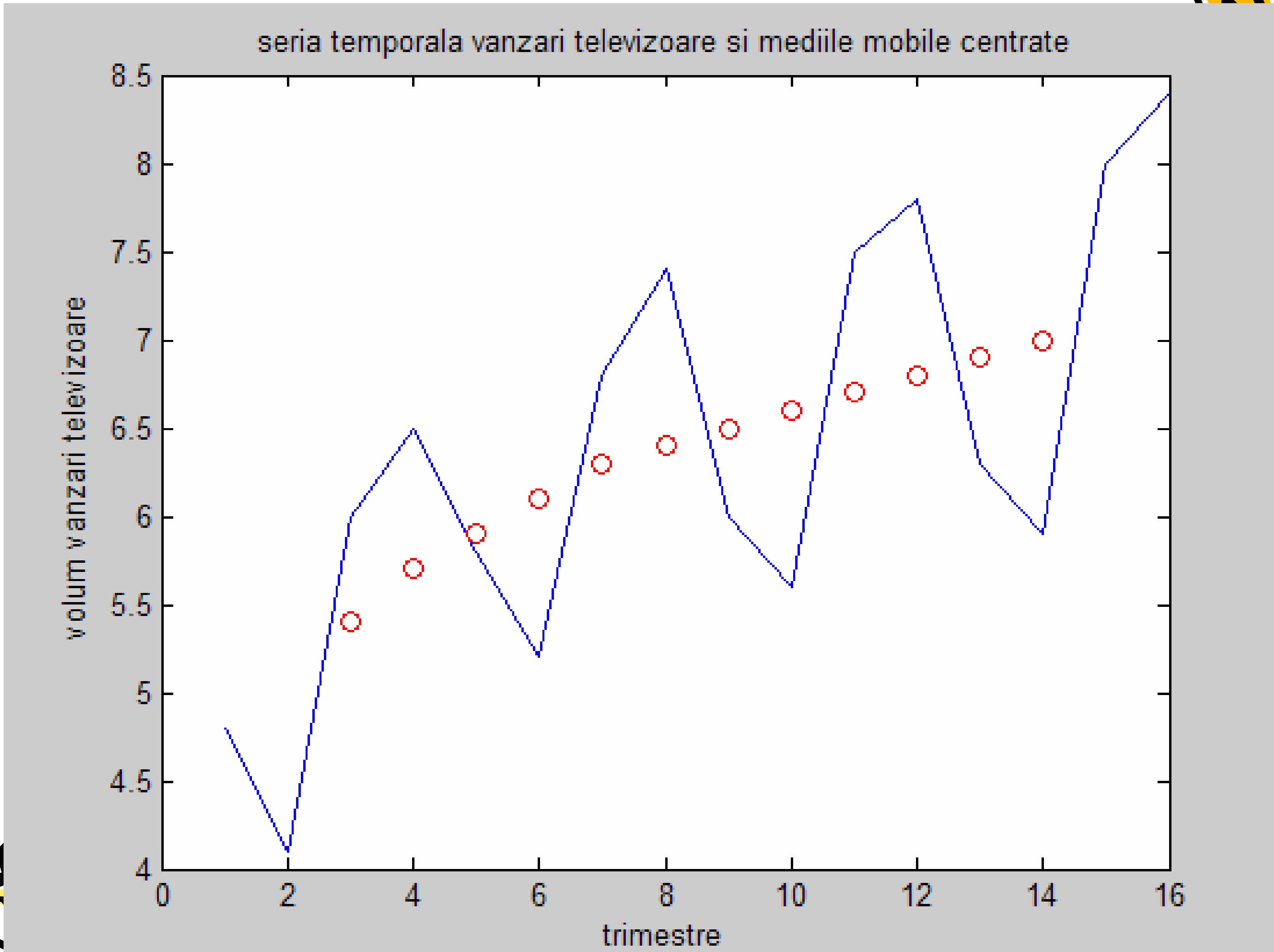
Analog asociem trimestrului al 4-lea media mobilă $\frac{5.6+5.875}{2}=5.7375$.

Aceste rezultate se numesc medii mobile centrate (*centered moving average*)
În tabelul următor prezentam mediile mobile şi mediile mobile centrate în cazul vânzărilor de televizoare.

anul	trimestrul	vânzări	media mobilă	media mobilă centrată
1	1 (1)	4.8		
	2 (2)	4.1		
	3 (3)	6	5.3500	5.4750
	4 (4)	6.5	5.600	5.7375
2	1 (5)	5.8	5.8750	5.9750
	2 (6)	5.2	6,0750	6.1875
	3 (7)	6.8	6.3000	6.3250
	4 (8)	7.4	6.3500	6.4000
3	1 (9)	6	6.4500	6.5375
	2 (10)	5.6	6.6250	6.6750
	3 (11)	7.5	6.7250	6.7625
	4 (12)	7.8	6.8000	6.8375
4	1 (13)	6.3	6,8750	6.9375
	2 (14)	5.9	7.0000	7.075
	3 (15)	8	7.1500	
	4 (16)	8.4		

Dacă numărul de puncte din media mobilă este impar punctul din mijloc va corespunde unei perioade din seria temporală, caz în care nu va mai fi nevoie să centrăm media mobilă.

Prezentăm graficul seriei temporale şi graficul seriei temporale corespunzătoare mediilor mobile centrate.


seria temporala vanzari televizoare si mediile mobile centrate
volum vanzari televizoare
trimestre
8.5
8
7.5
7
6.5
6
5.5
5
4.5
4
0
2
4
6
8
10
12
14
16

Se observă că seria temporală corespunzătoare mediilor mobile centrate este „netezită" în sensul că nu apar influențele sezoniere şi incidentale.
Impărțind fiecare observație din seria temporală cu media mobilă centrată corespunzătoare putem identifica efectul sezonier -incidental al seriei
 temporale

De exemplu in al 3-lea trimestru al primului an $\dfrac{6}{5.475}=1.0959$ este valoarea componentei sezonier -incidentale.

Prezentăm valorile componentei sezonier -incidentale pentru întreaga serie temporală

anul	trimestrul	vânzări	media mobilă centrată	componenta sezonier-incidentală
1	1 (1)	4.8		
	2 (2)	4.1		
	3 (3)	6	5.4750	1.0959
	4 (4)	6.5	5.7375	1.1329
2	1 (5)	5.8	5.9750	0.9707
	2 (6)	5.2	6.1875	0.8404
	3 (7)	6.8	6.3250	1.0751
	4 (8)	7.4	6.400	1.1563
3	1 (9)	6	6.5375	0.9178
	2 (10)	5.6	6.6750	0.8390
	3 (11)	7.5	6.7625	1.1091
	4 (12)	7.8	6.8375	1.408
4	1 (13)	6.3	6.9375	09081
	2 (14)	5.9	7.0750	0.8339
	3 (15)	8		
	4 (16)	8.4		

Să considerăm componenta sezonier-incidentală pentru al 3-lea trimestru din primii 3 ani: 1.0959, 1.0751, 1.1091.
În aceste cazuri componenta sezonier-incidentală are o influență peste medie. Deoarece putem presupune că fluctuațiile de la an la an ale componentei sezonier-incidentale se datorează în principal componentei incidentale, putem calcula media celor 3 valori, eliminând astfel influența incidentală și a obține o estimare a influenței sezoniere din al 3-lea trimestru.

$$efectul\ sezonier\ in\ trimestrul\ al\ 3-lea =$$

$$= \frac{1.0959 + 1.0751 + 1.1091}{3} = 1.0997$$

Efectul sezonier din al 3-lea trimestru este cunoscut sub numele de **indicele sezonier al trimestrului 3.**

Prezentăm indicii sezonieri ale celor 4 trimestre:

trimestru	valorile componentei sezonier-incidentale	indicele sezonier
1	0.9707, 0.9178, 0.9081	0.9322
2	0.8404; 0.8390; 0.8339	0.8378
3	1.0959; 1.0751; 1.1091	1.0933
4	1.1329; 1.1563; 1.1408	1.1433

Din tabel observăm că cele mai bune vânzări sunt în al 4-lea trimestru, cu 14.33% peste medie.

Cel mai slab trimestru pentru vânzări este al 2-lea cu 16.22% submedie.

Componenta sezonieră corespunde aşteptărilor intuitive: al 4-lea trimestru, cu vreme foarte rece, cu puține activități în afara casei predispune la urmărirea programelor TV, in timp ce al 2-lea trimestru presupune activități în aer liber şi ignorarea programelor TV.

Modelul multiplicativ cere ca media indicilor sezonieri să fie 1, condiție ce nu este îndeplinită în exemplul nostru.
Pentru a obține media indicilor sezonieri egală cu unitatea este necesară o ajustare, prin împărțirea fiecărui indice sezonier la media indicilor sezonieri.

trimestru	indicele sezonier	indicele sezonier ajustat
1	0.9322	0.9307
2	0.8378	0.8364
3	1.0933	1.0915
4	1.1433	1.1414

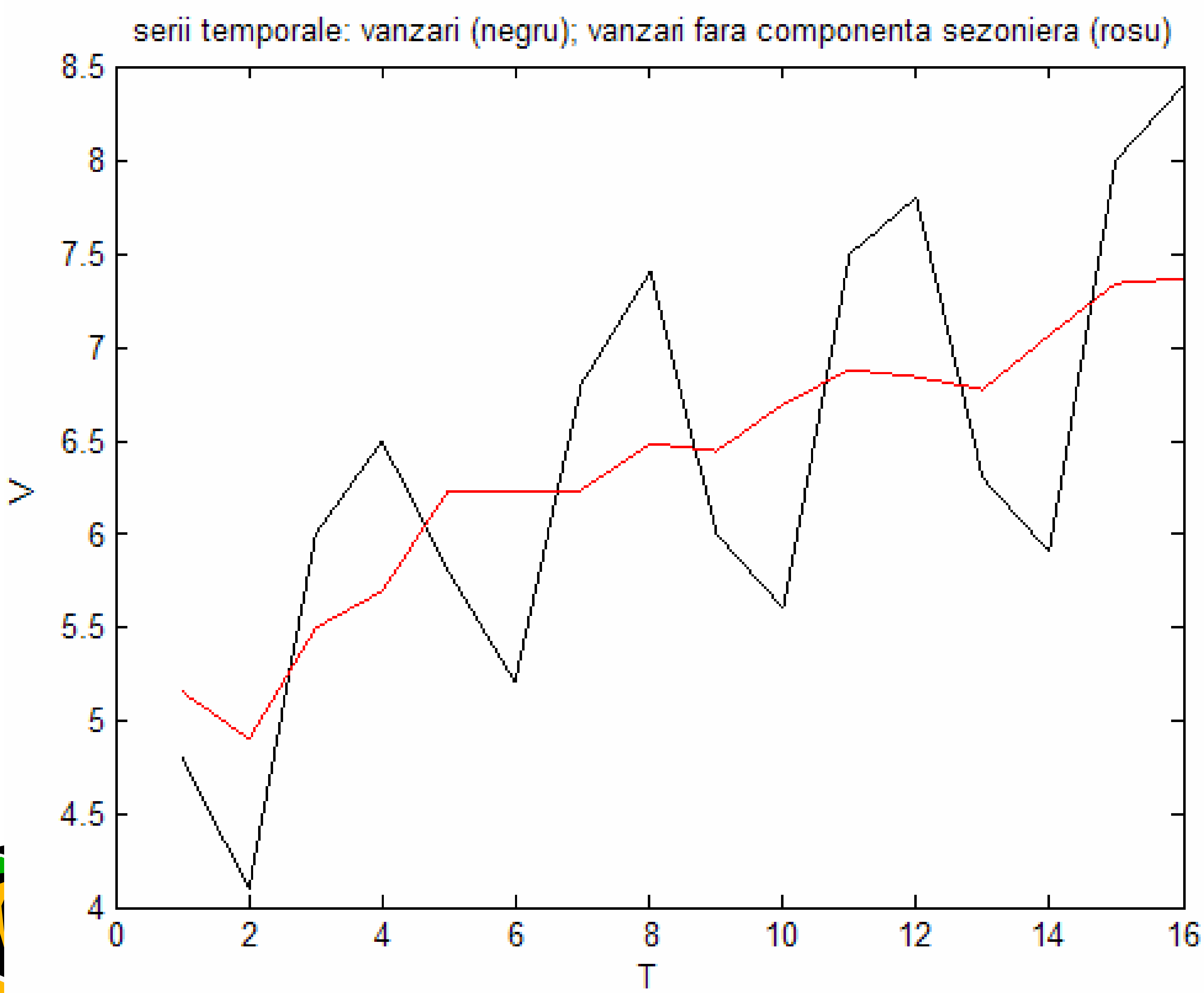
Suntem interesați să eliminăm efectele sezoniere ale unei serii temporale. Dacă reuşim acest lucru, putem de exemplu compara vânzările în perioade succesive. Dacă nu, putem face comparații intre vânzările dintr-o anumită perioadă din acest an şi aceiaşi perioadă a anului trecut.

În modelul multiplicativ avem:

$$\xi_t = T_t \cdot S_t \cdot I_t$$

Impărțind fiecare observație (termen) al seriei temporale la indicele sezonier corespunzător înlăturăm efectul sezonier al seriei temporale.
Prezentăm un tabel cu seria vânzărilor de televizoare, cu efectul sezonier înlăturat şi graficul acesteia.

anul	trimestrul	vânzări (ξ_i)	index sezonier (S_i)	seria temporală fără componenta sezonieră $\frac{\xi_i}{S_i} = T_i \cdot I_i$
1	1 (1)	4.8	0.9307	5.1567
	2 (2)	4.1	0.8364	4.9021
	3 (3)	6	1.0915	5.4968
	4 (4)	6.5	1.1414	5.6947
2	1 (5)	5.8	0.9307	6.2321
	2 (6)	5.2	0.8364	6.2173
	3 (7)	6.8	1.0915	6.2297
	4 (8)	7.4	1.1414	6.4832
3	1 (9)	6	0.9307	6.4470
	2 (10)	5.6	0.8364	6.6956
	3 (11)	7.5	1.0915	6.8710
	4 (12)	7.8	1.1414	6.8336
4	1 (13)	6.3	0.9307	6.7694
	2 (14)	5.9	0.8364	7.0542
	3 (15)	8	1.0915	7.3291
	4 (16)	8.4	1.1414	7.3593


serii temporale: vanzari (negru); vanzari fara componenta sezoniera (rosu)
V
T
8.5
8
7.5
7
6.5
6
5.5
5
4.5
4
0
2
4
6
8
10
12
14
16

Utilizând seria temporală fără componenta sezonieră putem scrie estima volumul vânzărilor ca funcție de timp.

$$T_t = b_0 + b_1 t$$

unde

$$b_1 = \frac{\sum_{t\in T} t \cdot \xi_t^1 - \frac{1}{n}\sum_{t\in T} t \cdot \sum_{t\in T} \xi_t^1}{\sum_{t\in T} t^2 - \frac{1}{n^2}\left(\sum_{t\in T} t\right)^2}\ ;\ b_0 = \bar{\xi} - b_1 \cdot \bar{t}$$

unde $\{\xi_t^1, t \in T\}$ este seria temporală fără componenta sezonieră.

$$T_t = 5.101 + 0.148 \cdot t$$

este trend-ul liniar al seriei temporale.

Putem prognoza volumul vânzărilor în primele două trimestre al celui de-a 5-lea an (al 17-lea şi al 18-lea trimestru)

$$T_{17} = 5.101 + 0.148 \cdot 17 = 7.617$$

$$T_{18} = 7.765; T_{19} = 7.913; T_{20} = 8.061$$

Aceste prognoze au fost obținute pe baza trend-ului.
Pentru a lua în considerare efectul sezonier asupra prognozei, înmulțim T_t cu indicele sezonier ajustat, de exemplu $T_{17} \cdot 0.9307 = 7.0973$

Prezentăm în următorul tabel prognoza vânzărilor în trimestrele celui de al 5-lea an:

anul	trimestru	prognoza (trend)	indice sezonier	prognoza
5	1	7.6135	0.9307	7.0973
	2	7.7609	0.8364	6.5018
	3	7.9083	1.0915	8.6465
	4	8.0557	1.1414	9.2101

Prezentăm rezolvarea problemei în Matlab:

>> V=[4.8 4.1 6 6.5 5.8 5.2 6.8 7.4 6 5.6 7.5 7.8 6.3 5.9 8 8.4];% **V vanzari trimestriale**

>> for i=5:17 M(i-2)=1/4*(V(i-4)+V(i-3)+V(i-2)+V(i-1));end % **M media mobila trimestriala**

>>M(16)=0

>> for i=3:14 MC(i)=1/2*(M(i)+M(i+1)); end % **MC media mobila centrata**

>> for i=15:16 MC(i)=0;end

>> for i=3:14 CS(i)=V(i)/MC(i);end

>> for i=15:16 CS(i)=0; end %**CS componenta sezoniera**

```
>> [V' M' MC' CS']
ans =
    4.8000         0         0         0
    4.1000         0         0         0
    6.0000    5.3500    5.4750    1.0959
    6.5000    5.6000    5.7375    1.1329
    5.8000    5.8750    5.9750    0.9707
    5.2000    6.0750    6.1875    0.8404
    6.8000    6.3000    6.3250    1.0751
    7.4000    6.3500    6.4000    1.1563
    6.0000    6.4500    6.5375    0.9178
    5.6000    6.6250    6.6750    0.8390
    7.5000    6.7250    6.7625    1.1091
    7.8000    6.8000    6.8375    1.1408
    6.3000    6.8750    6.9375    0.9081
    5.9000    7.0000    7.0750    0.8339
    8.0000    7.1500         0         0
    8.4000    7.1500         0         0
```

```
>> for i=1:2 I(i)=1/3*(CS(i+4)+CS(i+8)+CS(i+12));end
>> for i=3:4 I(i)=1/3*(CS(i)+CS(i+4)+CS(i+8));end % I indicele sezonier
>> I
I =
    0.9322    0.8378    1.0933    1.1433
>> Ia=I/mean(I) %Ia indicele sezonier ajustat
Ia =
    0.9307    0.8364    1.0915    1.1414
>>  IS=[Ia Ia Ia Ia];
>>  for i=1:16 VDS(i)=V(i)/IS(i);end % VDS vanzari cu factorul sezonier indepartat
```

```
>>  [V' VDS']
ans =
    4.8000    5.1576
    4.1000    4.9021
    6.0000    5.4968
    6.5000    5.6947
    5.8000    6.2321
    5.2000    6.2173
    6.8000    6.2297
    7.4000    6.4832
    6.0000    6.4470
    5.6000    6.6956
    7.5000    6.8710
    7.8000    6.8336
    6.3000    6.7694
    5.9000    7.0542
    8.0000    7.3291
    8.4000    7.3593
```

```
>> T=1:16;b1=(T*VDS'-16*mean(T)*mean(VDS))./(norm(T)^2-16*mean(T)^2)
b1 =
    0.1474
>> b0=mean(VDS)-b1*mean(T)
b0 =
    5.1080
>>  for t=17:20; Vp(t)=b0+b1*t;end %Vp vanzari prognozate
>>  for i=1:4 VP(i+16)=Vp(i+16)*I(i);end % VP vanzari prognozate tinand cont
de indicele sezonier
```

```
>>[ Vp'  VP']
ans =
   7.6135    7.0973
   7.7609    6.5018
   7.9083    8.6465
   8.0557    9.2101
```

In cazul analizei unei serii temporale trebuie să avem în vedere şi următoarele aspecte:

- *Distincţia* între observaţiile pe *timp scurt* şi cele pe *timp îndelungat* se referă la separarea dintre relaţiile persistente în timp, observate la datele culese şi relaţiile conjuncturale.
- *Relaţia de cauzalitate* se poate observa între două sau mai multe serii temporale.

  In cazul determinării unei relaţii de cauzalitate, se studiază factorul de defazare care intervine între cauză şi efect, privind seriile implicate.

Din multitudinea de *modele temporale* (*dinamice*) bazate pe seriile temporale, vom aminti doar trei tipuri principale:

- modelele de ajustare;
- modelele autopredictive;
- modelele explicative.

Modele de ajustare

Pe baza observațiilor obținute analizând datele reale, putem formula un model matematic ilustrat de o ecuație de forma:

$$\xi_t = f(t, u_t) ,$$

unde f este o funcție ce conține un număr finit de parametri necunoscuți, iar u_t reprezintă o variabilă aleatoare de medie zero, aleasă în funcție de situația reală modelată.

Ipotezele asupra variabilei u_t ca și estimarea funcției f se fac plecând de la așa-numitele *ajustări globale*, în care toate observațiile se bucură de aceeași considerație, având roluri egale în estimații, sau așa-numitele *ajustări locale*, în care fiecare observație își are rolul său în determinarea parametrilor modelului.

Modele autopredictive

Se presupune că prezentul este influențat de trecut, deci matematic vorbind, un asemenea model este ilustrat de o ecuație de forma:

$$\xi_t = f(\xi_{t-1}, \xi_{t-2}, \ldots, u_t),$$

unde u_t reprezintă și aici factorul de disturbare, fiind o variabilă aleatoare.

Modele explicative

Ecuația matematică este:

$\xi_t = f(x_t, u_t)$

unde x_t este o variabilă observabilă, numită și *variabilă exogenă*, iar u_t reprezintă din nou factorul disturbator.

În principal, aceste modele se împart în două tipuri:

- *statice*, în care variabila exogenă x_t nu conține informații din trecutul lui ξ_t, iar u_t sunt mutual independente;
- *dinamice*, în care fie x_t conține informații privind trecutul lui ξ_t, fie u_t sunt auto-corelate.

ARIMA

Modelul *ARIMA (Auto-Regressive Integrated Moving Average)* a fost dezvoltat de către Box şi Jenkins (1976), devenind unul dintre cele mai utilizate modele de prognoză, datorită puterii şi flexibilităţii dovedite în aplicaţiile concrete.

Menţionăm că modelul necesită totuşi o pregătire serioasă din partea utilizatorului, în ciuda simplităţii matematice aparente.

Aşa cum arată şi numele, ARIMA include atât o componentă autoregresivă cât şi una de tip medie mobilă.

Vom trece în revistă principalele caracteristici ale celor două componente (sub-modele).

procese autoregresive (autoregressive processes)

S-a observat că, în general, datele din seriile temporale sunt dependente secvențial, cu alte cuvinte prognoza coeficienților pentru comportamentul din perioada următoare depinde de cei precedenți (cu o anumită decalare).

Din punct de vedere formal, matematic, această aserțiune este ilustrată de ecuația:

$$x_t = \xi + \phi_1 \cdot x_{t-1} + \phi_2 \cdot x_{t-2} + \ldots + \varepsilon,$$

unde

- ξ reprezintă o constantă (*interceptorul*),
- ϕ_i reprezintă parametrii autoregresivi ai modelului,
- ε este eroarea (aleatoare) a modelului.

Parametrii modelului trebuie să se găsească în anumite intervale mărginite şi nu pe întreaga axă reală, de exemplu [-1, 1], altfel valorile prea mari ale acestora pot face ca valorile prognozate să tindă spre infinit.

procese cu medie mobila (moving average processes)

Fiecare element al seriei temporale poate fi influențat de erorile din trecut, făcute la fiecare pas, modelul matematic având ecuația:

$$x_t = \mu + \varepsilon_t - \theta_1 \cdot \varepsilon_{t-1} - \theta_2 \cdot \varepsilon_{t-2} - \ldots,$$

unde

- μ este o constantă,
- $\theta_1, \theta_2, \ldots,$ sunt parametrii medie mobilă ai modelului,
- ε_{t-i} reprezintă erorile la pasul respectiv,
- iar ε_t este eroarea (aleatoare) la pasul *t* a modelului.

După cum se observă, spre deosebire de cazul anterior când valoarea prognozată a seriei temporale la pasul t era o combinație liniară a valorilor anterioare, aici ea este o combinație liniară a erorilor anterioare.

În modelul *ARIMA* avem trei tipuri de parametri:

- parametri autoregresivi (*p*);
- defazajul transformării diferențiate (*d*);
- parametri medie mobilă (*q*).

Astfel, un model ARIMA în notația Box & Jenkins va fi de forma ARIMA (p, d, q).

De exemplu, un model de tip ARIMA (1, 2, 1) înseamnă că se referă la $p = 1$ parametri autoregresivi, $d = 2$ defazajul și $q = 1$ parametri medie mobilă.

Sunt cazuri în care nu este nevoie de transformarea diferențiată, adică. $d = 0$.

Cel mai adesea aceasta este aleasă $d = 1$ sau $d = 2$, atunci când este într-adevăr necesară.

În ceea ce privește alegerea celorlalți doi parametri, valorile uzuale rareori depășesc 2.

random walk

ARIMA (0,1,0) = *mers la întâmplare* (*random walk*).

Ecuația modelului este dată de:

$$x_t = \xi + x_{t-1}$$

ARIMA (0,1,0) poate fi considerat un model regresiv degenerat, în care

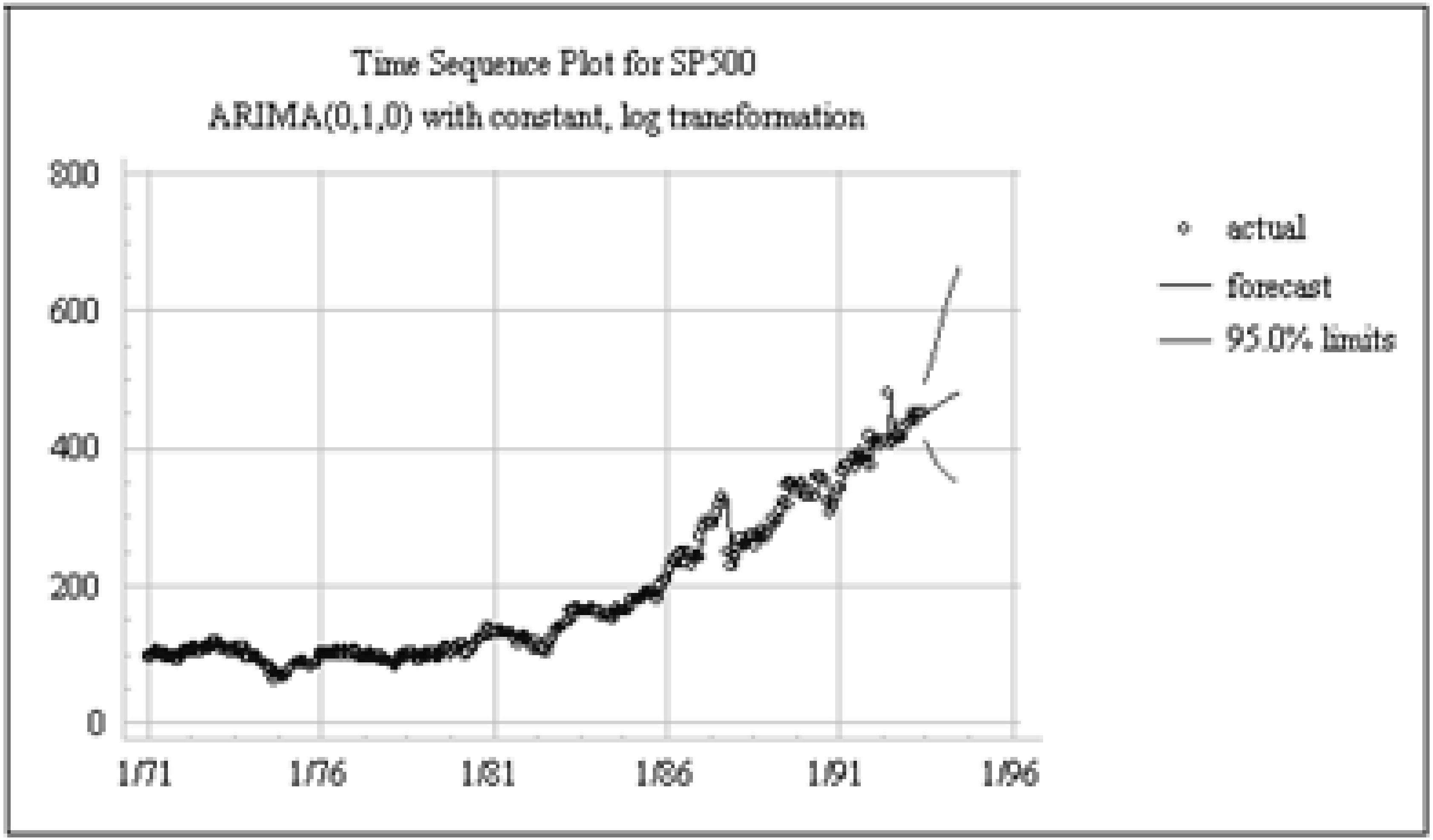
$$DIFF(x) = x_t - x_{t-1}$$

este variabila dependentă şi singura variabila independentă fiind constantă ξ .

Să presupunem că utilizăm un model de mers la întâmplare cu creştere (*random-walk-with-growth model*) sau mers la întâmplare geometric (*geometric random walk*) într-o problemă de prognoză a prețului acțiunilor după indexul Standard & Poors (S&P500 stock index) pe perioada 1971-1995, ilustrat în figura de mai jos.


S&P 500 Index Monthly Close
500
400
300
200
100
0
1/71
1/76
1/81
1/86
1/91
1/96
DATE
Variables
SP500

Prognoza dată de modelul ARIMA considerat în cazul prețului acțiunilor este prezentatăîn continuare, împreună cu intervalul corespunzător de încredere (95%).


Time Sequence Plot for SP500
ARIMA(0,1,0) with constant, log transformation
800
600
400
200
0
1/71
1/76
1/81
1/86
1/91
1/96
actual
forecast
95.0% limits

differenced first order auto-regressive model

ARIMA (1,1,0) = model autoregresiv cu diferențe de ordinul întâi

În cazul în care erorile obținute cu modelul mersului la întâmplare sunt autocorelate, se poate eventual rezolva problema prin adăugarea unei funcții a cărei variabilă este diferența dintre valorile variabilei dependente, cu o perioadă în urmă.

Obținem următoarea ecuație de predicție:

$$x_t - x_{t-1} = \xi + \phi(x_{t-1} - x_{t-2})$$

ecuație ce poate fi scrisă

$$x_t = \xi + x_{t-1} + \phi(x_{t-1} - x_{t-2}),$$

unde constanta este notată cu ξ și coeficientul autoregresiv cu ϕ.

ARIMA (0,1,1)

ARIMA (0,1,1) *fără constantă* = „netezire" simplu exponențială este un alt model utilizat pentru corectarea erorilor autocorelate din modelul mersului la întâmplare.

Pentru anumite serii temporale, spre exemplu o serie ce prezintă fluctuații de zgomot în jurul unei medii ce variază lent, utilizând modelul mersului la întâmplare nu se obțin rezultate la fel de bune cum ar fi cele obținute folosind media mobilă a valorilor trecute.

Aceasta înseamnă că în loc de a folosi cea mai recentă observație pentru a prognoza următoarea observație este preferabilă utilizarea mediei ultimelor observații în scopul de a filtra zgomotul şi a estima cu acuratețe media locală.

Ecuația de predicție a modelului este:

$$x_t = x_{t-1} - \theta \cdot \varepsilon_{t-1},$$

unde ε_{t-1} este eroarea la momentul $t-1$

ARIMA mixt (1,1,1)

Acest model are ecuația de prognoză:

$$x_t = \xi + x_{t-1} + \phi(x_{t-1} - x_{t-2}) + \theta \cdot \varepsilon_{t-1},$$

de unde se observă clar mixarea celor două componente (autoregresivă şi medie mobilă).

Uneori acest model poate duce la apariția fenomenului de „overfitting" al datelor.

În ceea priveşte estimarea parametrilor ce apar în ecuația de prognoză a modelului ARIMA, există mai multe metode.

Indiferent de metodă, în principiu se utilizează un algoritm, numit şi *metoda quasi-Newton*, ce maximizează verosimilitatea (probabilitatea) seriei prognozate în raport cu seria observată, în funcție de valorile propuse pentru parametri, producând estimări ale valorilor optime pentru parametri.

Practic, metodologia determinării valorilor parametrilor implică tot calculul reziduurilor urmat de obținerea sumei pătratelor reziduurilor, care trebuie minimizată pentru obținerea celor mai buni parametri.

exemple

- influențează în vreun fel (și dacă da, cum?) o nouă politică fiscală importurile?
- o nouă abordare a Ministerului Sănătății privind lista medicamentelor compensate influențează și în ce măsură profitabilitatea farmaciilor?

În acest context, ideea este de evalua impactul acțiunii evenimentului în cauză asupra valorilor seriei temporale analizate, privită ca *serie temporală întreruptă.*

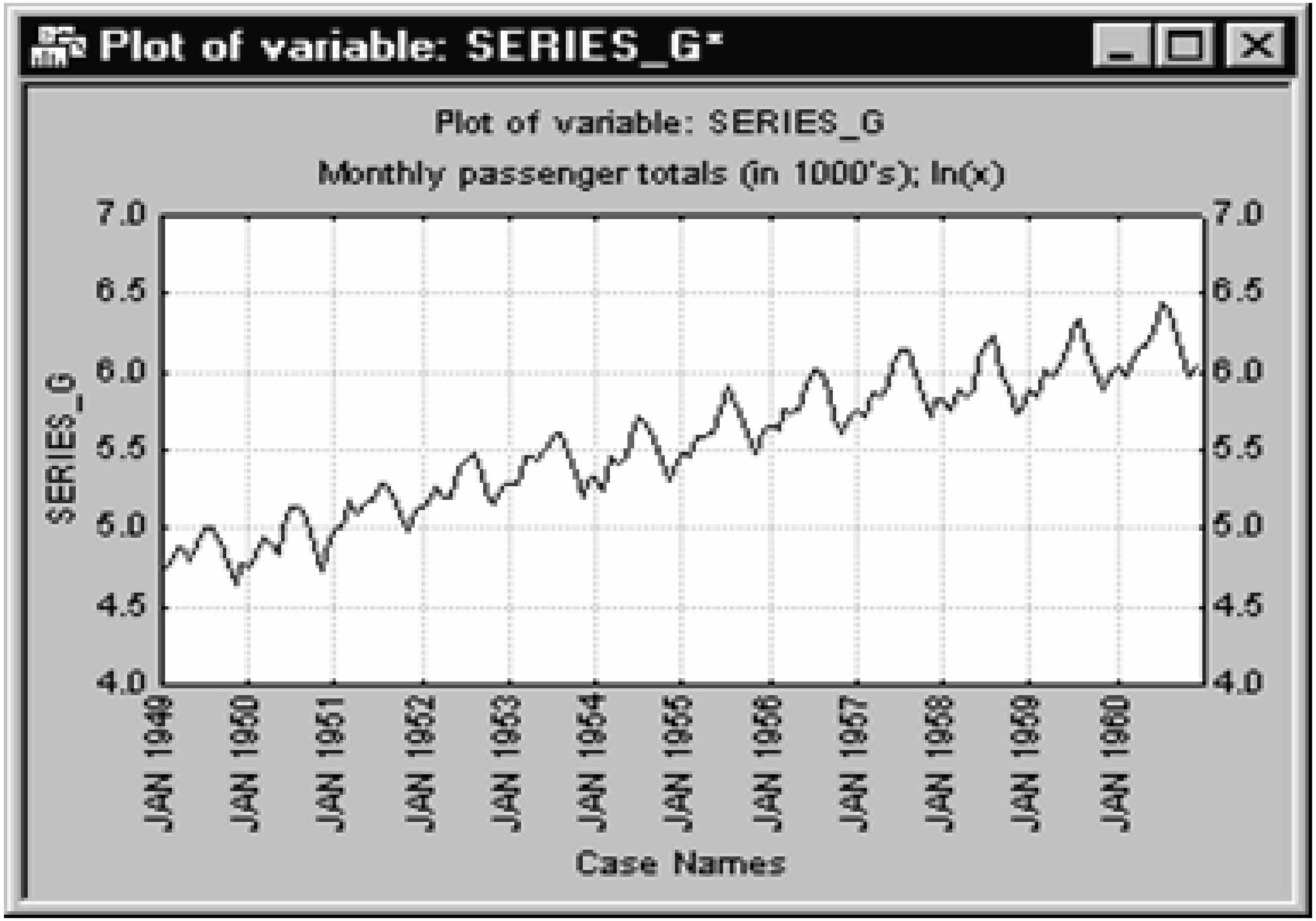
Există în această direcție modele de prognoză, dintre acestea menționând *ARIMA* pentru serii temporale întrerupte

aplicații

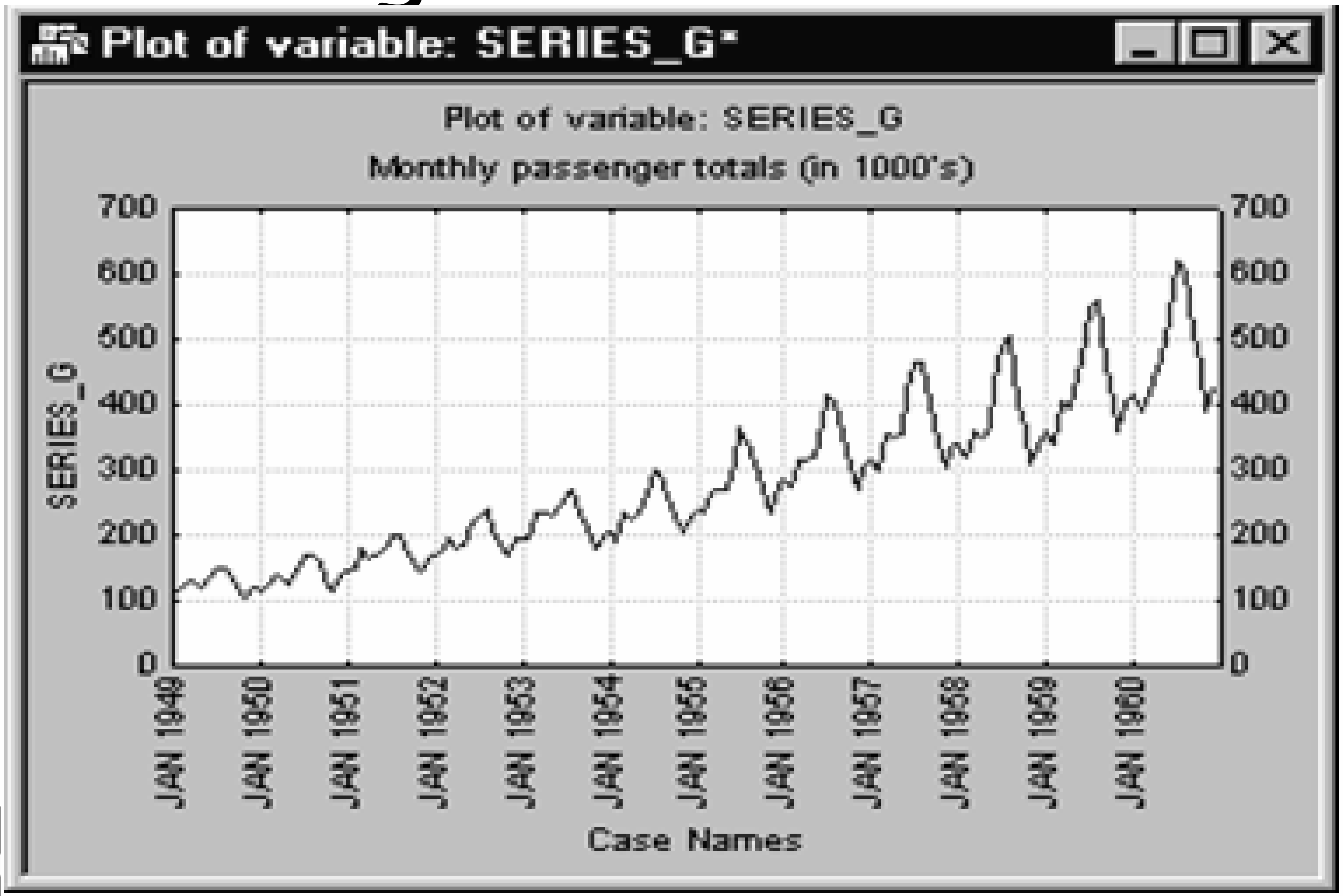
Reluăm cazul referitor la seria temporală reprezentând numărul de pasageri ai curselor aeriene internaționale, înregistrat în perioada 1949-1960

Mai întâi, seria temporală inițială va fi logaritmată pentru a stabiliza variabilitatea sa, generată de amplitudinea schimbărilor periodice, care este crescătoare și ar putea influența prognoza.

seria logaritmata

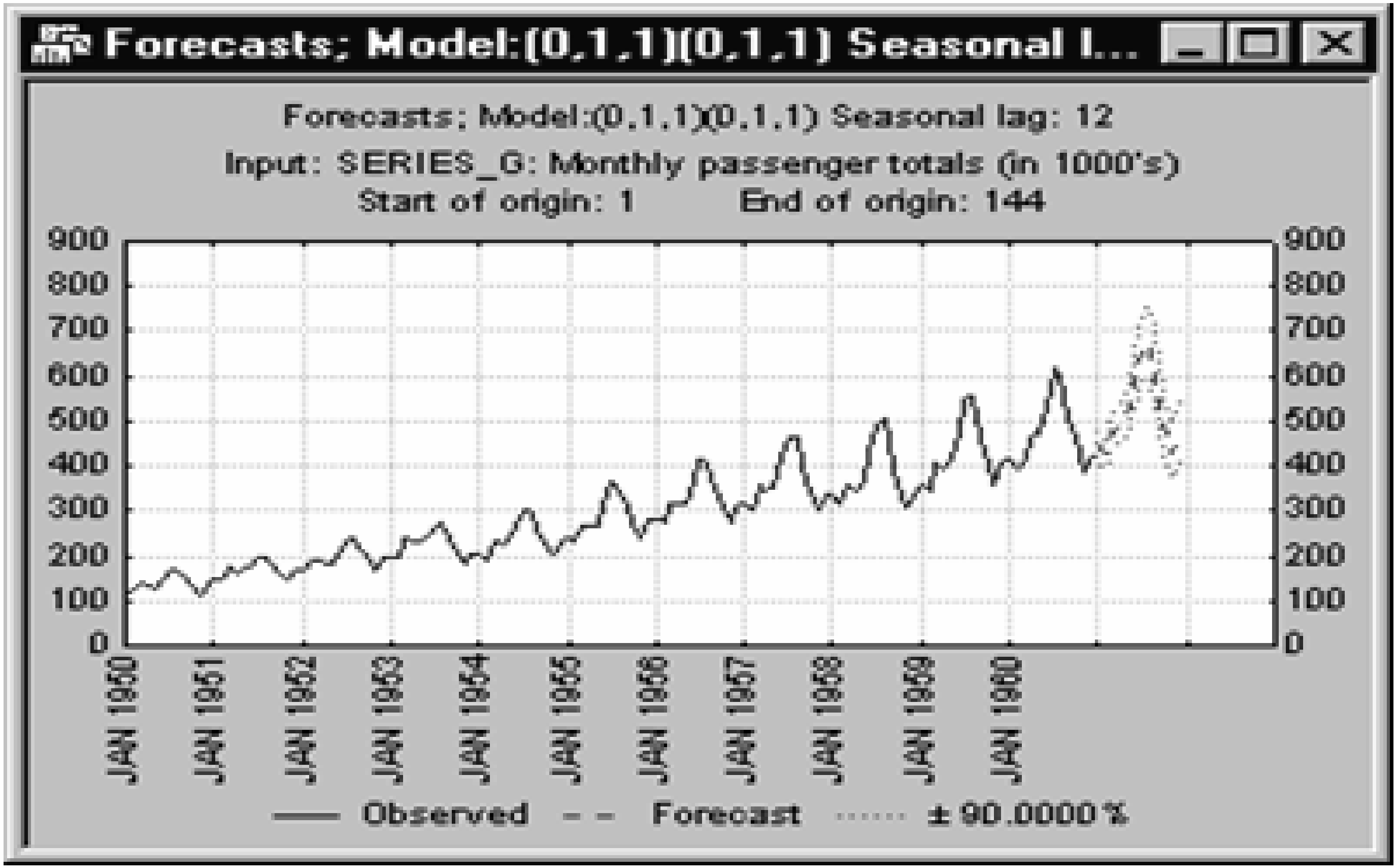


seria originara

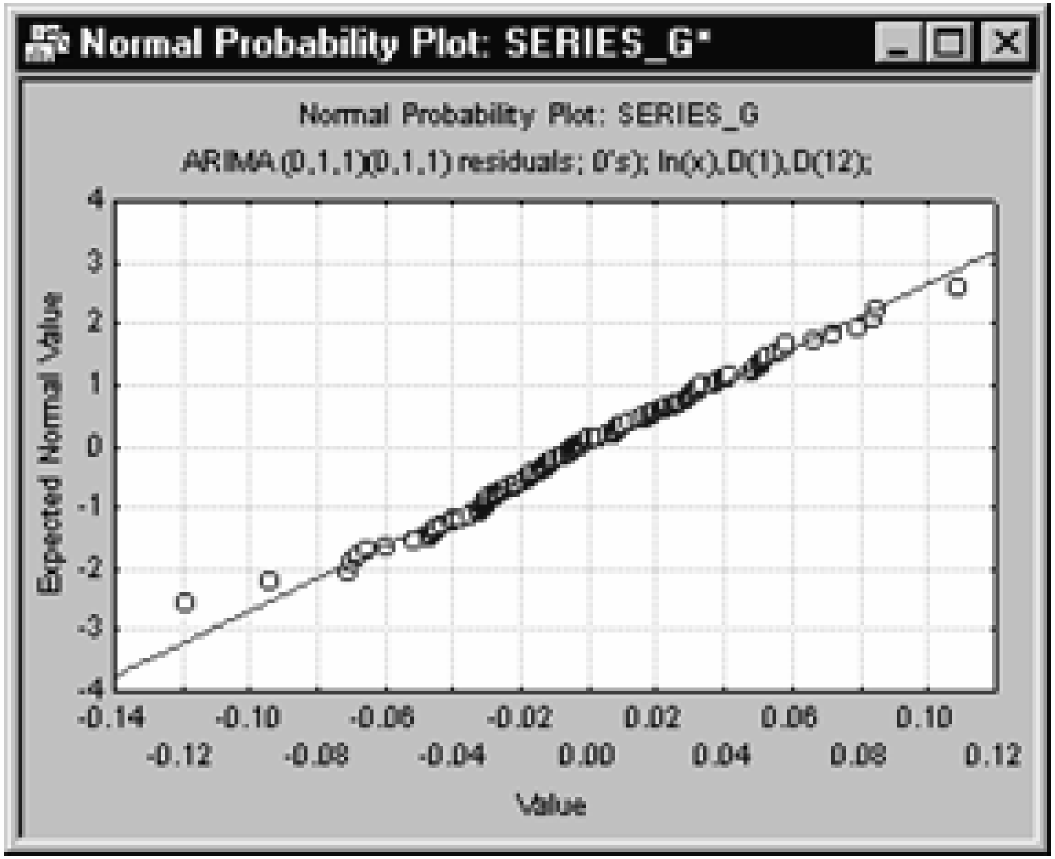


Comparând graficele,al seriei temporale originare cu cel al al seriei logaritmate, se observă fenomenul de stabilizare al ei, în sensul că la seria logaritmată amplitudinea creşterilor este stabilă.

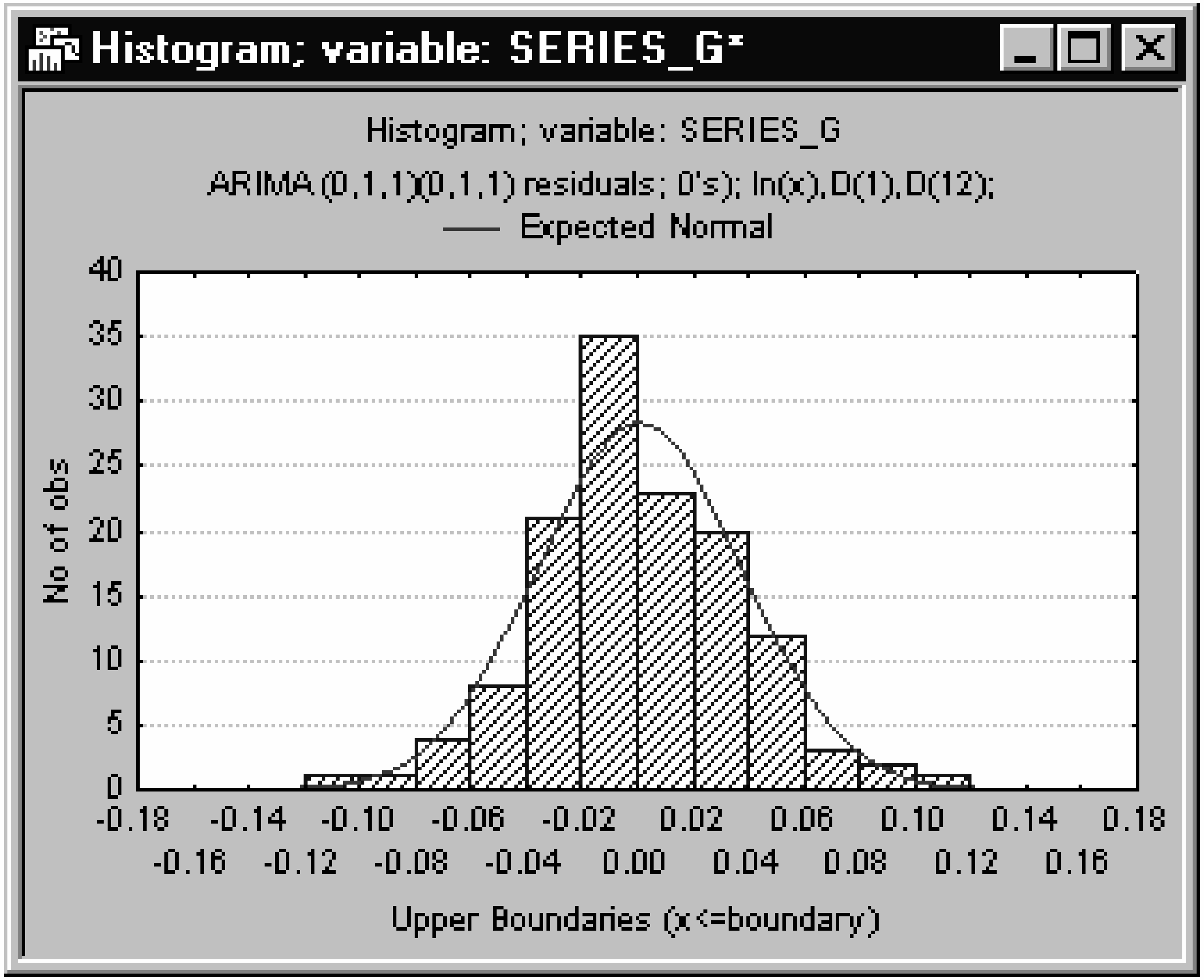
În vederea prognozei am aplicat modelul ARIMA (0, 1, 1) obținând secvența prognozată pentru perioada următoare, reprezentată în figură


Forecasts; Model:(0,1,1)(0,1,1) Seasonal l...
Forecasts; Model:(0,1,1)(0,1,1) Seasonal lag: 12
Input: SERIES_G: Monthly passenger totals (in 1000's)
Start of origin: 1
End of origin: 144
0
100
200
300
400
500
600
700
800
900
JAN 1950
JAN 1951
JAN 1952
JAN 1953
JAN 1954
JAN 1955
JAN 1956
JAN 1957
JAN 1958
JAN 1959
JAN 1960
Observed
Forecast
± 90.0000 %

Seria reziduurilor este normal repartizată :



Pentru o verificare „dublă", a fost desenată şi histograma reziduurilor.



Din cele două grafice de mai sus se observă că reziduurile sunt normal repartizate, deci prognoza este eficientă.

- A doua aplicație se referă la datele corespunzătoare apelurilor lunare către Cincinnati Directory Assistance, în perioada Ianuarie 1962- Decembrie 1976.

În *Martie* 1974, deci în a 147-a lună de la începerea înregistrării apelurilor, compania Cincinnati Bell a introdus o taxă suplimentară de 20 cenți pe fiecare apel către Cincinnati Directory Assistance.
Așa cum era de așteptat, această majorare a taxei pe fiecare apel a dus la scăderea numărului de apeluri.

Această acțiune exterioară asupra seriei temporale inițiale, considerată ca fiind în stare stabilă, a avut ca efect modificarea observațiilor ulterioare, ceea ce a determinat analiza seriei în ansamblul ei (adică atât până în Martie 1974, înainte de intrarea în aplicare a suprataxei, cât și după această dată).

Data: Director.st...

	Calls to directory assistance (in 100's)	
	1 MONTH	2 CALLS
1	22647	350
2	22678	339
3	22706	351
4	22737	364
5	22767	369
6	22798	331
7	22828	331
8	22859	340
9	22890	346
10	22920	341
11	22951	357
12	22981	398
13	23012	381
14	23043	367
15	23071	383
16	23102	375
17	23132	353

Data: Dire...

	Calls to directory a	
	1 MONTH	2 CALLS
139	26846	810
140	26877	762
141	26908	634
142	26938	626
143	26969	649
144	26999	697
145	27030	657
146	27061	549
147	27089	162
148	27120	177
149	27150	175
150	27181	162
151	27211	161
152	27242	165
153	27273	170
154	27303	172
155	27334	178
156	27364	186

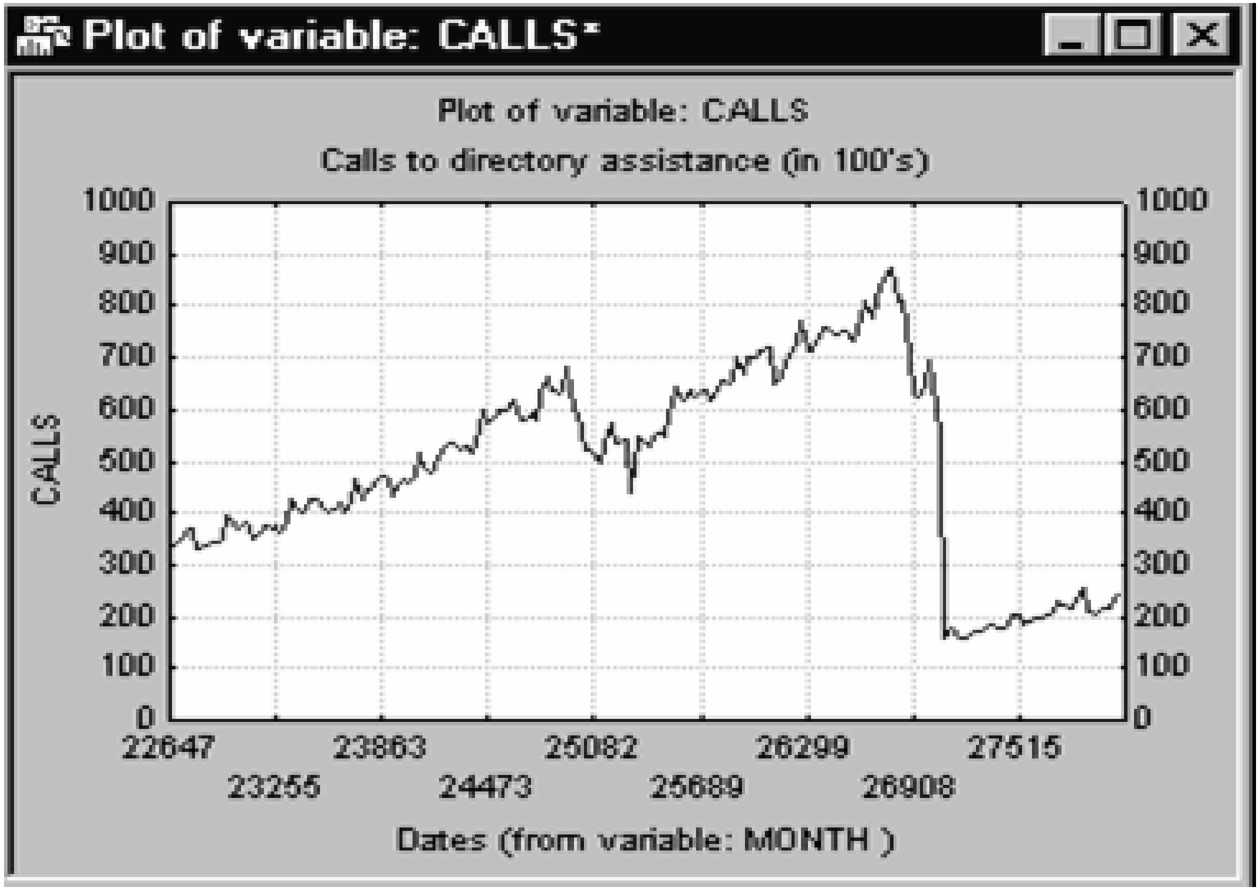
În cele două tabele sunt prezentate cele două eşantioane aparţinând celor două subserii (până la cazul 146 şi după acesta).

Este uşor de observat din figura de mai jos că după înregistrarea cu numărul 146 (tabelul din dreapta), asistăm la o scădere spectaculoasă a numărului apelurilor (de la 549 la numai 162).

Deoarece asistăm la intervenția externă a unui factor perturbator în luna a 146-a, va urma, logic, utilizarea unui model ARIMA pentru serii întrerupte.

seria temporala pana in luna 146

intreaga serie temporala

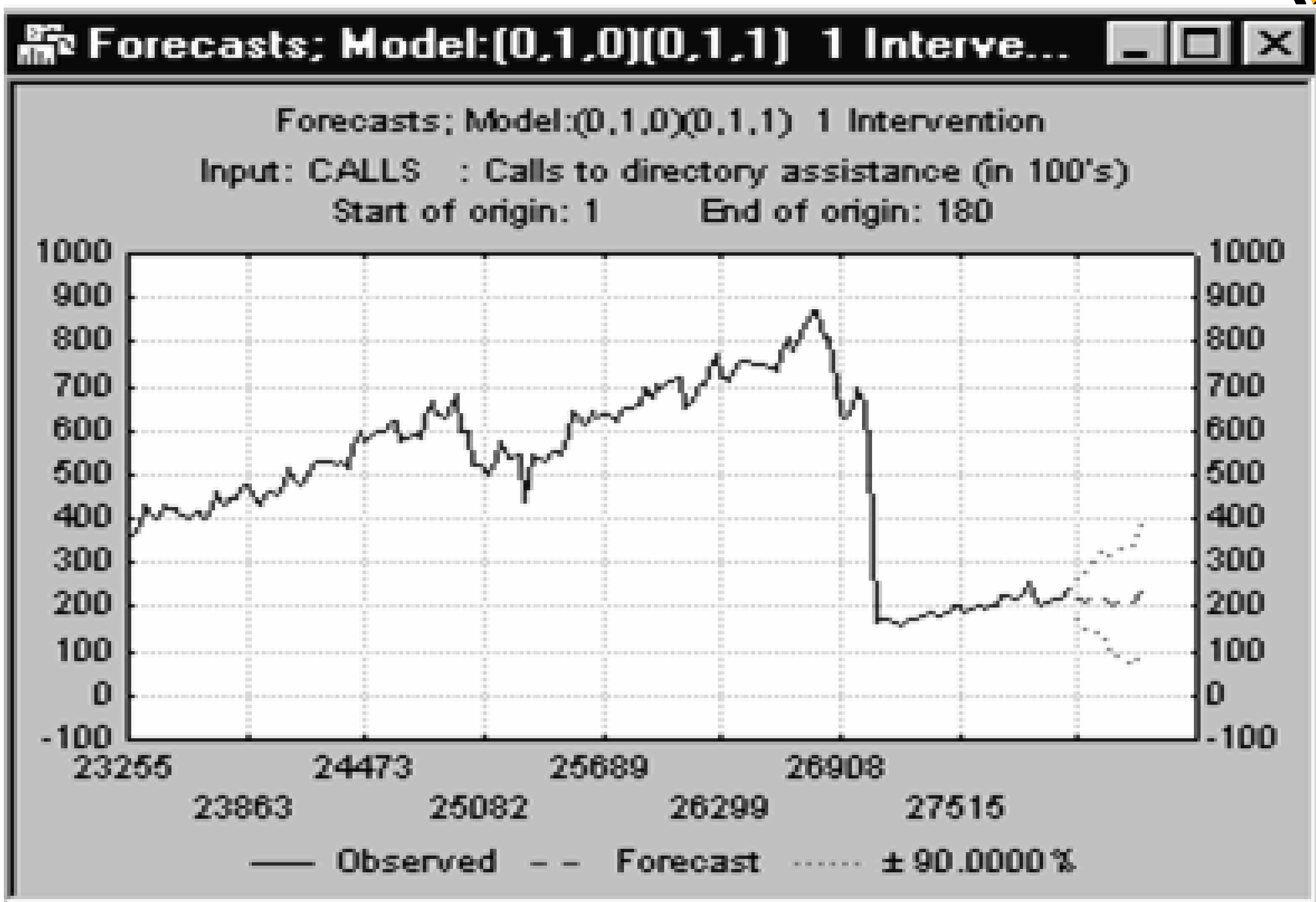


în cadrul modelului ARIMA pentru serii întrerupte există trei abordări:

- Intervenție abrupt-permanentă;
- Intervenție gradual-permanentă;
- Intervenție abrupt temporară.

În cazul de mai sus, este natural să alegem cazul intervenției abrupt permanente, deoarece scăderea numărului de apeluri a fost semnificativă (abruptă) şi supratAxarea care a produs-o a fost permanentă.

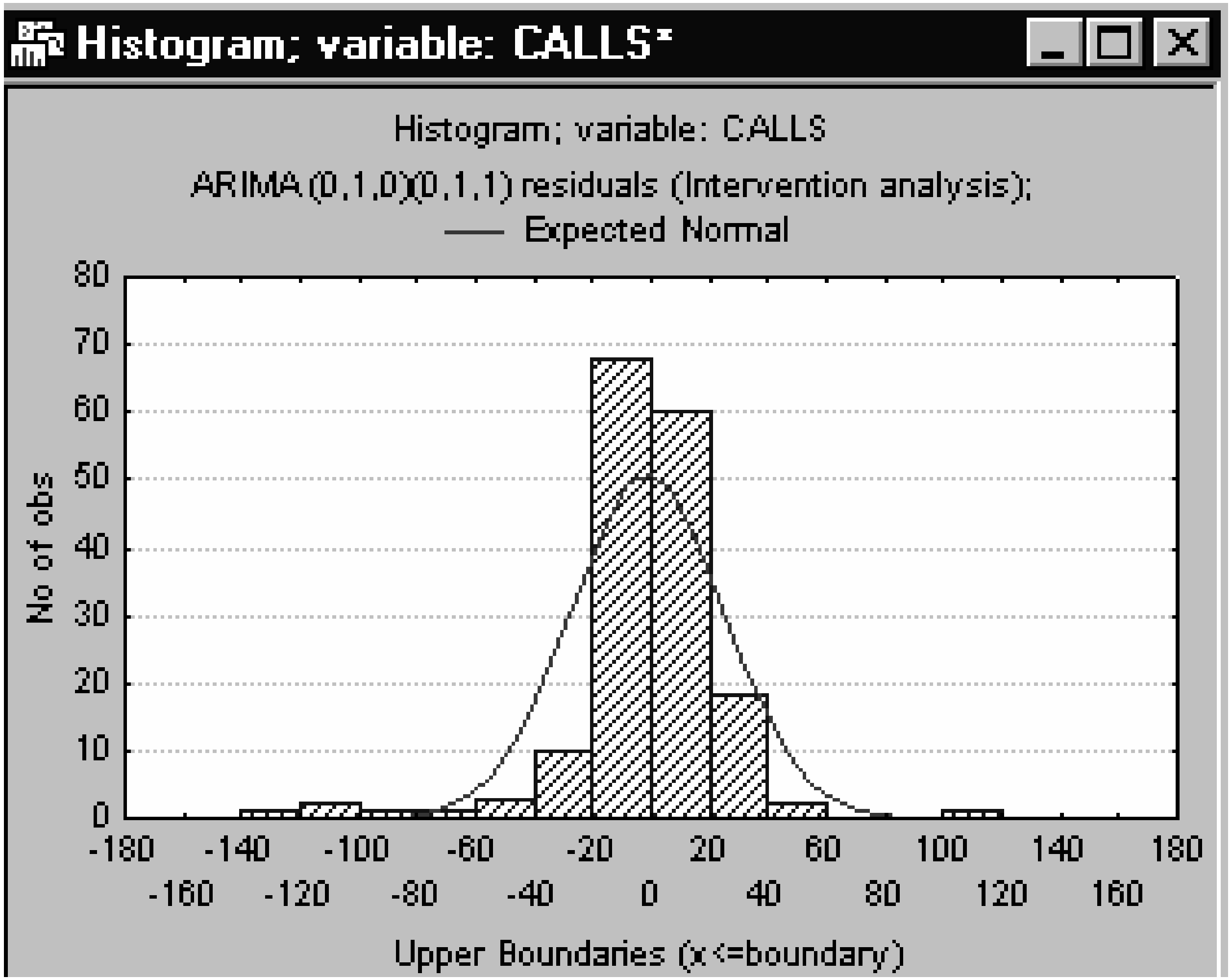
prognoza (un an) obtinuta aplicand modelul ARIMA



Pentru validarea modelului, utilizăm din

noua analiza reziduurilor.

Histograma corespunzătoare :



Histogram; variable: CALLS*
Histogram; variable: CALLS
ARIMA (0,1,0)(0,1,1) residuals (Intervention analysis);
Expected Normal
No of obs
80
70
60
50
40
30
20
10
0
-180
-140
-100
-60
-20
20
60
100
140
180
-160
-120
-80
-40
0
40
80
120
160
Upper Boundaries (x<=boundary)

Cu toate că din histograma nu rezultă o repartiție complet ,normală', totuşi putem considera modelul ales suficient de bun pentru prognoză.