Serii temporale

O mulţime de observaţii rezultate din măsurători efectuate în perioade succesive de timp, notată $\{ x_t , t \in T\}$, (unde $T \subseteq \mathbf{R}$, se referă la *timp*) se numeşte ***serie temporală*** sau *serie dinamică*.

Datele x_t se pot referi la observaţii în timp discret (zile, săptămâni, luni, trimestre, ani etc.) sau pot fi date obţinute în timp continuu.

Valorile succesive din baza de date reprezintă măsurători consecutive, prelevate la intervale de timp egale, măsurători ce ilustrează dinamica unui anumit proces în timp.

O caracterizare concisă a prognozei:

„Am văzut viitorul şi seamănă mult cu prezentul, numai că e mai lung.” (Kehlog Albran, *Profitul)*.

Suntem interesaţi de proprietăţile statistice ale seriilor temporale care vor descrie viitorul, în sensul descrierii comportamentului acestora în viitor, la fel de bine cum descriu şi prezentul.

Analiza seriilor temporale urmăreşte:

- înţelegerea evoluţiei seriei în trecut (identificarea naturii fenomenului reprezentat de secvenţa observaţiilor care sunt analizate);
- prognoza evoluţiei viitoare pe baza datelor cunoscute.

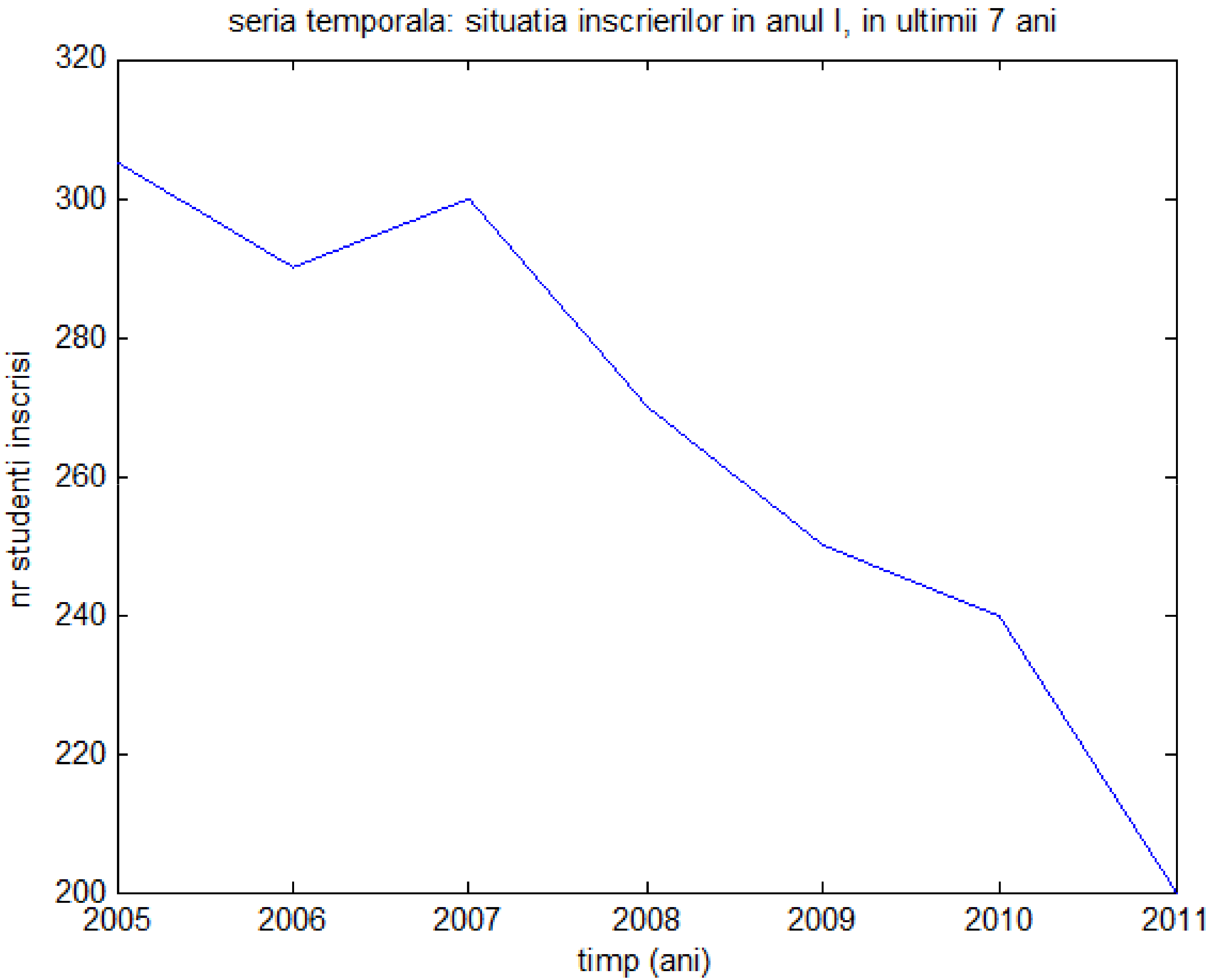
Seriile temporale se vizualizează prin tabele sau prin grafice.

1. Exemplu

Înscrierile în anul I al unei facultăţi, din ultimii 7 ani, sunt prezentate în tabelul de mai jos:

anul	2005	2006	2007	2008	2009	2010	2011
Studenţi înscrişi	305	290	300	270	250	240	200

```
>> t=2005:2011; x=[305 290 300 270 250 240 200];
>>plot(t,x,'k')
```


seria temporala: situatia inscrierilor in anul I, in ultimii 7 ani
320
300
280
260
240
220
200
nr studenti inscrisi
2005
2006
2007
2008
2009
2010
2011
timp (ani)

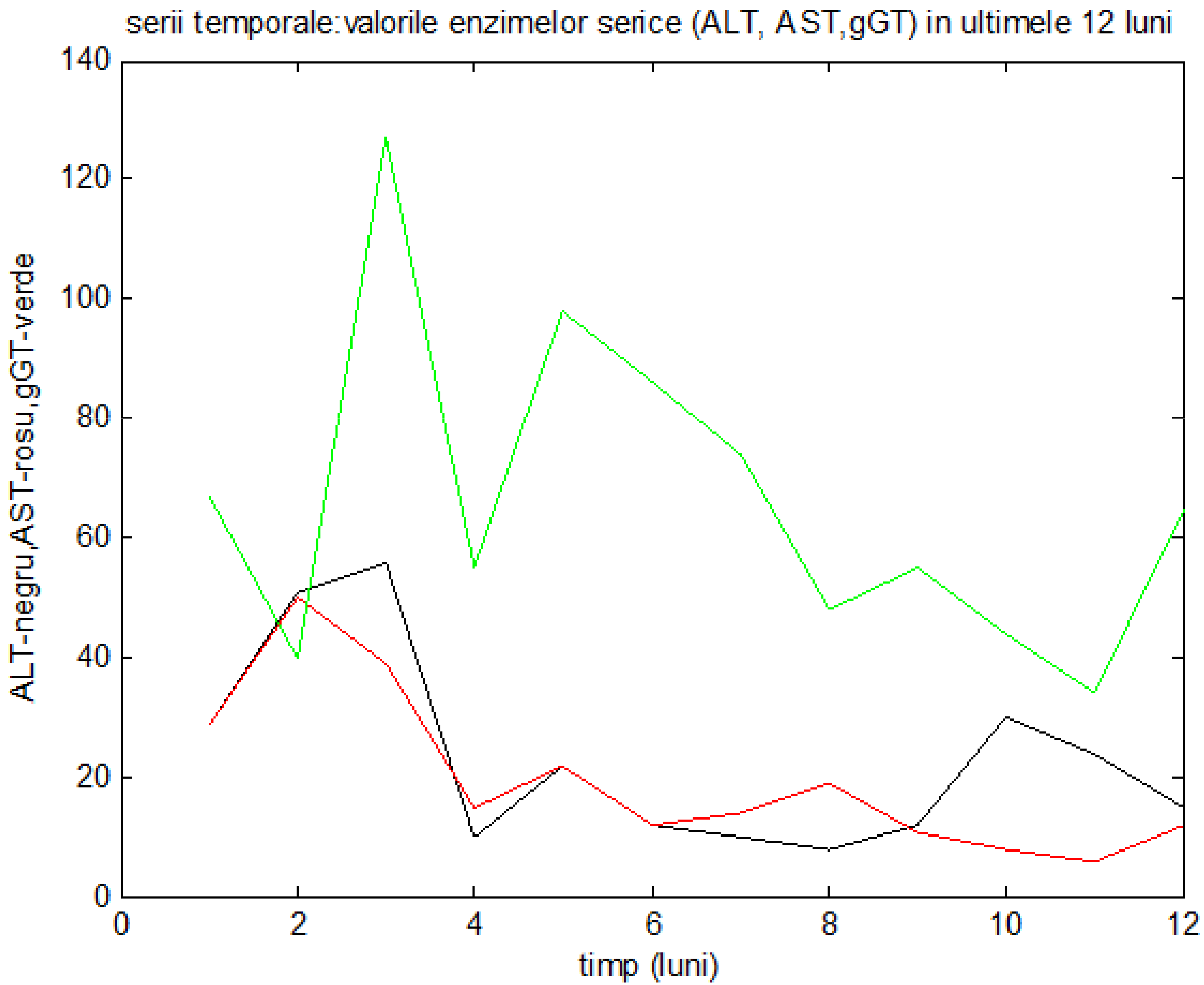
2. Exemplu

Tabelul următor prezintă valorile pentru principalele enzime serice prelevate timp de 12 luni consecutiv, de la un pacient diagnosticat cu hepatită cronică *C*: *alanina transaminasă* (*ALT*), *aspartate transaminase* (*AST*), *gamma glutamyl transpepdiase* (*gGT*):

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

Vom reprezenta în acelaşi sistem de axe cele trei serii temporale, vizualizare ce va permite medicilor să urmărească evoluţia simultană, în timpul unui an, a celor trei parametrii biochimici.

```
>> t=1:12; 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];
>>plot(t,ALT,'k',t,AST,'r',t,gGT,'g')
```



Comportarea (*pattern-ul*) datelor unei serii temporale este descrisă de către cele patru componente ale acesteia:

- trendul,
- componenta ciclică,
- componenta sezonieră,
- componenta incidentală.

Trendul

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

Deplasarea graduală a seriei temporale este percepută ca o direcție, **un trend**.

Trendul se datorează de exemplu schimbărilor în tehnologie, schimbărilor în preferințele consumatorilor, schimbărilor caracteristicilor demografice ale populației etc.

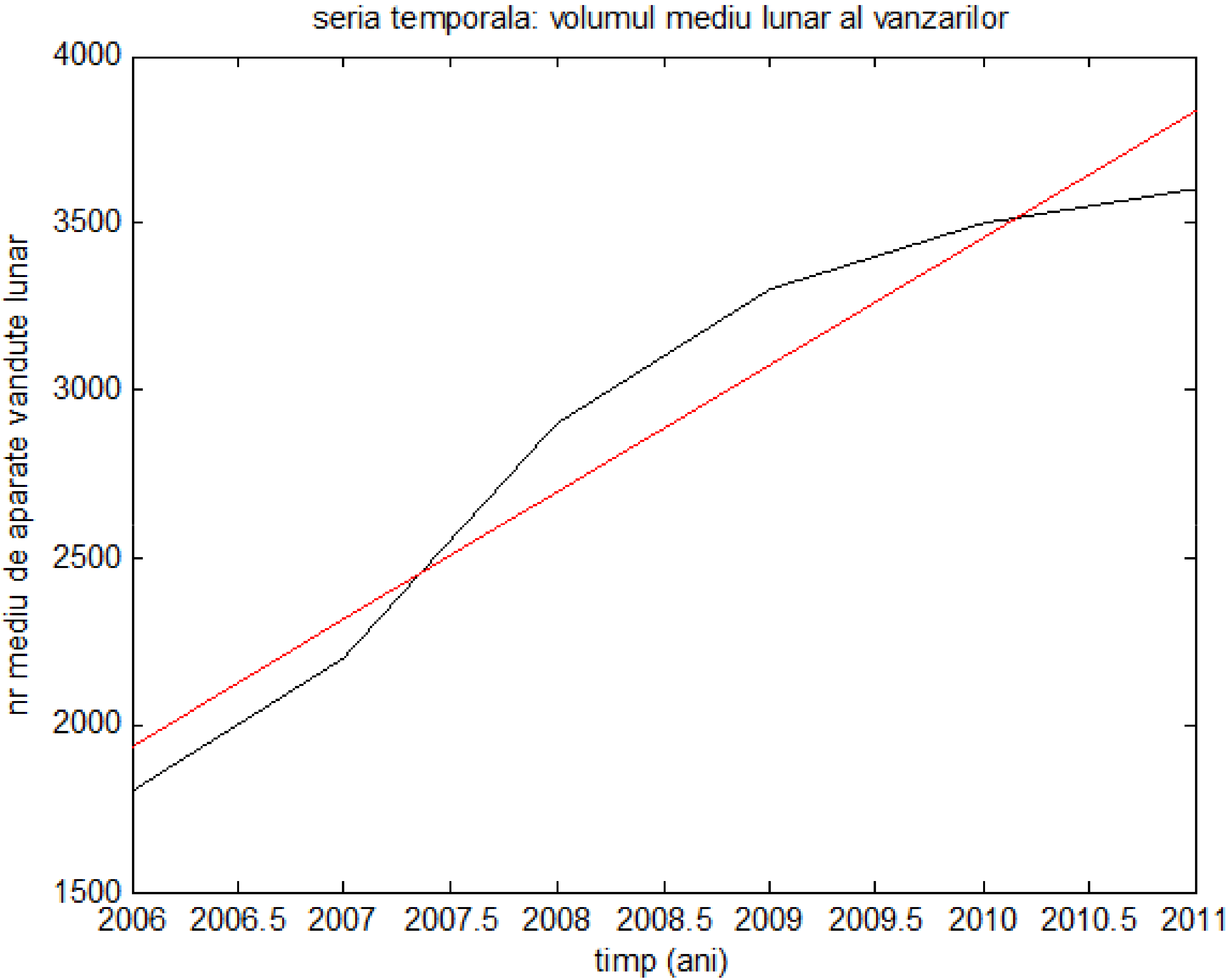
3. Exemplu

O firmă ce produce echipamente fotografice prezintă un **trend** ascendent al vânzărilor, după cum arată datele următoare care prezintă anul şi numărul mediu de aparate vândute lunar în anul respectiv:

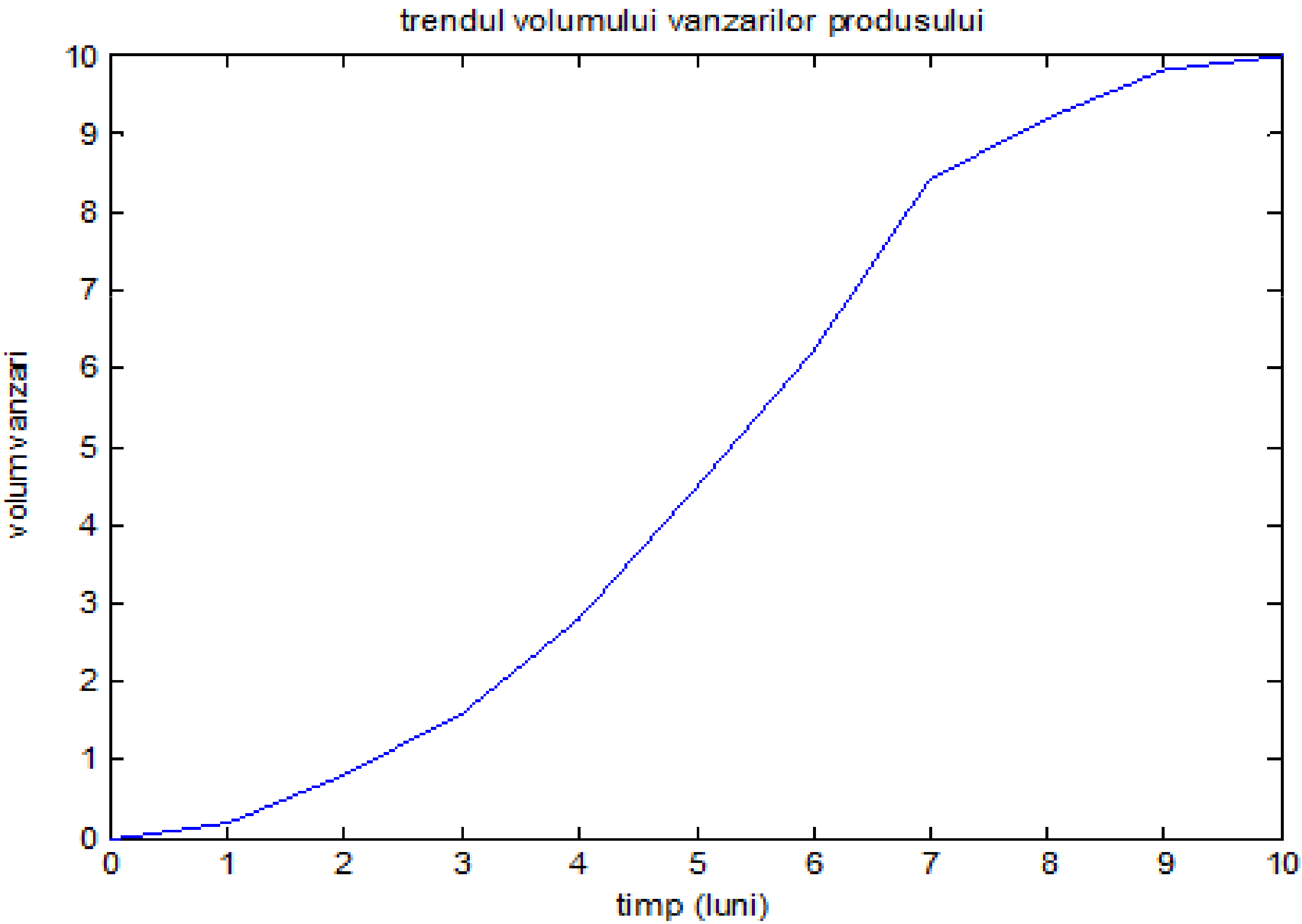
anul	2006	2007	2008	2009	2010	2011
Nr mediu	1800	2200	2900	3300	3500	3600

Pentru reprezentarea grafică am utilizat formula trendului liniar, ce va fi prezentată ulterior.

```
>>t=2006:2011;x=[1800 2200 2900 3300 3500 360];
>>b=(t*x'-6*mean(x)*mean(t))/(t*t'-6*(mean(t))^2);
>> a=mean(x)-b*mean(t);
>>plot(t,x,'k',t,a+b*t,'r')
```


seria temporala: volumul mediu lunar al vanzarilor
4000
3500
3000
2500
2000
1500
nr mediu de aparate vandute lunar
2006
2006.5
2007
2007.5
2008
2008.5
2009
2009.5
2010
2010.5
2011
timp (ani)

Prezentăm simularea vânzărilor unui produs, de la introducerea lui pe piaţă, urmată de perioada de creştere a vânzărilor şi în final de saturaţia pieţei.



Nu există tehnici de identificare automată a trendului în seriile temporale. Dacă tendinţa este monotonă (crescătoare sau descrescătoare), vizualizarea datelor este suficientă pentru identificarea acesteia.

Analiza trendului este necesară şi atunci când studiem mai multe serii temporale corespunzătoare aceleaşi perioade de timp.

Se analizează 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

Componenta ciclică: Valorile (datele) seriei temporale nu vor fi toate pe linia trendului. Orice secvenţă periodică de puncte, a cărei durată este cel puţin o unitate de măsură a timpului (de exemplu un an), se numeşte *componenta ciclică* a seriei temporale.

Asemenea componente pot apare de exemplu în economie: perioade de inflaţie mai modeste urmate de perioade de inflaţie rapidă.

Componenta sezonieră

Componenta sezonieră: O serie temporală poate prezenta o formă regulară într-o perioadă de timp mai mică de un an.

Componenta seriei temporale ce prezintă variaţii ale datelor influenţate de anotimpuri este *componenta sezonieră*.

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.

Astfel, volumul traficului prezintă variaţii „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 nopţii si moderate în rest.

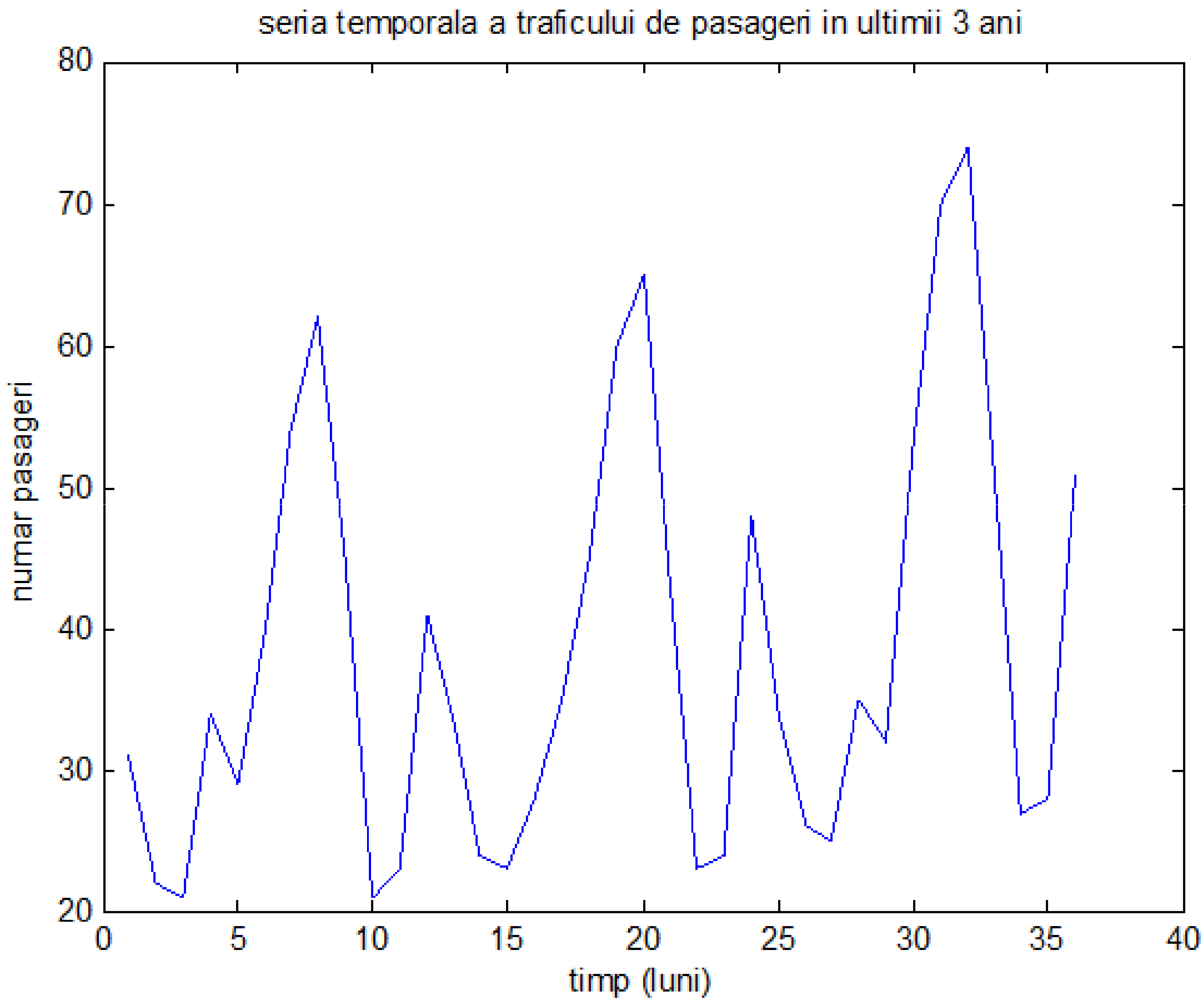
În domeniul vânzărilor se observă o creştere semnificativă în preajma Sărbătorilor de Crăciun şi de Paște. după Sărbătorile de Crăciun şi Paște creşte semnificativ numărul internărilor în spital.

4.Exemplu

Pe un aeroport, numărul de pasageri (în zeci de mii) pentru zborurile internaționale, înregistrat lunar în ultimii trei ani, este dat în următorul tabel:

luna	I	F	M	A	M	I	I	A	S	O	N	D
I	31	22	21	34	29	40	54	62	45	21	23	41
II	33	24	23	28	35	45	60	65	44	23	24	48
III	34	26	25	35	32	53	70	74	50	27	28	51

```
>> t=1:36;x=[31 22 21 34 29 40 54 62 45 21 23 41 33 24 23 28 35 45
60 65 44 23 24 48 34 26 25 35 32 53 70 74 50 27 28 51];
>> plot(t,x)
```


seria temporala a traficului de pasageri in ultimii 3 ani
numar pasageri
timp (luni)
80
70
60
50
40
30
20
0
5
10
15
20
25
30
35
40

Componenta incidentală

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

Metode de netezire (smoothing methods)

Într-o serie temporală întâlnim uneori variații întâmplătoare, caz în care apelăm la tehnici de reducere a efectelor acestor variații.

Tehnicile ce utilizează *metode de netezire* (smoothing methods) se aplică seriilor temporale *aproape stabile*, (serii ce nu prezintă modificări semnificative ale trend-ului, ale componentei ciclice sau ale celei sezoniere).

Tehnicile de netezire permit, pe de o parte, prognoza comportamentului viitor al seriei temporale şi, pe de altă parte, relevă mai clar trendul acesteia, componentele sezoniere şi ciclice. Vom prezenta două tehnici de netezire:

1. Mediile mobile
2. Metodele de netezire exponențială

Mediile mobile

Calculul mediilor mobile este cea mai simplă metodă de netezire a datelor, uşor de înţeles şi de folosit, care permite şi prognoza comportării ulterioare a seriei temporale.

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 pe cea mai veche dată din media mobilă, calculându-se astfel o nouă medie. Astfel media se va schimba o dată cu apariţia unei noi date, de unde şi termenul de medie mobilă.

Pentru a obţine o cât mai bună prognoză în cazul seriei temporale $\{x_t, t \in \{1, ..., p\}\}$, avem de găsit numărul optim de date ce vor fi introduse în media mobilă.

O abordare posibilă constă în a încerca diferite variante şi de a o alege pe aceea care minimizează media pătratelor erorii (*mean squared error_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.

Etapele rezolvării problemei sunt:

- calcularea mediilor mobile pentru $n \in \{3, 4, 5, ...\}$, $M_{n+i}, 1 \le i \le p-n$
- presupunând că M_{n+i} este prognoza pentru săptămâna numerotată $n+i$,

calculăm pătratul erorii comise cu această prognoză:

$$er_i = (M_{n+i} - x_{n+i})^2$$

- pentru a aprecia acurateţea prognozei vom lua în considerare media pătratelor erorii – *MSE*

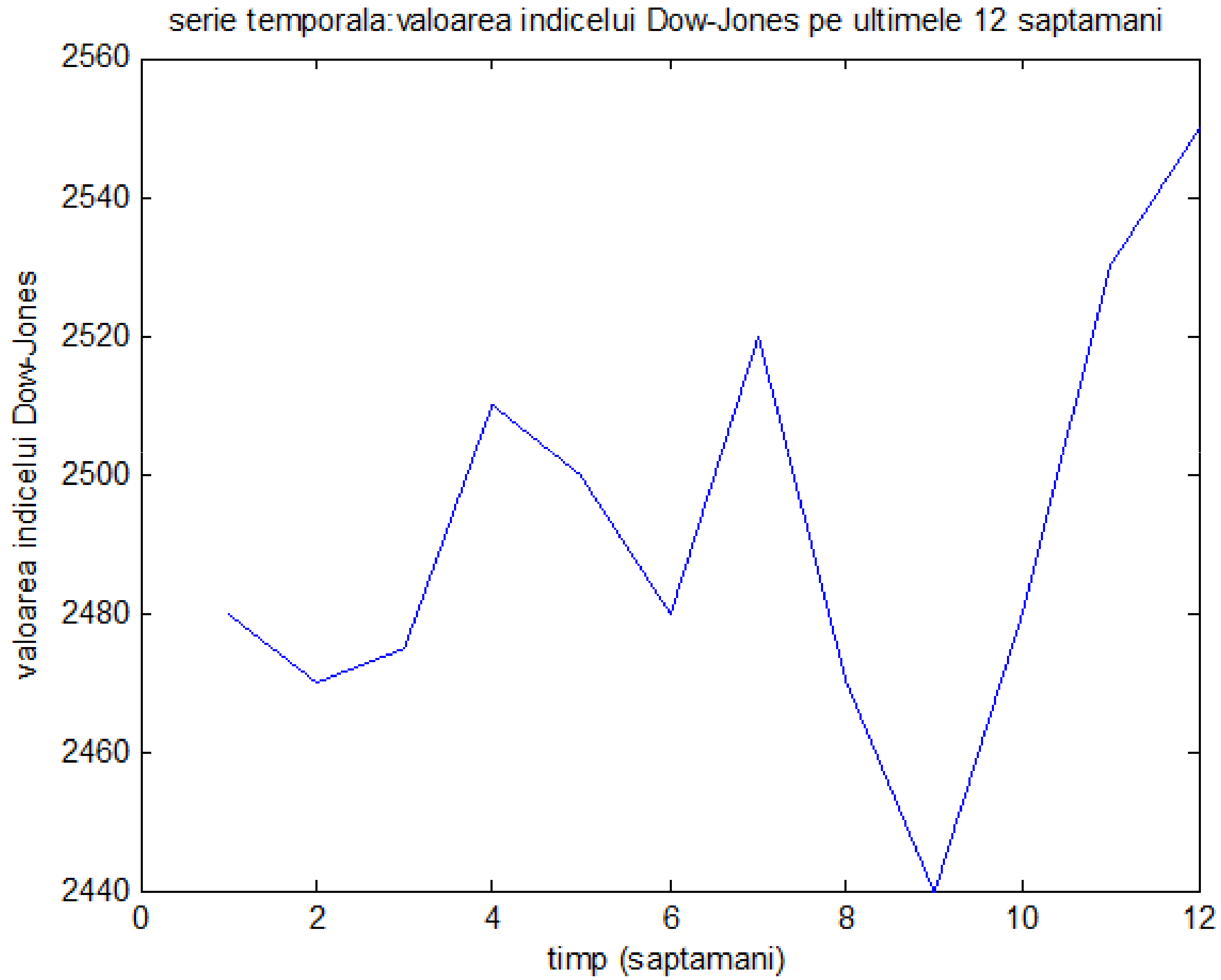
Pentru prognoză, dacă seria temporală nu are trend, este folosită media mobilă.

5.Exemplu

Indicele Dow Jones (Dow Jones Industrial Average sau Dow 30) este cel mai vechi indice al burselor din New York şi cel mai vechi indice bursier din lume (1884). Indicele urmăreşte evoluţia pe piaţa bursieră a 30 de firme americane importante, care s-au schimbat de-a lungul timpului (numai General Electric a rămas din 1884 până azi). Dow Jones permite studiul evoluţiei economiei americane în ultimii 120 ani. Prezentăm indicele Dow Jones la închidere, pe 12 săptămâni

săptămâna	Dow Jones		săptămâna	Dow Jones
1	2480		7	2520
2	2470		8	2470
3	2475		9	2440
4	2510		10	2480
5	2500		11	2530
6	2480		12	2550

```
>> t=1:12;
>>x=[2480 2470 2475 2510 2500 2480 2520 2470 2440 2480 2530 2550];
>> plot(t,x)
```



Vom calcula media mobilă pentru 3, 4 şi respectiv 5 săptămâni şi media pătratelor erorii (*MSE*) în cele trei cazuri:

```
>> x=[2480 2470 2475 2510 2500 2480 2520 2470 2440 2480 2530 2550];
>> for t=4:12; m3(t)=(x(t-1)+x(t-2)+x(t-3))/3;end
>> for t=4:12; e3(t)=(x(t)-m3(t))^2;end
>> for t=5:12; m4(t)=(x(t-1)+x(t-2)+x(t-3)+x(t-4))/4;end
>> for t=5:12; e4(t)=(x(t)-m4(t))^2;end
>> for t=6:12; m5(t)=(x(t-1)+x(t-2)+x(t-3)+x(t-4)+x(t-5))/5;end
>> for t=6:12; e5(t)=(x(t)-m5(t))^2;end
```

```
>> [x' m3' e3' m4' e4' m5' e5']
ans =
  1.0e+003 *
    2.4800         0         0         0         0         0         0
    2.4700         0         0         0         0         0         0
    2.4750         0         0         0         0         0         0
    2.5100    2.4750    1.2250         0         0         0         0
    2.5000    2.4850    0.2250    2.4838    0.2641         0         0
    2.4800    2.4950    0.2250    2.4888    0.0766    2.4870    0.0490
    2.5200    2.4967    0.5444    2.4913    0.8266    2.4870    1.0890
    2.4700    2.5000    0.9000    2.5025    1.0562    2.4970    0.7290
    2.4400    2.4900    2.5000    2.4925    2.7563    2.4960    3.1360
    2.4800    2.4767    0.0111    2.4775    0.0063    2.4820    0.0040
    2.5300    2.4633    4.4444    2.4775    2.7563    2.4780    2.7040
    2.5500    2.4833    4.4444    2.4800    4.9000    2.4880    3.8440
```

Din tabelul generat de MATLAB, se observă că e3, e4, e5 care sunt vectori cu 9, respectiv 8, respectiv 7 componente, sunt considerați a fi vectori cu 12 componente, scufundarea în spațiul 12-dimensional fiind realizată prin atribuirea valorii zero componentelor lipsă. Dacă vom calcula mediile pătratelor erorii aplicând direct instrucțiunea `mean`, obținem rezultate viciate, deoarece se face împărțirea la 12 componente și nu la 9, respectiv 8, respectiv 7.

Pentru a nu renunţa la utilizarea instrucţiunii **mean**, vom ţine seama că:

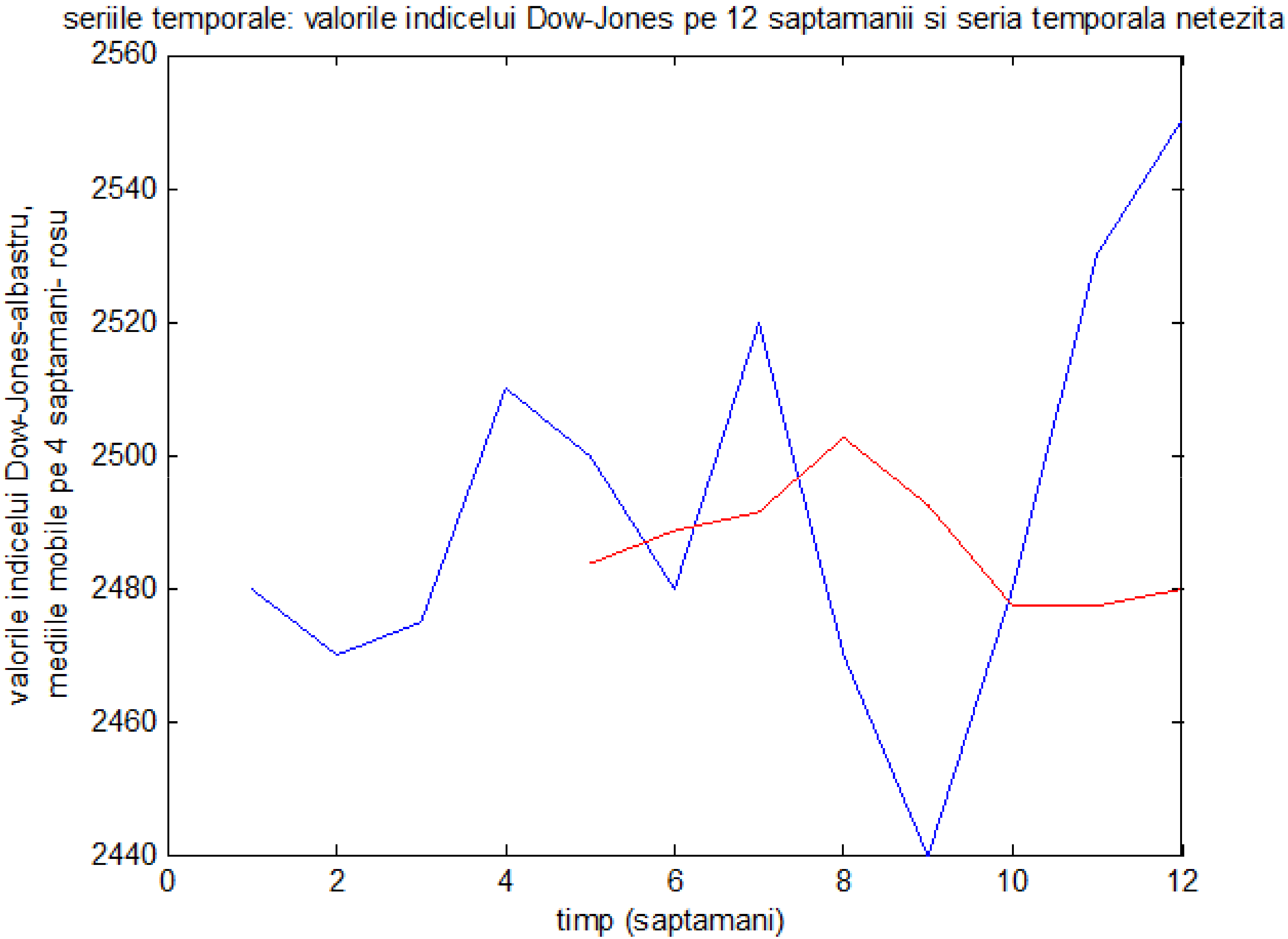
MSEk=12/(12-k)*mean(ek), $k \in \{3,4,5\}$

Astfel:

```
>>MSE3=12/9*mean(e3);
>> MSE4=12/8*mean(e4);
>> MSE5=12/7*mean(e5);
>> [MSE3 MSE4 MSE5]
ans =
  1.0e+003 *
    1.6133    1.5803    1.6507
```

Remarcăm că media mobilă pentru $n = 4$ conduce la cea mai mică valoare a mediei pătratelor erorii şi astfel aceasta va fi folosită în prognoză.

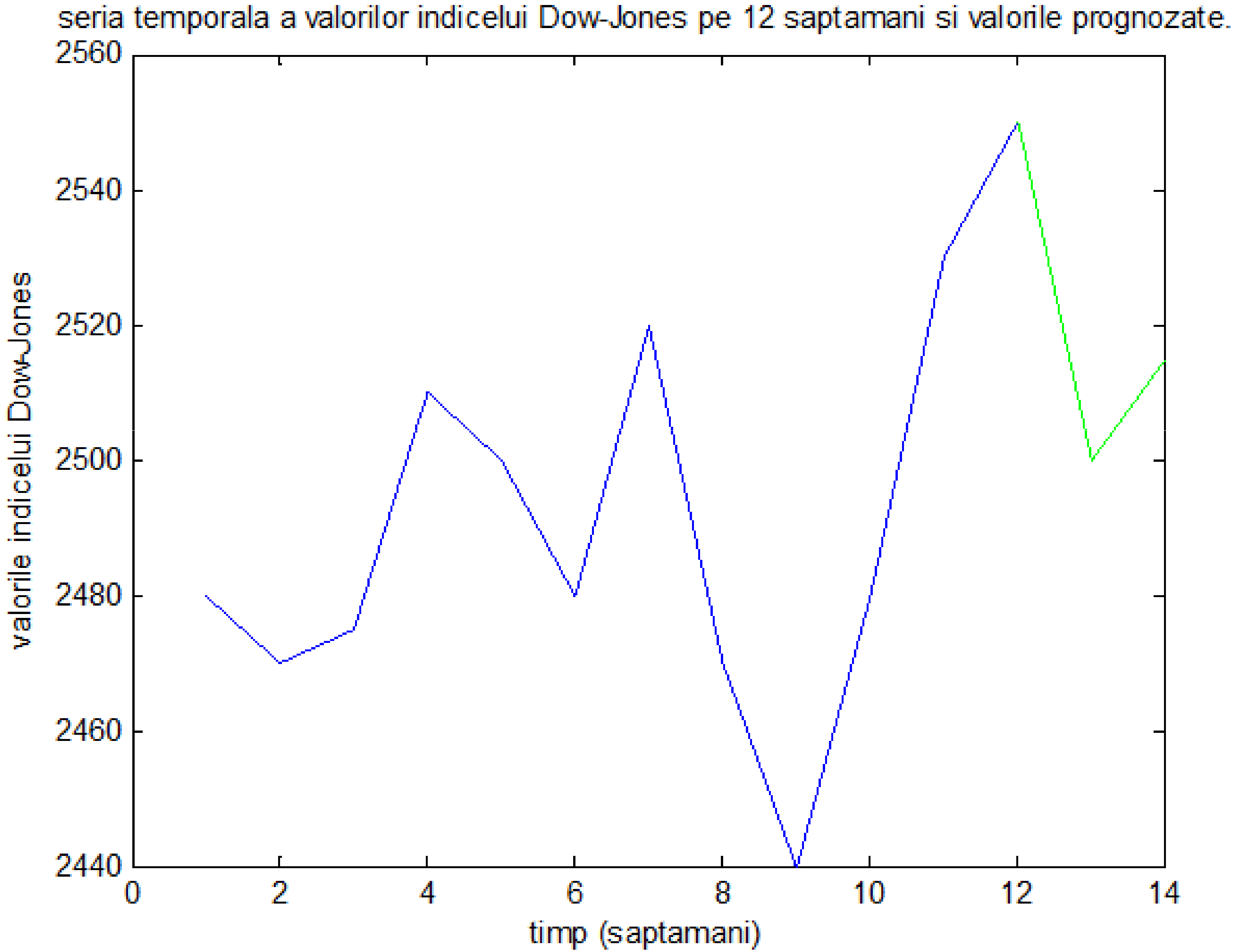
```
>>t1=5:12; plot(t,x,'b',t1,m4(t1),'r')
```



Astfel prognoza indicelui Dow-Jones pentru săptămâna a 13-a, respectiv a 14-a va fi:

```
>> p13=(x(9)+x(10)+x(11)+x(12))/4;
>> p14=(x(10)+x(11)+x(12)+p13)/4;
>> [p13 p14]
ans =
       2500       2515
```

```
>> t1=12:14;x1=[ x(12) p13 p14];
>> plot(t,x,'b', t1,x1,'g')
```



Medie mobilă ponderată

O formă mai sofisticată de medie mobilă, care ţine cont de importanţa 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.

Problemă: care este numărul n (numărul de date introduse în media mobilă) care permite cea mai bună prognoză?

Etapele rezolvării problemei sunt aceleaşi ca în cazul mediei mobile.

Reluăm datele din exemplul nr 5, referitoare la valorile indicelui Dow-Jones pe ultimele 12 săptămâni. Vom calcula mediile mobile ponderate pentru $n \in \{3, 4, 5\}$, pătratele erorilor comise şi cea mai mică valoare a mediei pătratelor erorii ne va da numărul n optim.

```
>> x=[2480 2470 2475 2510 2500 2480 2520 2470 2440 2480 2530 2550];
>> for t=4:12; mp3(t)=(3*x(t-1)+2*x(t-2)+x(t-3))/6;end
>> for t=4:12; ep3(t)=(x(t)-mp3(t))^2;end
>> for t=5:12; mp4(t)=(4*x(t-1)+3*x(t-2)+2*x(t-3)+x(t-4))/10;end
>> for t=5:12; ep4(t)=(x(t)-mp4(t))^2;end
>> for t=6:12; mp5(t)=(5*x(t-1)+4*x(t-2)+3*x(t-3)+2*x(t-4)+x(t-5))/15;end
>> for t=6:12; ep5(t)=(x(t)-mp5(t))^2;end
```

```
>> [x' mp3' ep3' mp4' ep4' mp5' ep5']
ans =
  1.0e+003 *
    2.4800         0         0         0         0         0         0
    2.4700         0         0         0         0         0         0
    2.4750         0         0         0         0         0         0
    2.5100    2.4742    1.2840         0         0         0         0
    2.5000    2.4917    0.0694    2.4885    0.1323         0         0
    2.4800    2.4992    0.3674    2.4950    0.2250    2.4923    0.1521
    2.5200    2.4917    0.8028    2.4915    0.8123    2.4900    0.9000
    2.4700    2.5033    1.1111    2.5030    1.0890    2.5010    0.9610
    2.4400    2.4883    2.3361    2.4900    2.5000    2.4920    2.7040
    2.4800    2.4633    0.2778    2.4690    0.1210    2.4733    0.0444
    2.5300    2.4650    4.2250    2.4700    3.6000    2.4727    3.2871
    2.5500    2.4983    2.6694    2.4910    3.4810    2.4900    3.6000
```

Folosim aceeaşi tehnică pentru calculul mediei pătratelor erorilor:

```
>> MSEp3=12/9*mean(ep3);
>> MSEp4=12/8*mean(ep4);
>> MSEp5=12/7*mean(ep5);
>> [ MSEp3 MSEp4 MSEp5]
ans =
  1.0e+003 *
    1.4603    1.4951    1.6641
```

Valoarea minimă a mediei pătratelor erorii este în cazul $n = 3$ şi astfel vom folosi media ponderată pe trei săptămâni pentru a prognoza valoarea indicelui Dow Jones pe următoarele săptămâni.

```
>> p13=(x(10)+2*x(11)+3*x(12))/6; p14=(x(11)+2*x(12)+3*p13)/6
>> [p13 p14]
ans =
  1.0e+003 *
      2.5317  2.5375
```

Observăm că în acest exemplu este preferabilă utilizarea mediei mobile ponderate şi nu a mediei mobile, (MSEp3 ia cea mai mică valoare).

Netezirea exponenţială

O metodă simplă, deseori utilizată, care permite prognoza pentru perioadă de timp $t+1$ în funcţie de valoarea seriei şi valoarea prognozei pentru perioadă de timp t, este *netezirea exponenţială*.

Prezentăm modelul matematic:

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

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

Alegerea valorii atribuite lui P_1 este importantă.

Metode des întâlnite de iniţializare:

- consideram $P_1 = x_1$.
- considerăm P_1 ca fiind media aritmetică a primelor patru sau cinci observaţii.

Se poate demonstra (prin inducţie) că:

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

Această metodă de netezire se numeşte *exponenţială*, deoarece ponderile sunt proporţionale cu termenii unei progresii geometrice

$\{1, 1-\alpha, \ (1-\alpha)^2, ...\}$, $\alpha \in [0,1]$.

Reamintim că o progresie geometrică este versiunea discretă a unei funcţii exponenţiale.

În alegerea constantei de netezire $\alpha \in [0,1]$,vom ţine seama că valoarea prognozată P_{t+1} depinde de prognoza P_t şi de eroarea prognozei anterioare $x_t - P_t$:

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

În general, alegem valoarea lui α care minimizează media pătratelor erorii (*mean squared error_MSE*). 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 pătratelor erorii corespunzătoare *_MSE;*
- alegem acel α pentru care obţinem cel mai mic *MSE.*

6. Exemplu

Reluăm datele privitoare la indicele Dow-Jones din exemplul nr 5 şi folosim metoda incrementală:

```
>> x=[2480 2470 2475 2510 2500 2480 2520 2470 2440 2480 2530 2550];
>> for j=1:9 p(1,j)=x(1);end
>> for j=1:9
for t=2:12 p(t,j)=0.1*j*x(t-1)+0.1*(10-j)*p(t-1,j);end;
end
>> p
p =
  1.0e+003 *
    2.4800   2.4800   2.4800   2.4800   2.4800   2.4800   2.4800   2.4800   2.4800
    2.4800   2.4800   2.4800   2.4800   2.4800   2.4800   2.4800   2.4800   2.4800
    2.4790   2.4780   2.4770   2.4760   2.4750   2.4740   2.4730   2.4720   2.4710
    2.4786   2.4774   2.4764   2.4756   2.4750   2.4746   2.4744   2.4744   2.4746
    2.4817   2.4839   2.4865   2.4894   2.4925   2.4958   2.4993   2.5029   2.5065
    2.4836   2.4871   2.4905   2.4936   2.4962   2.4983   2.4998   2.5006   2.5006
    2.4832   2.4857   2.4874   2.4882   2.4881   2.4873   2.4859   2.4841   2.4821
    2.4869   2.4926   2.4972   2.5009   2.5041   2.5069   2.5098   2.5128   2.5162
    2.4852   2.4881   2.4890   2.4885   2.4870   2.4848   2.4819   2.4786   2.4746
    2.4807   2.4784   2.4743   2.4691   2.4635   2.4579   2.4526   2.4477   2.4435
    2.4806   2.4788   2.4760   2.4735   2.4718   2.4712   2.4718   2.4735   2.4763
```

```
>>for j=1:9 for t=1:12 er(t,j)=(p(t,j)-x(t))^2;end;end
>>MSE=mean(er)
MSE =
  1.0e+003 *
    0.9769    0.9854    1.0014    1.0106    1.0092    0.9990    0.9842    0.9688
0.9556
```

Aşadar pentru prognoza valorilor indicelui Dow-Jones pentru săptămâna a 13-a vom considera $\alpha = 0.9$ şi astfel:

```
>>p13=0.9*x(12)+0.1*p(12,9)
p13 =
  2.5475e+003
```

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

Diferența între *MSE* și *MAD* constă în faptul că *MSE* este influențată de erorile mari de prognoză, fiind media pătratelor erorilor.

Specialiștii au păreri contradictorii asupra alegerii uneia sau alteia dintre cele două metode. Preferăm utilizarea *MSE*.

În cazul unei serii temporale care are trend nu poate fi folosită netezirea exponențială prezentată.

Situația este remediată prin introducerea unei a doua ecuații cu a doua constantă γ, aleasă coroborat cu prima constantă de netezire α.

Matematic, următoarele două ecuații descriu *netezirea dublu exponențială*:

$$P_t = \alpha \cdot x_t + (1-\alpha)(P_{t-1} + b_{t-1}),\ \alpha \in [0,1],$$
$$b_t = \gamma \cdot (P_t - P_{t-1}) + (1-\gamma) \cdot b_{t-1},\ \gamma \in [0,1].$$

Prima ecuație exprimă prognoza P_t în funcție de valoarea x_t a seriei temporale, de valoarea anterioară P_{t-1} și de valoarea anterioară a trendului b_{t-1}.

A doua ecuație actualizează trendul, care este exprimat ca diferența dintre ultimele două valori prognozate P_{t-1}, P_t.

Există mai multe metode pentru alegerea valorilor iniţiale.

Dacă în general alegem $P_1 = x_1$, pentru b_1 avem mai multe variante:

$$b_1 = x_2 - x_1;$$

$$b_1 = \frac{(x_2 - x_1) + (x_3 - x_2) + (x_4 - x_3)}{3};$$

$$b_1 = \frac{x_n - x_1}{n - 1}.$$

Problema ce urmează a fi rezolvată este alegerea acelor parametrii α şi γ care corespund valorii minime a mediei pătratelor erorii (*MSE*).

Propunem utilizarea metodei incrementale.

Pentru a prognoza valoarea următoare din seria temporală se utilizează formula:

$$F_{t+1} = P_t + b_t \,,$$

în timp ce pentru prognoza următoarelor m perioade avem:

$$F_{t+m} = P_t + m \cdot b_t$$

Exemplu

Reluăm exemplul nr 1, referitor la înscrierile în anul I al unei facultăţi, din ultimii 7 ani şi aplicăm tehnica de netezire dublu exponenţială, folosind metoda incrementală:

- calculăm probabilităţile pentru $\alpha \in \{0.1 * j, j = 1:9\}$ şi $\gamma \in \{0.1 * k, k = 1:9\}$;
- evaluăm media pătratelor erorilor în cele 81*7 cazuri;
- alegem acele valori ale lui α şi γ care au determinat cea mai

mică valoare a MSE

```
>> x=[305 290 300 270 250 240 200];
>> for j=1:9
for k=1:9 P(1,j,k)=x(1);end;
end
for j=1:9
for k=1:9 b(1,j,k)=x(2)-x(1);end;
end
>> for j=1:9
for k=1:9 b(1,j,k)=x(2)-x(1);end;
end
>> for t=2:7
for j=1:9
for k=1:9
P(t,j,k)=0.1*j*x(t)+0.1*(10-j)*(P(t-1,j,k)+b(t-1,j,k));
b(t,j,k)=0.1*k*(P(t,j,k)-P(t-1,j,k))+0.1*(10-k)*b(t-1,j,k);
end
end
end
>> for t=1:7
for j=1:9
for k=1:9
er(t,j,k)=(x(t)-P(t,j,k))^2;
end
end
end
er
```

Vor fi afişate 9 matrice cu 7 linii şi 9 coloane, notate **er**(:,:,j), care reprezintă eroarea pentru $\gamma = 0.1 * j$. Vom calcula media erorilor pe cei 7 ani, pentru fiecare parametru în parte şi apoi vom stabili care este cea mai mică medie:

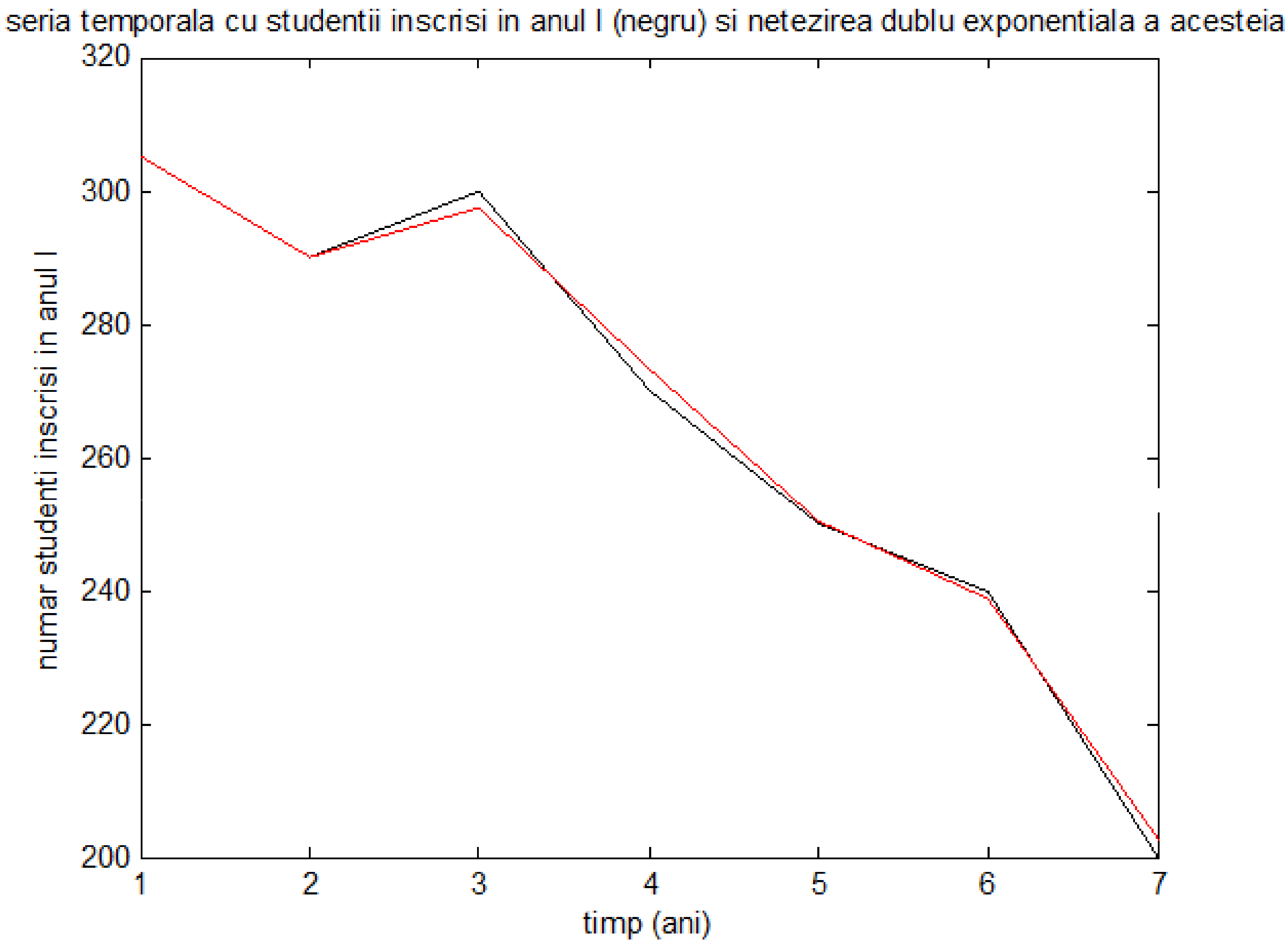
```
>> for j=1:9
for k=1:9
mean(er(:,j,k));
end
end
>> mean(er)
```

Obţinem 9 vectori de 9 componente, notaţi **ans**(:,:,j), corespunzători lui $\gamma = 0.1 * j$

```
>> min(min(mean(er)))

ans =

   2.2131
```

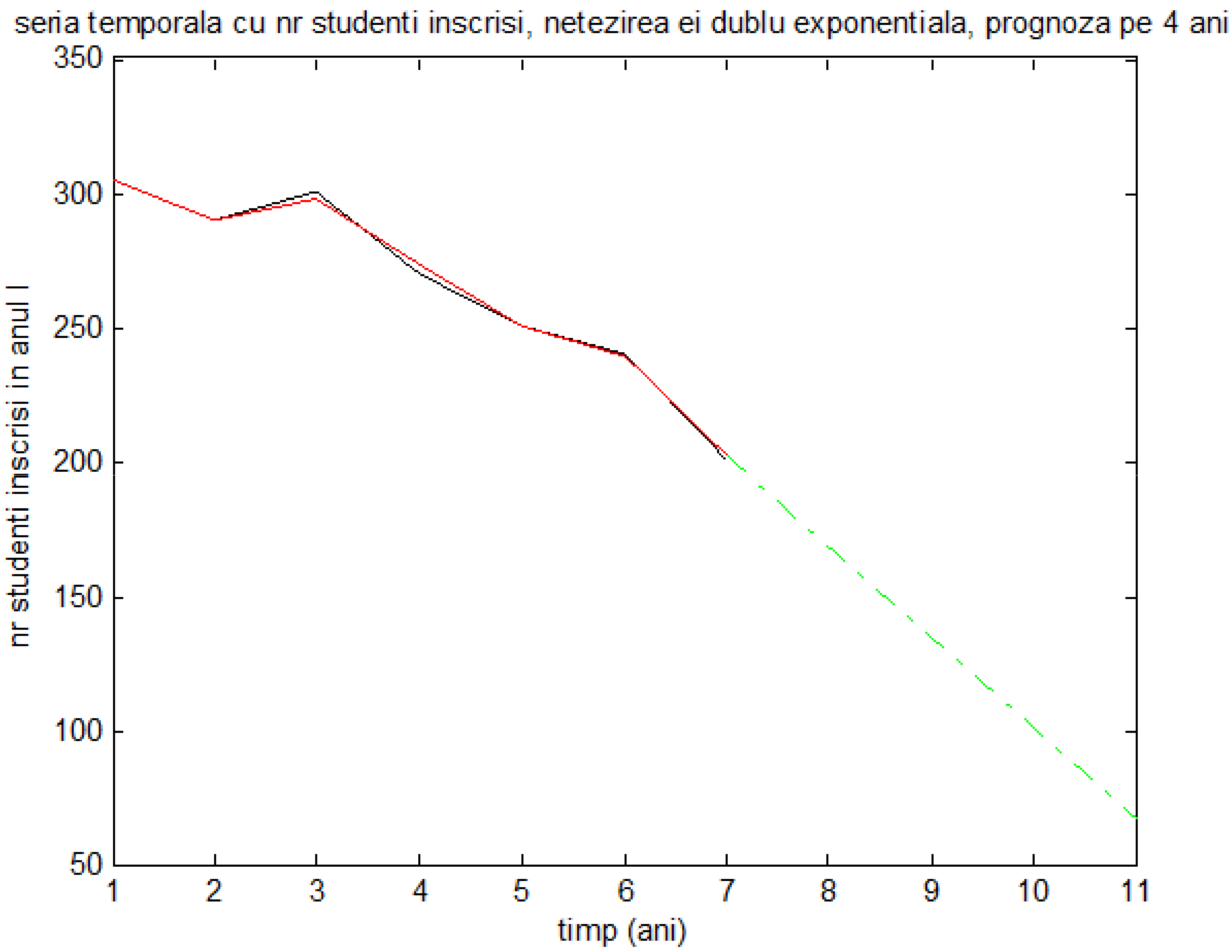
Valoarea minimă a mediei pătratelor erorilor este obţinută pentru $\alpha = 0.9$ şi $\gamma = 0.1$ şi astfel am determinat valorile parametrilor ce vor fi utilizaţi în netezirea dublu exponentială a seriei temporale.

```
>> for t=2:7
P(t)=0.9*x(t)+0.1*(P(t-1)+b(t-1));
b(t)=0.9*(P(t)-P(t-1))+0.1*b(t-1);
end
>> t=1:7;
>> plot(t,x,'k', t,P,'-.k')
```


seria temporala cu studentii inscrisi in anul I (negru) si netezirea dublu exponentiala a acesteia
320
300
280
260
240
220
200
numar studenti inscrisi in anul I
1
2
3
4
5
6
7
timp (ani)

Facem prognoza pe următorii patru ani:

```
>> F(8)=P(7)+b(7); for t=2:4 F(7+t)=P(7)+t*b(7);end % F provine de la
forecasting
>> F'
ans =
  168.8439
  135.0760
  101.3081
   67.5402

>> t=1:7;t1=7:11; F=[ P(7) F(8) F(9) F(10) F(11)];
>> plot (t,x,'k',t,P,'r', t1,F,'g-.')
```


seria temporala cu nr studenti inscrisi, netezirea ei dublu exponentiala, prognoza pe 4 ani
nr studenti inscrisi in anul I
timp (ani)
350
300
250
200
150
100
50
1
2
3
4
5
6
7
8
9
10
11

Prognoza utilizând trendul

În cazul în care valorile unei serii temporale prezintă timp îndelungat un trend liniar, putem prognoza valorile viitoare folosind regresia liniară pentru identificarea trendului.

Componenta trendului nu va reflecta orice variaţie a seriei temporale ca funcţie de timp, ci, direcţia în care variază seria temporală, în cazul nostru creşterea valorilor în funcţie de timp.

Matematic, avem:

$$T_t = a + bt ,$$

unde:

- T_t este valoarea trend-ului seriei temporale în perioada de timp *t;*
- b_0 este interceptorul liniei trendului;
- b_1 este panta liniei trend-ului.

Conform formulelor pentru calculul dreptei de regresie, avem:

$$b = \frac{\sum_{t \in T} t \cdot x_t - \frac{1}{n} \sum_{t \in T} t \cdot \sum_{t \in T} x_t}{\sum_{t \in T} t^2 - \frac{1}{n} \left(\sum_{t \in T} t \right)^2}; \quad a = \bar{x} - b \cdot \bar{t}$$

unde:

- x_t este valoarea actuală a seriei temporale in perioada de timp t;
- n este numărul perioadelor de timp, $n = card(T)$;
- $\bar{x}$ este media valorilor seriei temporale, $\bar{x} = \frac{1}{n} \sum_{t \in T} x_t$;
- $\bar{t}$ este media valorilor lui t, adică $\bar{t} = \frac{1}{n} \sum_{t \in T} t$

Considerând că valorile seriei temporale din trecut sunt un bun indicator pentru valorile viitoare, putem spune că ecuaţia:

$$T_t = a + bt$$

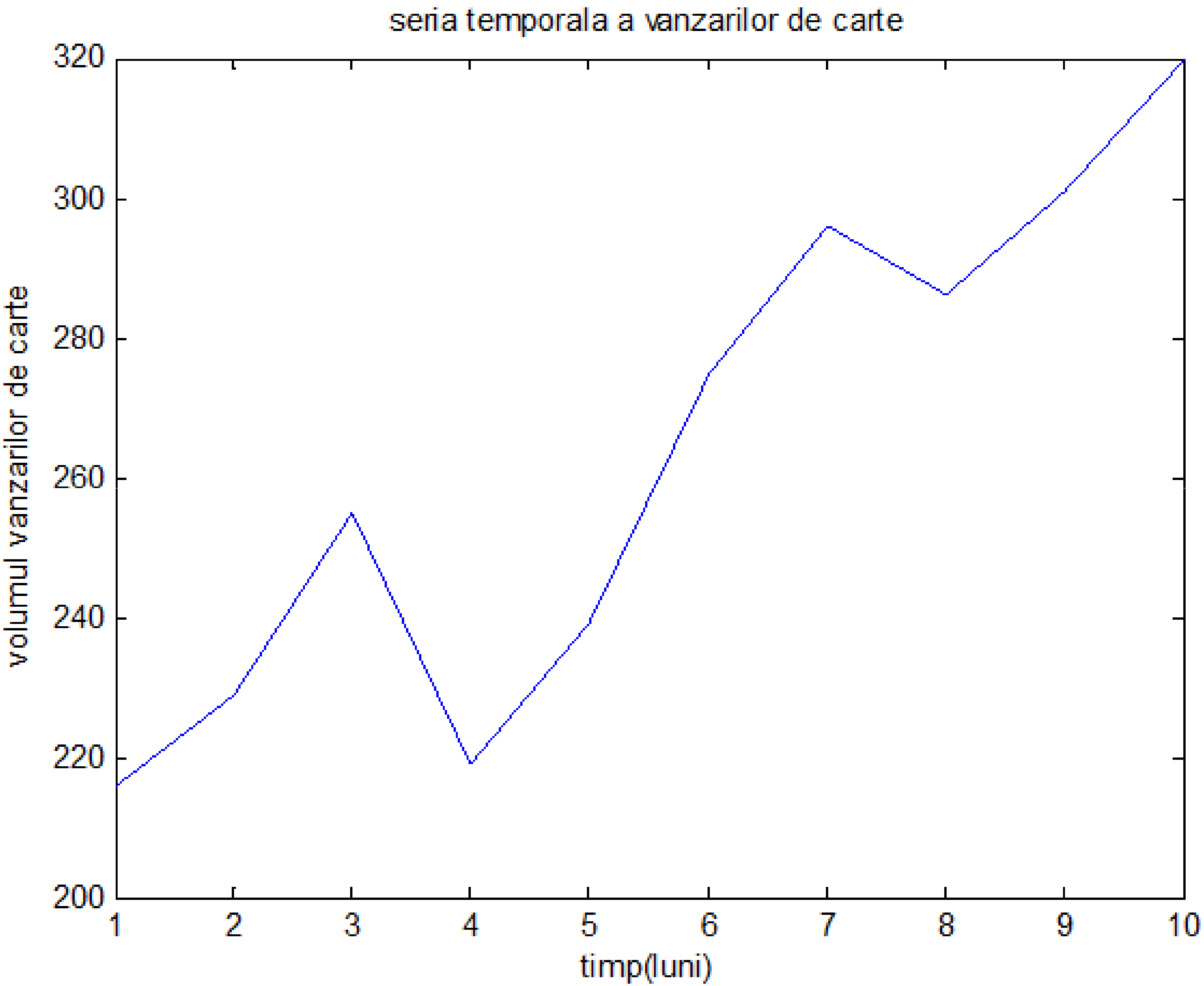
poate fi utilizată pentru a prognoza valori viitoare.

7.Exemplu

Vânzările de carte (număr de cărţi) dintr-o librărie virtuală în ultimele 10 luni sunt:

luna	1	2	3	4	5	6	7	8	9	10
vânzări	216	229	255	219	239	275	296	286	301	320

```
>> t=1:10;
>>x=[216 229 255 219 239 275 296 286 301 320];
>>plot(t,x,'k')
```

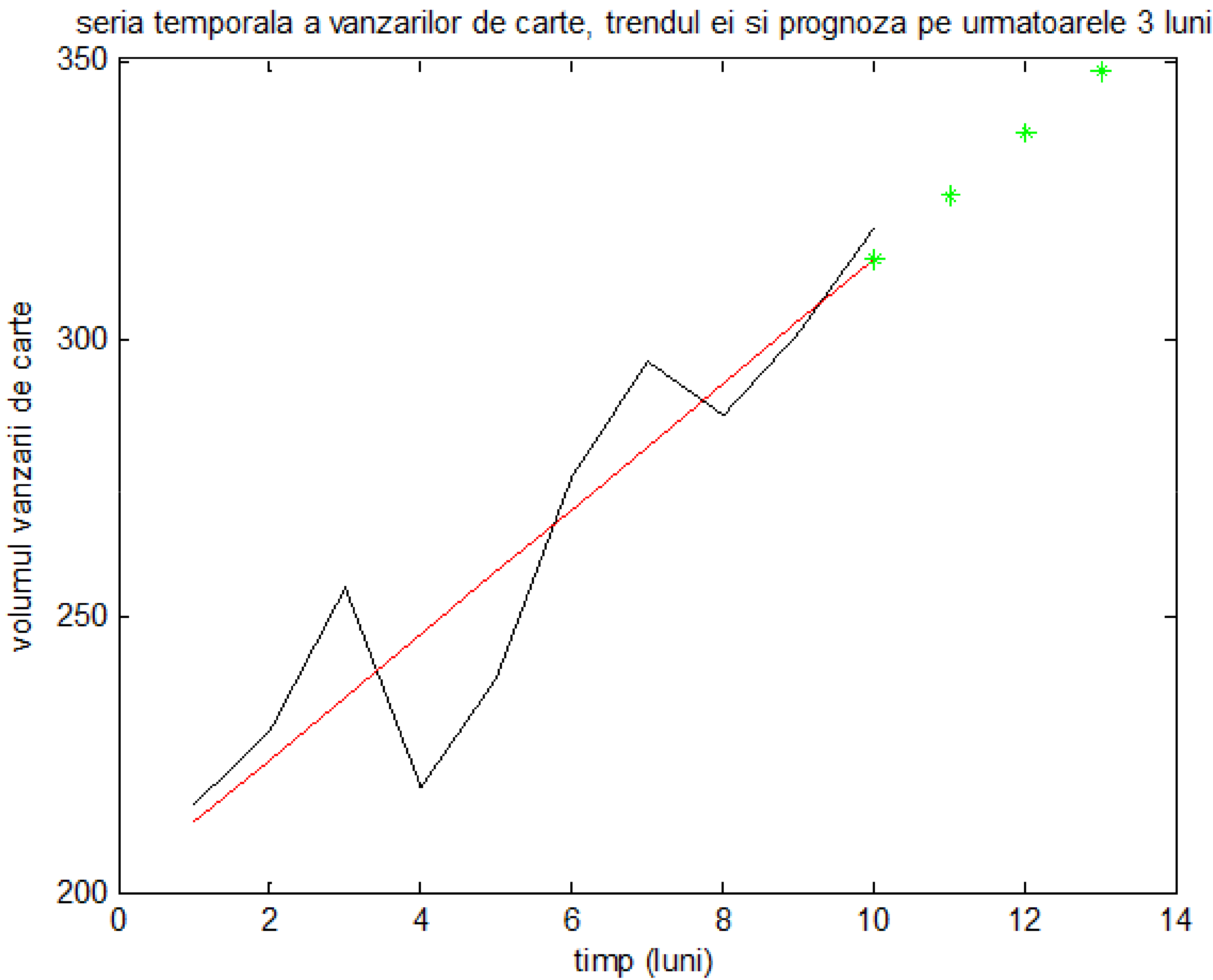

seria temporala a vanzarilor de carte
volumul vanzarilor de carte
timp(luni)
320
300
280
260
240
220
200
1
2
3
4
5
6
7
8
9
10

Determinăm ecuaţia trendului acestei serii temporale:

```
>>b=(t*x'-10*mean(t)*mean(x))./(t*t'-10*mean(t)^2)
b =
       11.2848
>>a=mean(x)-b*mean(t)
a =
     201.5333
```

şi astfel folosind trendul putem prognoza volumul vânzărilor pe următoarele 3 luni:

```
>> T11=a+11*b
T11 =
       325.6667
>> T12=a+12*b
T12 =
       336.9515
>> T13=a+13*b
T13 =
       348.2364

>>T=a+b*t;
>> plot(t,x,'k',t,T,'r', 10,T(10),'g*',11,T11,'g*',12,T12,'g*',13,T13,'g*')
```



seria temporala a vanzarilor de carte, trendul ei si prognoza pe urmatoarele 3 luni
volumul vanzari de carte
timp (luni)
350
300
250
200
0
2
4
6
8
10
12
14

Prognoza utilizând trendul şi componenta sezonieră

Considerăm că valoarea reală a seriei temporale, $\{x_t, t \in T\}$ poate fi descrisă prin *modelul multiplicativ* de o serie temporală:

$$x_t = T_t \cdot S_t \cdot I_t \,,$$

unde T_t, S_t, I_t este trendul, componenta sezonieră, respectiv cea incidentală corespunzătoare perioadei de timp t.

Componentă incidentală este cea care dă efectele aleatoare din seria temporală, efecte care nu pot fi explicate de trend sau componenta sezonieră.

Tehnica de lucru constă în:

- eliminarea componentei sezoniere;
- realizarea prognozei utilizând trendul;
- reintroducerea componentei sezoniere în scopul ajustării prognozei.

Dacă reuşim să eliminăm efectele sezoniere ale unei serii temporale, putem de exemplu compara valorile seriei temporale în perioade succesive.
Dacă nu, putem face comparaţii între valorile seriei temporale dintr-o anumită perioadă din acest an şi aceeaşi perioadă a anului trecut.

Considerăm o serie temporală $\{x_t, t \in T\}$, unde $T = \{1, 2, ..., 4 \cdot n\}$, unde n este numărul de ani luaţi în considerare ($4 \cdot n$ fiind numărul de trimestre), de forma $x_t = T_t \cdot S_t \cdot I_t$.

1. Existenţa pattern-ului sezonier impune identificarea computaţională a fiecărei influenţe trimestriale sezoniere. Calculăm o medie mobilă pentru a izola componenta sezonieră S şi componenta incidentală I. Folosim datele din fiecare an şi anume cele 4 trimestre:

$$media\ mobila\ (1) = \frac{x_1 + x_2 + x_3 + x_4}{4},$$

$$media\ mobila\ (2) = \frac{x_2 + x_3 + x_4 + x_5}{4},...$$

Valorile calculate nu corespund trimestrelor din seria temporală. Valoarea *media mobila* (1) corespunde ultimei jumătăţi a trimestrului al 2-lea şi primei jumătăţi a trimestrului al 3-lea în timp ce *media mobila* (2) corespunde ultimei jumătăţi a trimestrului al 3-lea şi primei jumătăţi a trimestrului al 4-lea.

În aceste condiții vom considera media mobilă corespunzătoare trimestrului al 3-lea ca fiind media aritmetică a celor două medii calculate, așa numita *medie mobilă centrată* (*centered moving average- CMA*):

$$CMA(3) = \frac{media\ mobila(1) + media\ mobila(2)}{2}$$

Continuăm calculul asociind următoarelor trimestre ($4 \le i \le 4 \cdot n - 2$) mediile mobile centrate corespunzătoare $CMA(i)$, $4 \le i \le 4 \cdot n - 2$, obținând astfel seria temporală corespunzătoare mediilor mobile centrate.

Această serie temporală este „netezită" în sensul că nu apar influențele sezoniere și incidentale.

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.

2. Împărţind fiecare observaţie din seria temporală cu media mobilă centrată corespunzătoare putem identifica efectul sezonier – incidental al seriei temporale.
Astfel *componenta sezonier-incidentală* corespunzătoare celui de-al i-lea trimestrului este:

$$CSI(i) = \frac{x_i}{CMA(i)},\ 3 \le i \le 4 \cdot n - 2\,.$$

Să considerăm componenta sezonier-incidentală pentru al 3-lea trimestru din primii $n-1$ ani:
Deoarece putem presupune că fluctuaţiile de la an la an ale componentei sezonier-incidentale se datorează în principal componentei incidentale, putem calcula media celor $n-1$ valori, eliminând astfel influenţa incidentală.

Se obţine astfel o estimare a influenţei sezoniere din al 3-lea trimestru, estimare cunoscută sub numele de *indicele sezonier în trimestrul* al treilea - $IS(3)$:

$$IS(3) = \frac{\sum_{j=0}^{n-2} CS(4j+3)}{n-1}.$$

Analog calculăm indicii sezonieri pentru fiecare trimestru în parte:

$$IS(4) = \frac{\sum_{j=0}^{n-2} CS(4j+4)}{n-1};$$

$$IS(1) = \frac{\sum_{j=1}^{n-1} CS(4j+1)}{n-1};$$

$$IS(2) = \frac{\sum_{j=1}^{n-1} CS(4j+2)}{n-1};$$

Modelul multiplicativ cere ca media indicilor sezonieri să fie 1, condiţie ce nu este îndeplinită în general.
Pentru a obţine media indicilor sezonieri egală cu unitatea este necesară o ajustare, prin împărţirea fiecărui indice sezonier la media indicilor sezonieri:

$$IS_a(i) = \frac{IS(i)}{\frac{1}{4}\sum_{i=1}^{4} IS(i)}, 1 \leq i \leq 4 .$$

3. Împărţind fiecare observaţie (termen) al seriei temporale la indicele sezonier corespunzător ajustat înlăturăm efectul sezonier al seriei temporale:

$$x'_{4k+i} = \frac{x_{4k+i}}{IS_a(i)}, 0 \le k \le n-1, 1 \le i \le 4 .$$

4. Putem estima trendul liniar al seriei temporale din care am eliminat componenta sezonieră $\{x_j', j \in T = \{1,2,....,4 \cdot n\}\}$.

$$T_t = a + bt,$$

unde:

$$b = \frac{\sum_{j=1}^{4n} j \cdot x_j' - \frac{1}{4n}\sum_{j=1}^{4n} j \cdot \sum_{i=1}^{4n} x_j'}{\sum_{j=1}^{4n} j^2 - \frac{1}{4n}\left(\sum_{j=1}^{4n} j\right)^2}\ ;\ a = \frac{1}{4n}\left(\sum_{j=1}^{4n} x_j' - b \cdot \sum_{j=1}^{4n} j\right).$$

Putem prognoza, pe baza trend-ului, comportarea seriei temporale pentru cele patru trimestre din următorul an (al $n+1$-lea):

$$T_{4n+i} = a + b(4n+i),\ i \in \{1,2,3,4\};$$

Luăm în considerare efectul sezonier asupra prognozei, înmulţind T_i cu indicele sezonier ajustat:

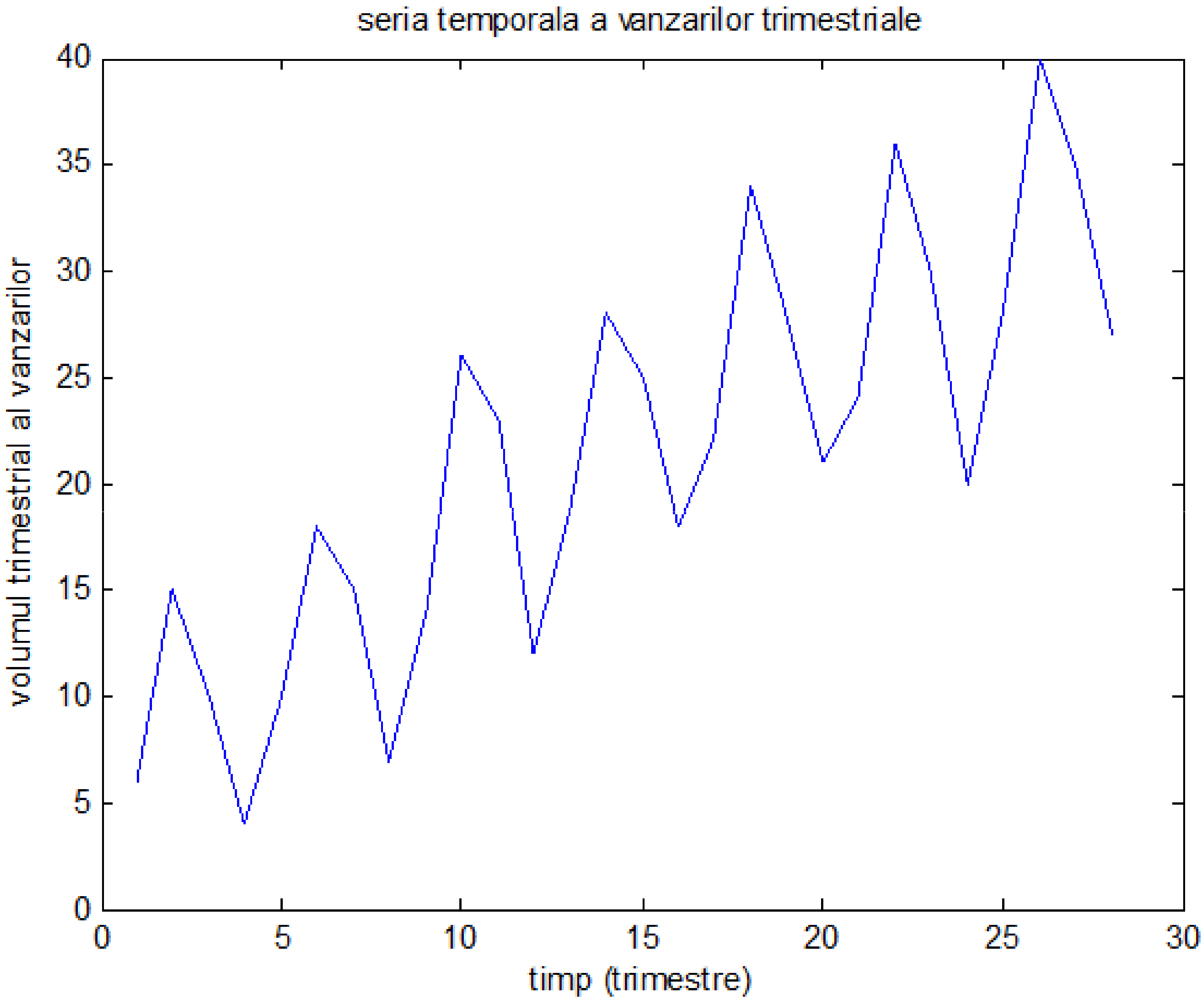
$$T'_{4n+i} = IS_a(i) \cdot T_{4n+i},\ i \in \{1,2,3,4\}.$$

8. Exemplu

Firma X este dealer autorizat al firmei pentru vânzarea de automobile. Următorul tabel prezintă situaţia vânzărilor trimestriale pe ultimii 7 ani:

anul	trim I	trim II	trim III	trim IV	total
1	6	15	10	4	35
2	10	18	15	7	50
3	14	26	23	12	75
4	19	28	25	18	90
5	22	34	28	21	105
6	24	36	30	20	110
7	28	40	35	27	130

```
>> t=1:28;
>>V=[6 15 10 4 10 18 15 7 14 26 23 12 19 28 25 18 22 34 28 21 24 36 30
20 28 40 35 27];% V=vanzari trimestriale
>> plot(t,V)
```


seria temporala a vanzarilor trimestriale
volumul trimestrial al vanzarilor
timp (trimestre)
0
5
10
15
20
25
30
35
40

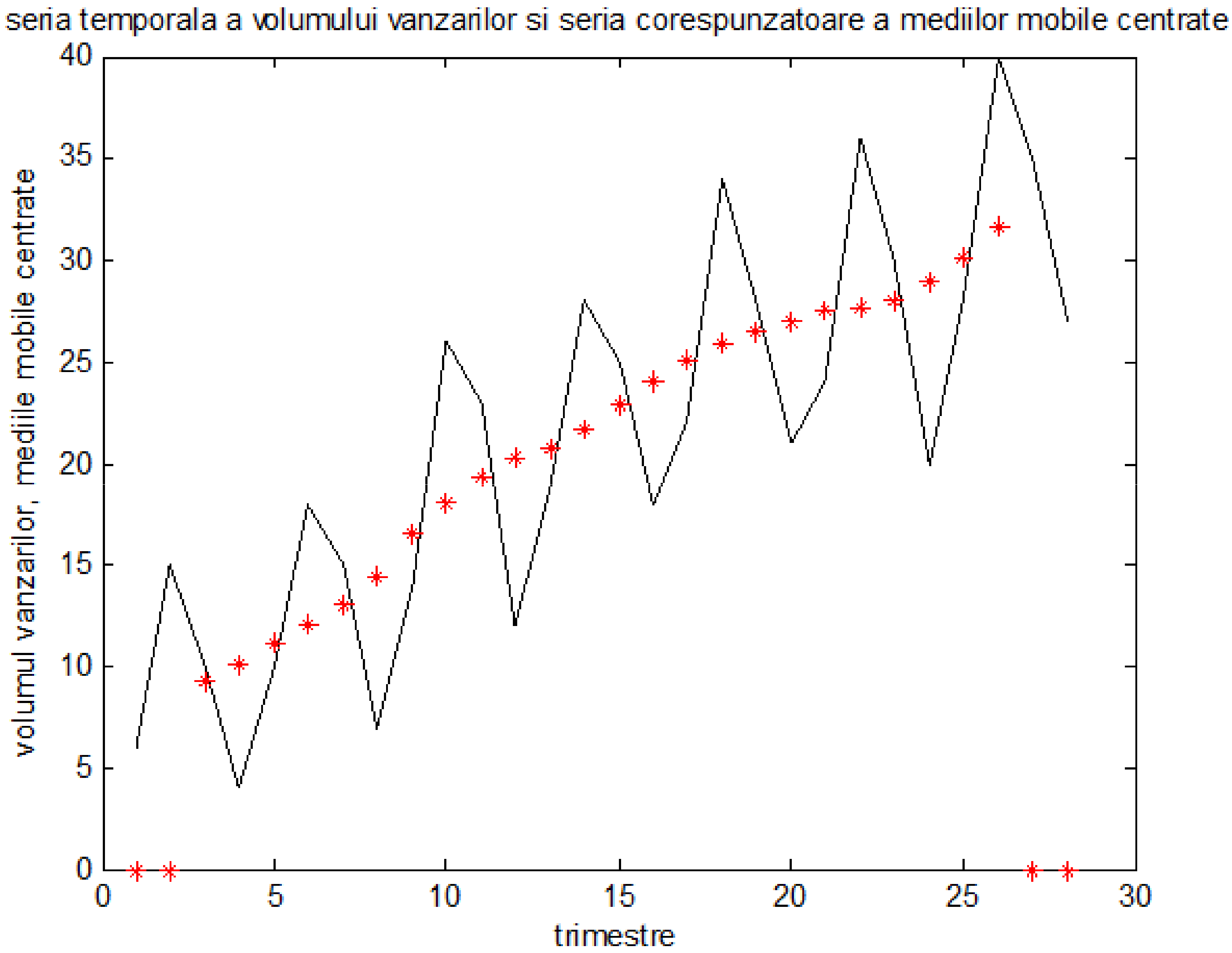
Se observă că în trimestrul al patrulea al fiecărui an volumul vânzărilor este minim, în timp ce în trimestrul al doilea al fiecărui an volumul creşte, ceea ce indică existenţa unui pattern sezonier.

Vom determina mediile mobile pe 4 trimestre, mediile mobile centrate.

```
>> t=1:28;V=[6 15 10 4 10 18 15 7 14 26 23 12 19 28 25 18 22 34 28 21 24
36 30 20 28 40 35 27];% V=vanzari trimestriale
>> for i=5:29 M(i-2)=1/4*(V(i-4)+V(i-3)+V(i-2)+V(i-1));end % M media mobila
trimestriala
>> M(28)=0;
>> for i=3:26 MC(i)=1/2*(M(i)+M(i+1)); end % MC media mobila centrata
>> for i=27:28 MC(i)=0;end
```

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

```
>> plot(t,V,'k',t,MC,'r*')
```


seria temporala a volumului vanzarilor si seria corespunzatoare a mediilor mobile centrate
volumul vanzarilor, mediile mobile centrate
trimestre
0
5
10
15
20
25
30
35
40

Calculăm componenta sezonier-incidentală:

```
>> for i=3:26 CS(i)=V(i)/MC(i);end
>> for i=27:28 CS(i)=0;end %CS componenta sezoniera
>> [V' M' MC' CS'];
```

Calculăm indicii sezonieri pentru fiecare trimestru, şi indicii sezonieri ajustaţi:

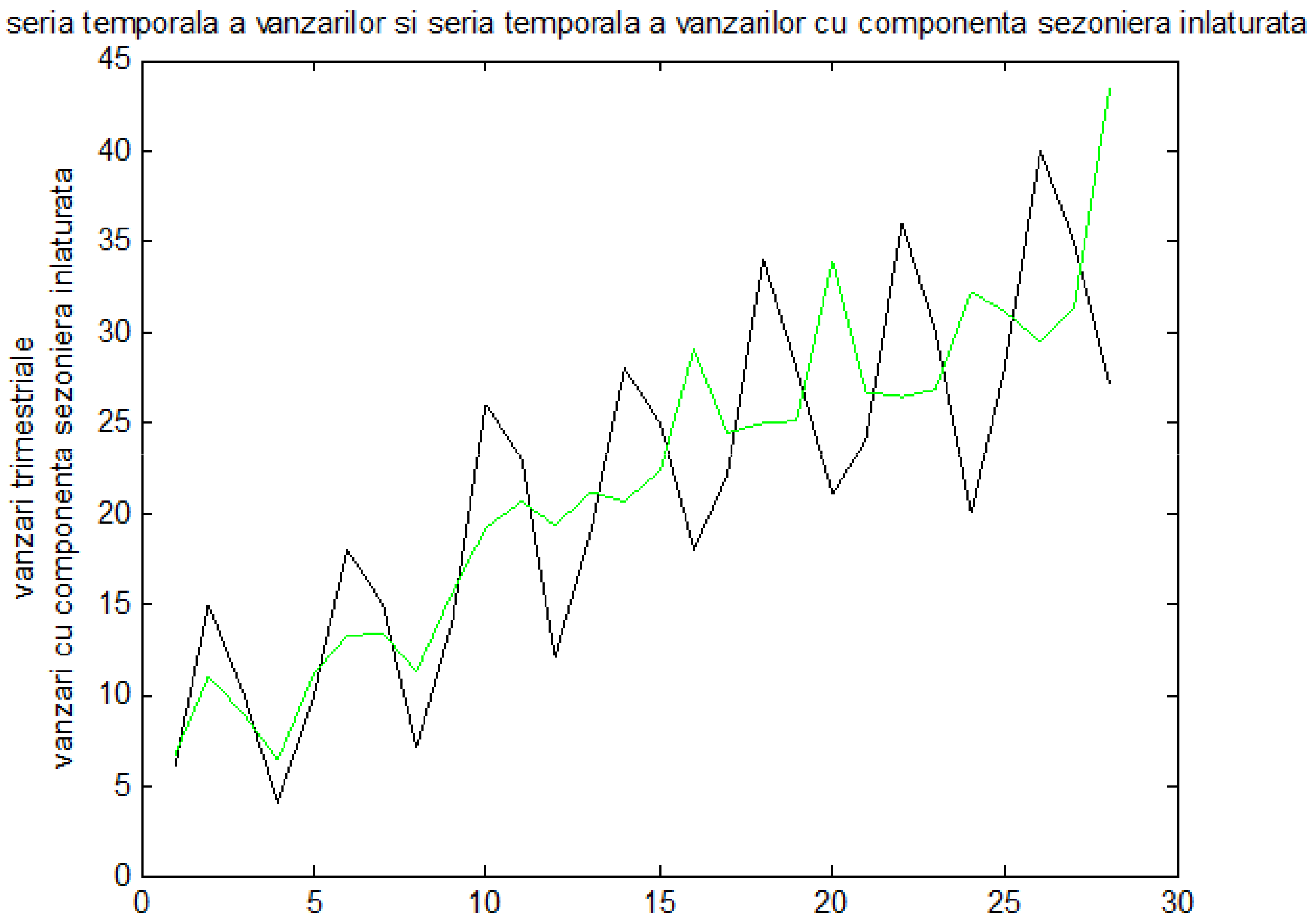
```
>> for i=1:2
I(i)=1/6*(CS(i+4)+CS(i+8)+CS(i+12)+CS(i+16)+CS(i+20)+CS(i+24));end
>>for i=3:4
I(i)=1/6*(CS(i)+CS(i+4)+CS(i+8)+CS(i+12)+CS(i+16)+CS(i+20));end % I
indicele sezonier
>> I
I =
    0.8901    1.3481    1.1072    0.6146
>> Ia=I/mean(I) %Ia indicele sezonier ajustat
Ia =
    0.8991    1.3617    1.1183    0.6208
```

Observăm că cele mai bune vânzări sunt în al 2-lea trimestru, cu 36.17% peste medie. Cel mai slab trimestru pentru vânzări este al 4-lea cu 37.92% sub medie.

Împărţind fiecare observaţie (termen) al seriei temporale la indicele sezonier corespunzător ajustat, înlăturăm efectul sezonier al seriei temporale.

Vom prezenta graficele seriei vânzărilor trimestriale şi al seriei vânzărilor trimestriale cu efectul sezonier înlăturat.

```
>> IS=[Ia Ia Ia Ia Ia Ia Ia];
>> for i=1:28 VDS(i)=V(i)/IS(i);end % VDS vanzari cu factorul sezonier indepartat
>>plot(t, V,'k',t,VDS,'g.')
```


seria temporala a vanzarilor si seria temporala a vanzarilor cu componenta sezoniera inlaturata
vanzari trimestriale
vanzari cu componenta sezoniera inlaturata
0
5
10
15
20
25
30
35
40
45
0
5
10
15
20
25
30

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

```
>> b=(t*VDS'-28*mean(t)*mean(VDS))./(norm(t)^2-28*mean(t)^2)
b =
   1.0547
>> a=mean(VDS)-b*mean(t)
a =
   6.3315
```

Astfel ecuaţia trendului liniar al seriei temporale este:

$$T_t = 6.3315 + 1.0547 \cdot t$$

Prognozăm pe baza trendului volumul vânzărilor în cel de-al 8-lea an şi luăm apoi în considerare şi efectul sezonier:

```
>> for t=29:32; Vp(t)=a+b*t;end %Vp vanzari prognozate
>> for i=1:4 VP(i+28)=Vp(i+28)*I(i);end % VP vanzari prognozate tinand
 cont de indicele sezonier
>> [ Vp'  VP']
ans =
   36.9166   32.8609
   37.9712   51.1876
   39.0259   43.2079
   40.0806   24.6352
```

Considerăm o serie temporală $\{x_t, t \in T\}$ unde $T = \{1, 2, ..., 12 \cdot n\}$, unde n este numărul de ani luați în considerare ($12 \cdot n$ fiind numărul de luni), de forma $x_t = T_t \cdot S_t \cdot I_t$, serie ce prezintă un pattern sezonier.

1. Vom identifica fiecare influență lunară sezonieră, pentru a o izola, calculând o medie mobilă pe 12 luni:

$$media\ mobila\ (1) = \frac{\sum_{i=1}^{12} x_i}{12},$$

$$media\ mobila\ (2) = \frac{\sum_{i=2}^{13} x_i}{12}, \ldots$$

Vom considera media mobilă corespunzătoare trimestrului al 11-lea ca fiind media aritmetică a celor două medii calculate, aşa numita *medie mobilă centrată* (*centered moving average- CMA*):

$$CMA(11) = \frac{media\ mobila(1) + media\ mobila(2)}{2}$$

Asociem următoarelor trimestre ($12 \le i \le 12 \cdot n - 2$) mediile mobile centrate corespunzătoare $CMA(i)$, $12 \le i \le 12 \cdot n - 2$.

Am obţinut astfel seria temporală corespunzătoare mediilor mobile centrate, serie care este „netezită" în sensul că nu apar influenţele sezoniere şi incidentale.

2. Fiecare observație din seria temporală se împarte cu media mobilă centrată corespunzătoare pentru a identifica efectul sezonier –incidental al seriei temporale.

Componenta sezonier-incidentală corespunzătoare celui de-al i-lea trimestru este:

$$CSI(i) = \frac{x_i}{CMA(i)},\ 11 \le i \le 12 \cdot n - 2$$

Să considerăm componenta sezonier-incidentală pentru luna a 11-a din primii $n-1$ ani. Fluctuațiile de la an la an ale componentei sezonier-incidentale se datorează în principal componentei incidentale, astfel putem calcula media celor $n-1$ valori, eliminând astfel influența incidentală.

Se obține o estimare a influenței sezoniere din al luna a 11-a, estimare cunoscută sub numele de *indicele sezonier în luna a* 11-a- $IS(11)$:

$$IS(11) = \frac{\sum_{j=0}^{n-2} CS(12j+11)}{n-1}$$

Analog calculăm indicii sezonieri pentru fiecare lună în parte:

$$IS(12) = \frac{\sum_{j=0}^{n-2} CS(12j+12)}{n-1};$$

$$IS(k) = \frac{\sum_{j=1}^{n-1} CS(4j+k)}{n-1}, 1 \le k \le 10;$$

Pentru a obține media indicilor sezonieri egală cu unitatea, împărțim fiecare indice sezonier la media indicilor sezonieri.

$$IS_a(i) = \frac{IS(i)}{\frac{1}{12}\sum_{i=1}^{12} IS(i)}, 1 \le i \le 12$$

3. Împărţind fiecare observaţie (termen) al seriei temporale la indicele sezonier corespunzător ajustat, înlăturăm efectul sezonier al seriei temporale.

$$x'_{12k+i} = \frac{x_{12k+i}}{IS_a(i)}, 0 \le k \le n-1, 1 \le i \le 12$$

4. Putem estima trendul liniar al seriei temporale căreia i-am eliminat componenta sezonieră $\{x_j', j \in T = \{1,2,....,12 \cdot n\}\}$:

$$T_t = a + bt,$$

unde:

$$b = \frac{\sum_{j=1}^{12n} j \cdot x_j' - \frac{1}{12n} \sum_{j=1}^{12n} j \cdot \sum_{i=1}^{12n} x_j'}{\sum_{j=1}^{12n} j^2 - \frac{1}{12n}\left(\sum_{j=1}^{12n} j\right)^2}; \; a = \frac{1}{12n}\left(\sum_{j=1}^{12n} x_j' - b \cdot \sum_{j=1}^{12n} j\right)$$

$\{x_j', j \in T = \{1,2,....,12 \cdot n\}\}$ este seria temporală fără componenta sezonieră.

Putem prognoza, pe baza trendului, comportarea seriei temporale pentru următoarele 12 luni:

$$T_{4n+i} = a + b(4n+i), 1 \leq i \leq 12\,;$$

Luăm în considerare efectul sezonier asupra prognozei, înmulţind T_t cu indicele sezonier ajustat:

$$T'_{4n+i} = IS_a(i) \cdot T_{4n+i}\,,\ 1 \leq i \leq 12$$

Exemplu

Reluăm exemplul nr 4 cu numărul de pasageri (în mii) înregistrat pentru zborurile internaţionale pe un aeroport, din ultimele 36 de luni. Folosind tehnica prezentată dorim să prognozăm numărul de pasageri pentru următorul an.

```
>>N=[31 22 21 34 29 40 54 62 45 21 23 41 33 24 23 28 35 45 60 65 44 23
24 48 34 26 25 35 32 53 70 74 50 27 28 51];% N numarul de pasageri (mii)
>> for i=13:37
M(i-2)=1/12*(N(i-12)+N(i-11)+N(i-10)+N(i-9)+N(i-8)+N(i-7)+N(i-6)+N(i-5)+N(i-
4)+N(i-3)+N(i-2)+N(i-1));
end % M media mobila lunara
>> M(36)=0;
>> for i=11:34
MC(i)=1/2*(M(i)+M(i+1));
end % MC media mobila centrata
>> for i=35:36
MC(i)=0;
end
```

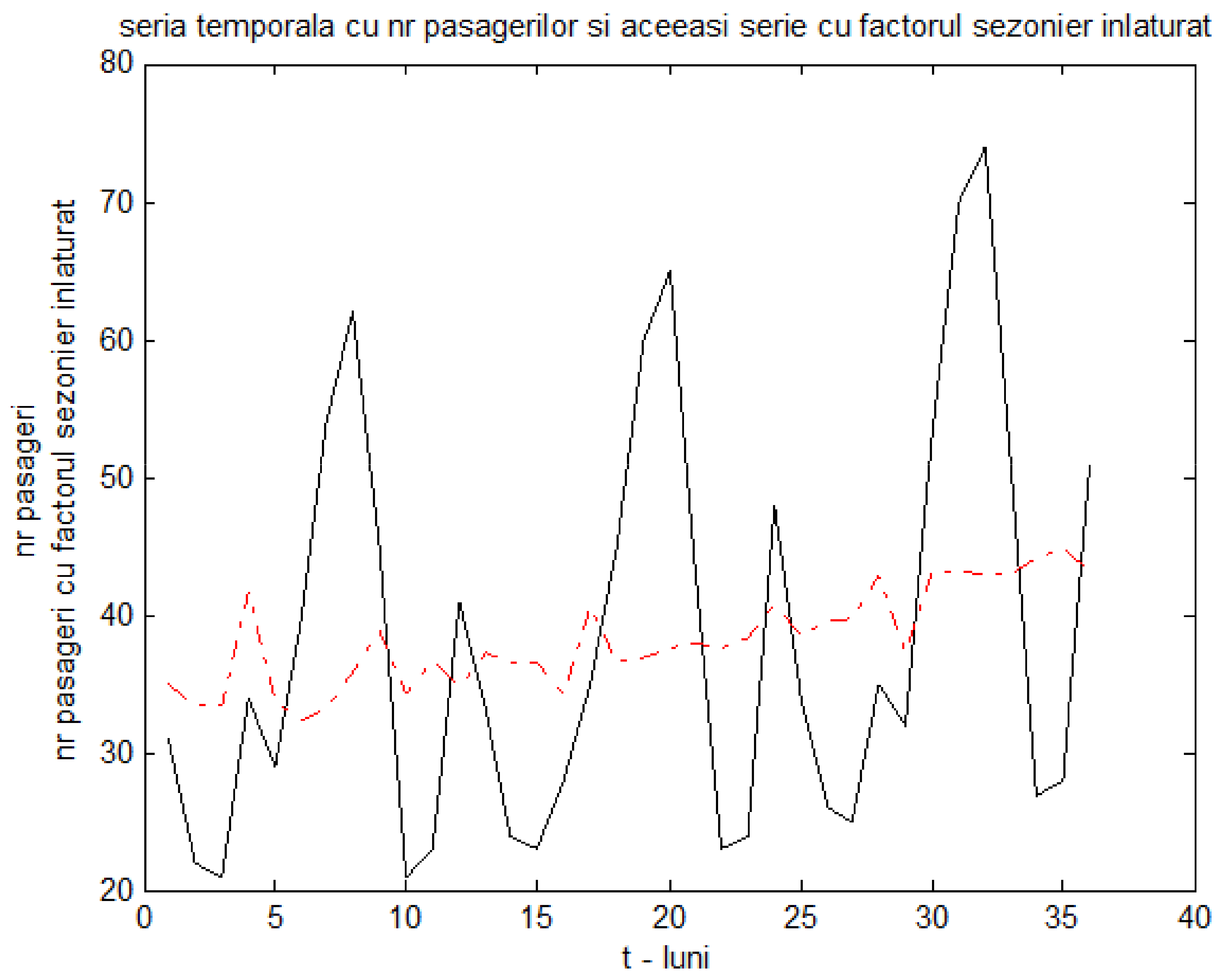
Exemplu

Reluăm exemplul nr 4 cu numărul de pasageri (în mii) înregistrat pentru zborurile internaţionale pe un aeroport, din ultimele 36 de luni. Folosind tehnica prezentată dorim să prognozăm numărul de pasageri pentru următorul an.

```
>>N=[31 22 21 34 29 40 54 62 45 21 23 41 33 24 23 28 35 45 60 65 44 23
24 48 34 26 25 35 32 53 70 74 50 27 28 51];% N numarul de pasageri (mii)
>> for i=13:37
M(i-2)=1/12*(N(i-12)+N(i-11)+N(i-10)+N(i-9)+N(i-8)+N(i-7)+N(i-6)+N(i-5)+N(i-
4)+N(i-3)+N(i-2)+N(i-1));
end % M media mobila lunara
>> M(36)=0;

>> for i=11:34
MC(i)=1/2*(M(i)+M(i+1));
end % MC media mobila centrata
>> for i=35:36
MC(i)=0;
end
```

```
>> for i=11:34
NS(i)=N(i)/MC(i);
end

>> for i=35:36
 NS(i)=0;
end %NS componenta sezoniera

>> [N' M' MC' NS'];

>> for i=1:10
 I(i)=1/2*(NS(i+12)+NS(i+24));
end

>> for i=11:12
 I(i)=1/2*(NS(i)+NS(i+12));
end % I indicele sezonier

>> I
    0.9100    0.6768    0.6483    0.8410    0.8856    1.2693    1.6690    1.7755
1.1939    0.6294    0.6437    1.2118
```

```
>> Ia=I/mean(I) ; %Ia indicele sezonier ajustat
Ia =
   0.8839    0.6574    0.6297    0.8168    0.8602    1.2329    1.6211    1.7246
1.1597    0.6114    0.6252    1.1771
>> IS=[Ia Ia Ia ];
>> for i=1:36
NDS(i)=N(i)/IS(i);
end % NDS nr pasageri cu factorul sezonier indepartat
>> [N' NDS'];

>> t=1:36; plot(t, N,'k',t,NDS,'k -.')
```


seria temporala cu nr pasagerilor si aceeasi serie cu factorul sezonier inlaturat
nr pasageri
nr pasageri cu factorul sezonier inlaturat
t - luni
0
5
10
15
20
25
30
35
40
20
30
40
50
60
70
80

Calculăm trendul:

```
>> b=(t*NDS'-28*mean(t)*mean(NDS))./(norm(t)^2-28*mean(t)^2)
b =
    1.0161
>> a=mean(NDS)-b*mean(t)
a =
   19.4582
```

Facem prognoza pe urmărtorul an:

```
>> for t=37:48;
Np(t)=a+b*t;
end %Np numar depasageri prognozat
>> for i=1:12
 NP(i+36)=Np(i+36)*I(i);
end % NP numar de pasageri prognozat tinand cont de indicele sezonier
```

```
>> [ Np'  NP']
ans =
   57.0525   51.9169
   58.0685   39.3002
   59.0846   38.3027
   60.1007   50.5418
   61.1167   54.1257
   62.1328   78.8624
   63.1488  105.3927
   64.1649  113.9262
   65.1810   77.8212
   66.1970   41.6671
   67.2131   43.2653
   68.2291   82.6819
```

Modele dinamice, bazate pe serii temporale

Seriile temporale mai sunt cunoscute şi sub numele de *serii dinamice*, datorită faptului că ilustrează cinetica (dinamica) unui fenomen evoluând în timp real.

De aceea, în conexiune cu seriile temporale, putem vorbi şi de *modelele dinamice*, adică modelele ce captează *mişcarea* unui anumit fenomen în raport cu timpul.

Modelele dinamice explică comportamentul actual al variabilei dependente cu ajutorul valorilor actuale şi din trecut ale variabilelor.

De exemplu modelul

$$y_t = \beta_1 + \beta_2 \cdot x_t + \beta_3 \cdot x_{t-1} + u_t$$

arată că y_t este explicitat prin valoarea actuală a variabilei x, x_t şi prin valoarea din trecut x_{t-1}.

Se iau în considerare:

- modele cu valori din trecut ale variabilelor independente; aceste modele sunt generalizări ale modelelor regresive, prin adăugarea valorii x_{t-1} în lista variabilelor independente (predictorilor). Specificul acestui tip de model este introducerea unei variabile aleatoare în lista variabilelor independente.
- modele cu valori din trecut ale variabilelor dependente, modele ce includ valori din trecut ale variabilei dependente, de exemplu:

$$y_t = \gamma \cdot y_{t-1} + \beta_1 + \beta_2 \cdot x_{2t} + \ldots + \beta_k \cdot x_{kt} + u_t \quad (*).$$

Modele explicative

Aceste modele sunt caracterizate de ecuaţia:

$$y_t = f(x_t, u_t)$$

unde f este o funcţie determinată de un număr finit de parametri necunoscuţi, x_t este o variabilă observabilă, numită şi *variabilă exogenă*, iar u_t reprezintă o variabilă aleatoare de medie zero, aleasă în funcţie de situaţia reală modelată (factor de disturbare/zgomot/eroare).

În principal, aceste modele se împart în două cazuri:

- modele *statice*, în care variabila exogenă x_t nu conţine informaţii din trecutul lui y_t, iar u_t sunt mutual independente,
- modele *dinamice* în care fie x_t conţine informaţii privind trecutul lui y_t, fie u_t sunt autocorelate.

9. Exemplu

Agenţii economici iau decizii asupra viitorului, bazându-se pe propriile previziuni asupra unor condiţii necunoscute.

De exemplu, dacă se decide reabilitarea sistemului de încălzire, bazându-se pe preţul preconizat al combustibililor şi al echipamentelor, cel mai simplu model fiind

$$y_t = \beta_1 + \beta_2 \cdot x_t^p + u_t \quad (1)$$

unde x_t^p este preţul preconizat.

Valorile preconizate ale prețului sunt stabilite pe baza situațiilor existente în trecut și astfel putem modela prețul preconizat x_t^p astfel:

$$x_t^p = x_{t-1}^p + \lambda(x_{t-1} - x_{t-1}^p),\ \lambda \in [0,1] \quad (2)$$

unde x_{t-1} este prețul echipamntelor și combustibilor în momentul t-1 și x_{t-1}^p este prețul preconizat pentru momentul t-1.

Așadar prețul preconizat este suma prețului preconizat pentru momentul anterior și eroarea observată în prognozarea prețului în momentul anterior, ponderată cu parametrul $\lambda \in [0,1]$.

Dinamica din construcţia pretului prognozat, duce la o dinamică în datele observate:

Conform (1) avem:

$$y_{t-1} = \beta_1 + \beta_2 \cdot x^p_{t-1} + u_{t-1}$$

şi astfel:

$$x^p_{t-1} = \frac{1}{\beta_2} \cdot (y_{t-1} - \beta_1 - u_{t-1})$$

Conform (2):

$$x^p_t = \lambda \cdot x_{t-1} + (1-\lambda) \cdot x^p_{t-1} = \lambda \cdot x_{t-1} + \frac{1-\lambda}{\beta_2}(y_{t-1} - \beta_1 - u_{t-1}) \quad (3)$$

Înlocuind în (1) valoarea obţinută pentru preţul preconizat în (3), obţinem:

$$y_t = (1-\lambda) \cdot y_t + \lambda\beta_1 + \lambda\beta_2 \cdot x_{t-1} + (u_t + (1-\lambda)u_{t-1})$$

Notând $\varepsilon_t = u_t + (1-\lambda)u_{t-1}$, obţinem ecuaţia unui model de tip (*).

Modele de ajustare

În acest caz, pe baza observaţiilor obţinute analizând datele reale, putem formula un model matematic ilustrat de o ecuaţie de forma:

$$y_t = f(t, u_t)$$

unde f este o funcţie determinată de un număr finit de parametri necunoscuţi, iar u_t reprezintă o variabilă aleatoare de medie zero, aleasă în funcţie de situaţia reală modelată.

Vom menţiona că ipotezele asupra variabilei u_t ca şi estimarea funcţiei f se fac plecând de la:

1. *ajustările globale*, în care toate observaţiile se bucură de aceeaşi consideraţie, având roluri egale în estimaţii, sau
2. *ajustările locale*, în care fiecare observaţie îşi are rolul său în determinarea parametrilor modelului.

10. Exemplu

În cazul în care agenții economici nu se pot adapta pe deplin la schimbarea condițiilor , are loc un fenomen similar cu cel din exemplul nr 9.

De exemplu, o dată cu trecerea anilor, o firmă are nevoie de lărgirea sediului. Costurile necesare modificărilor dorite restricționează mărimea construcției. Putem modela dimensiunile dorite utilizând regresia liniară:

$$y_t^* = \beta_1 + \beta_2 \cdot x_t + u_t$$

Vom face legătura între dimensiunea dorită şi cea observată printr-un proces de ajustare parţială:

$$y_t - y_{t-1} = \lambda \cdot (y_t * - y_{t-1}) + v_t, \quad \lambda \in [0,1]$$

ceea ce va conduce la o relaţie între y şi x:

$$y_t = (1-\lambda) y_{t-1} + \lambda \beta_1 + \lambda \beta_2 \cdot x_t + \varepsilon_t, \quad \lambda \in [0,1]$$

unde am notat $\varepsilon_t = v_t + \lambda \cdot u_t$

Modele autopredictive

În aceste modele se presupune că prezentul este influenţat de trecut, deci matematic vorbind, un asemenea model este ilustrat de o ecuaţie de forma:

$$y_t = f(y_{t-1}, y_{t-2}, \ldots, u_t),$$

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

Să considerăm cel mai simplu model autoregresiv

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t$$

care poate fi scris astfel:

$$\begin{aligned} y_t &= \beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-2} + u_{t-1}) + u_t = \\ &= \beta_0 + \beta_1\beta_0 + \beta_1^{\,2} y_{t-2} + \beta_1 u_{t-1} + u_t = \\ &= \beta_0(1 + \beta_1 + \ldots + \beta_1^{\,n}) + \beta_1^{n-1} y_{t-n} + u_t + \beta_1 u_{t-1} + \beta_1^n u_{t-n} \end{aligned}$$

Dacă $|\beta_1| \geq 1$, atunci y_t va creşte exponenţial

Dacă $|\beta_1| < 1$, procesul se *stabilizează* deoarece:

$$\beta_0(1+\beta_1+\ldots+\beta_1^{\,n}) \underset{n\to\infty}{\longrightarrow} \frac{\beta_0}{1-\beta_1}$$

$$\beta_1^{n-1} y_{t-n} \underset{n\to\infty}{\longrightarrow} 0$$

$$u_t+\beta_1 u_{t-1}+\beta_1^n u_{t-n} \underset{n\to\infty}{\longrightarrow} N(0,\frac{\sigma^2}{1-\beta_1^2})$$

In cazul modelelor complicate, varianta cea mai simplă pentru studiul stabilităţii, este simularea modelului.

Dacă modelul este instabil, simulările vor avea o formă explozivă.

Dacă însă este stabil, simulările vor converge.

Modelul *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.
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.

Modelul ARIMA include trei tipuri de parametri:

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

Astfel, un model ARIMA în notaţia 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 şi q = 1 parametri medie mobilă.
Sunt cazuri în care nu este nevoie de transformarea diferenţiată, adică. d = 0.
Cel mai adesea aceasta este aleasă d = 1 sau d = 2, atunci când este într-adevăr necesară.
În ceea ce priveşte alegerea celorlalţi doi parametri, valorile uzuale rareori depăşesc 2.

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

Ecuaţia modelului este dată de:

$$\hat{x}_t = \mu + x_{t-1}$$

unde $\hat{x}_t$ reprezintă prognoza pentru seria temporală la momentul t şi μ este o constantă

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

$$DIFF(x) = \hat{x}_t - x_{t-1}$$

este variabila dependentă și singura variabila independentă fiind constantă μ

Să remarcăm că ecuaţia mersului la întâmplare fără creştere este cea de sus, dar fără constanta μ.

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

(differenced first-order autoregressive model).

În cazul în care erorile obţinute cu modelul mersului la întâmplare sunt autocorelate, se poate eventual rezolva problema prin adăugarea unei funcţii a cărei variabilă este diferenţa dintre valorile variabilei dependente, cu o perioadă în urmă. Obţinem următoarea ecuaţie de predicţie:

$$\hat{x}_t - x_{t-1} = \mu + \phi(x_{t-1} - x_{t-2})$$

ecuaţie ce poate fi scrisă

$$\hat{x}_t = \mu + x_{t-1} + \phi(x_{t-1} - x_{t-2}),$$

unde constanta este notată cu μ şi coeficientul autoregresiv cu ϕ .

ARIMA (0,1,1) fără constantă = *netezire* simplu exponenţială

ARIMA (0,1,1) fără constantă = *netezire* simplu exponenţială (modelul SES-simple exponential smoothing)

este un alt model utilizat pentru corectarea erorilor autocorelate din modelul mersului la întâmplare.

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

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

Ecuaţia de predicţie a modelului este:

$$\hat{x}_t = x_{t-1} - \theta \cdot \varepsilon_{t-1},$$

unde ε_{t-1} este eroarea la momentul , iar θ reprezintă coeficientul erorii de decalaj al prognozei şi corespunde acelui coeficientului $1-\alpha$ din formula netezirii exponenţiale.

ARIMA (0,1,1) cu constantă = *netezire* simplu exponenţială cu creştere

Implementând netezirea simplu exponențială ca un model ARIMA, câștigăm o oarecare flexibilitate deoarece:

- In primul rând este acceptat un coeficient negativ, corespunzător unui factor de netezire supraunitar, care nu este acceptat în procedura de netezire exponenţială.
- In al doilea rând aveţi posibilitatea de a introduce un factor constant în modelul ARIMA, cu scopul de a estima un trend ce nu are media nulă.

ARIMA (0,1,1) cu constantă are următoarea ecuaţie de prognoză:

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

Traiectoria prognozelor pe termen lung este o dreaptă de pantă μ, faţă de cazul netezirii exponenţiale când este o linie orizontală.

ARIMA mixt (1, 1, 1).

Acest model are ecuaţia de prognoză:

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

unde $\hat{x}_t$ reprezintă prognoza pentru seria temporală la momentul t, μ este o constantă, ϕ este coeficientul de regresie, ε_{t-1} este eroarea la momentul t, iar θ reprezintă coeficientul erorii de decalaj al prognozei

Se observă clar mixarea celor două componente (autoregresivă şi medie mobilă).
Uneori acest model poate duce la apariţia fenomenului de *overfitting* al datelor şi nu este exclus ca să nu obţinem coeficienţi unici.

11. Exemplu

Vom ilustra modelul de mai sus printr-o aplicaţie concretă privind un tratament clasic în bolile hepatice.
Considerămstudiul tratamentului cu interferon al hepatitei cronice *C* .

Astfel, vom fi interesaţi de prognoza evoluţiei stării unui pacient având hepatită cronică *C*, care a urmat un tratament standard de 6 luni cu interferon.

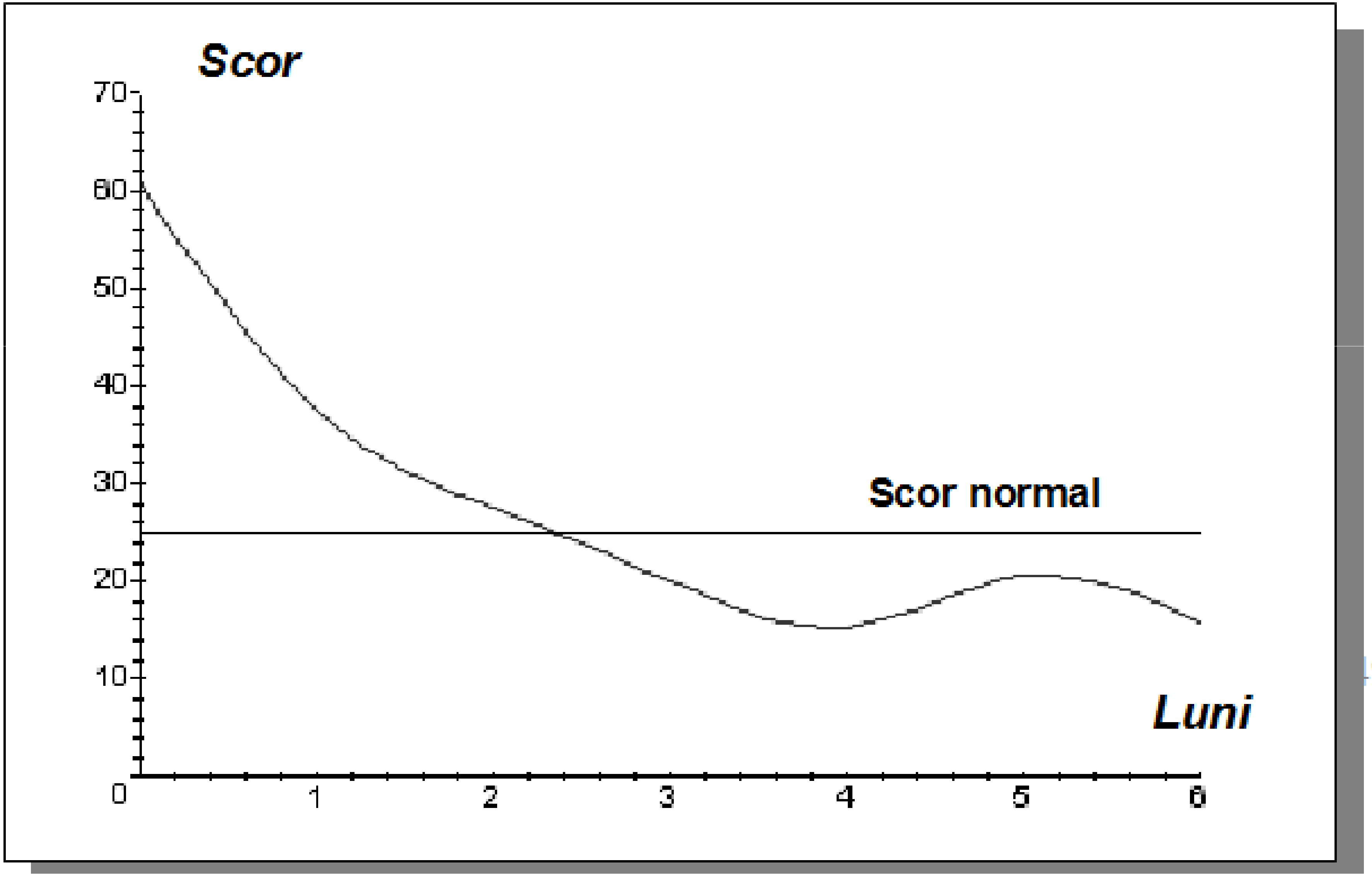
Importanţa prognozei rezidă în faptul că, pe de-o parte tratamentul cu interferon este foarte costisitor şi, pe de altă parte, o fracţiune deloc neglijabilă dintre pacienţi nu răspund pozitiv la acest tratament, continuarea acestuia nefiind eficientă din toate punctele de vedere.

Din această cauză este importantă crearea unui model predictiv, suficient de sigur, care să prevadă comportamentul clinic al unui pacient în următoarele 3 – 6 luni, pentru a se decide astfel dacă este vorba de un aşa-numit *responder*, adică un pacient cu răspuns pozitiv la tratament, sau un *non-responder*, adică un pacient cu răspuns negativ la tratament, pentru care continuarea tratamentului nu-şi mai are rostul.

Pentru modelarea comportamentului clinic al pacienţilor a fost ales clasicul model *mixt ARIMA* de tipul (1, 1, 1).

Deoarece procesarea computerizată a acestui model, ca de altfel şi gradul de încredere în prognoză, cere un număr relativ ridicat de date temporale (de la 20 în sus), a fost folosit procedeul de *interpolare B-spline cubică*, care induce curbe suficient de netede la noduri, fiind astfel compatibile cu procesul clinic asociat tratamentului. Pornind de la datele înregistrate în primele 6 luni, prin divizarea perioadei de timp respective şi interpolarea nodurilor obţinute, s-a creat baza de date suficientă pentru aplicarea modelului *ARIMA*.

Prezentăm mai jos graficul curbei de interpolare *B*-spline pentru scor, care reprezintă singura posibilitate de prognozare a evoluție stării de sănătate a unui pacient.



Plecând de la aceste date procesate prin interpolarea *B*-spline, prezentăm mai jos curba de prognoză *ARIMA* şi intervalul de încredere (95%) corespunzător pentru următoarele 6 luni (săptămânile 25-49), privind comportamentul clinic virtual al unui pacient de tip responder.