Serii temporale

Marina Gorunescu
mgorun@inf.ucv.ro

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.

In cazul seriilor temporale analiza se bazează pe presupunerea că valorile succesive din baza de date reprezintă măsurători consecutive, prelevate la intervale de timp egale (de aici şi denumirea de serii temporale).

Deoarece aceste măsurători temporale ilustrează dinamica unui anumit proces în timp, ele mai sunt cunoscute şi ca serii dinamice.

scopul 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.

analiza seriilor temporale

- identificarea prealabilă a pattern-ului (formei) seriei temporale a datelor și descrierea sa mai mult sau mai puțin formalizată.

-interpretarea pattern-ului seriei datelor analizate şi integrarea sa în anumite clase de alte pattern-uri de serii temporale deja cunoscute

- utilizarea pattern-ului sa în analiza fenomenului studiat.

- extrapolarea pattern-ului în vederea prognozei unor valorii viitoare, deci a unor evenimente ulterioare ce guvernează aceste valori.

exemplu

prognoza vremii care se face pe baza înțelegerii și extrapolării unor date meteo semnificative, prelevate până la data producerii prognozei, și utilizarea modelului (pattern-ului) astfel construit la prognoza vremii pentru perioada următoare.

O asemenea secvență temporală de date

(observații secvențiale), notată (ξ_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.

Din punct de vedere grafic, seriile temporale se vizualizează cel mai adesea prin tabele sau prin grafice

exemple

- 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.

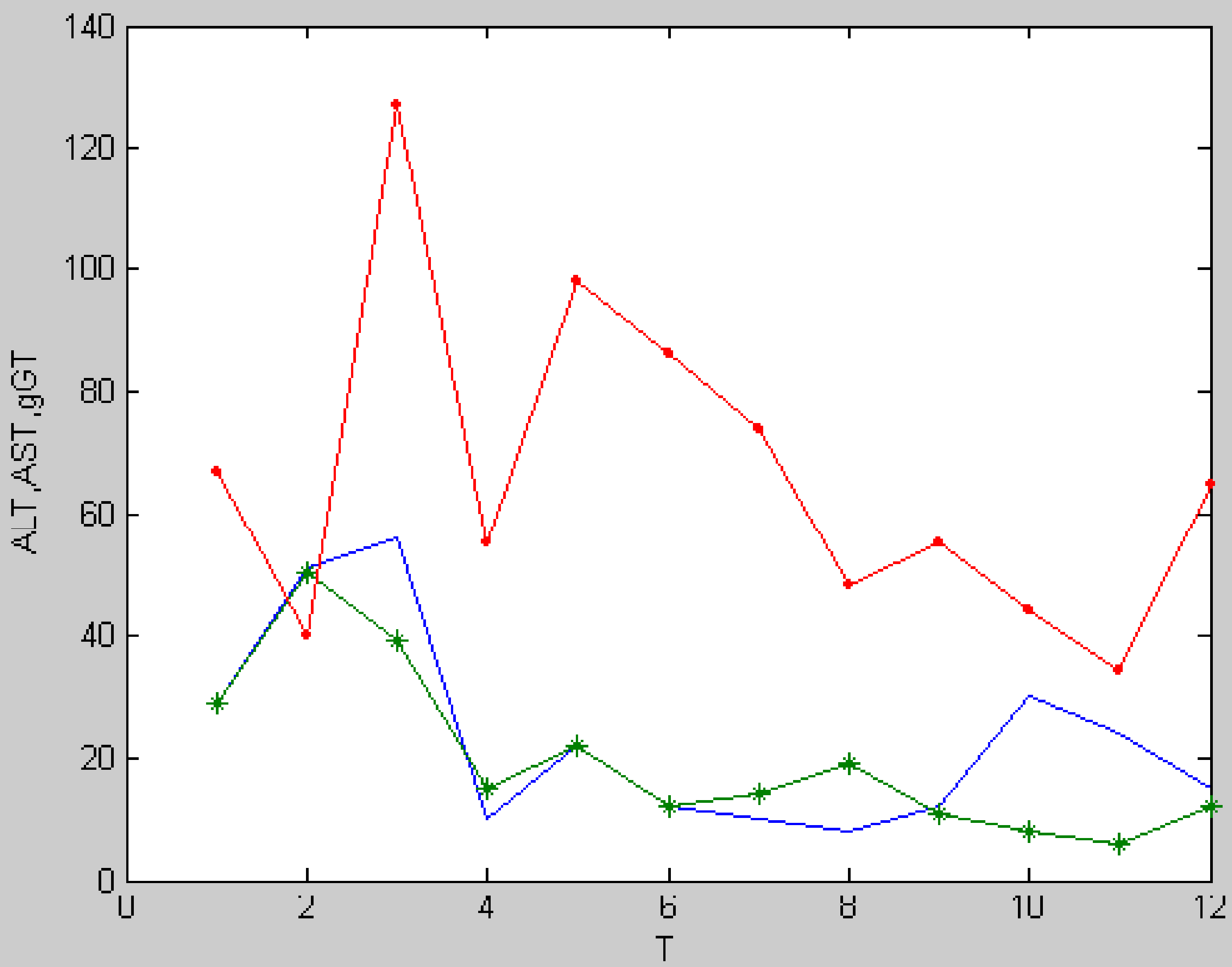
exemplu

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).*

	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,'.-')
```



ALT,AST,gGT
T
0
20
40
60
80
100
120
140
0
2
4
6
8
10
12

analiza unei serii temporale

- *Prognoza* se referă la evaluarea unor valori viitoare ξ_{T+h}, $h > 1$, pe baza datelor cunoscute $\xi_1, \xi_2, \ldots, \xi_T$.

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

- *Analiza orientării* (a *trend-ului*) se referă la faptul că atunci când studiem mai multe serii temporale corespunzătoare aceleaşi perioade de timp, este uneori necesar să analizăm motivele pentru care două sau mai multe asemenea serii au aceeaşi direcție (sunt puternic corelate), chiar dacă, la prima vedere, acest fapt pare inexplicabil.

- *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).

- *Distincția* între observațiile pe *timp scurt* și cele pe *timp îndelungat* se referă la separarea dintre relațiile persistente în timp, observate la datele culese și relațiile conjuncturale.

- *Relația de cauzalitate* se poate observa între două sau mai multe serii temporale.

  In cazul determinării unei relații de cauzalitate, se studiază factorul de defazare care intervine între cauză și 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.

1. 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 şi estimarea funcţiei f se fac plecând de la aşa-numitele *ajustări globale*, în care toate observaţiile se bucură de aceeaşi consideraţie, având roluri egale în estimaţii, sau aşa-numitele *ajustări locale*, în care fiecare observaţie îşi are rolul său în determinarea parametrilor modelului.

2. 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.

3. modele explicative

ecuația matematică capătă forma:

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

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.

In cazul seriilor temporale se pleacă de la ipoteza că datele analizate conțin o componentă definită de un pattern (formă) sistematic, adică o mulțime de elemente ce pot fi identificate fără dificultate, precum și o componentă de tip „zgomot” (*random noise*) sau „eroare”, care se suprapune peste pattern-ul definitoriu al seriei, făcând astfel dificilă identificarea sa cu precizie.

ce se poate face in aceste cazuri?

filtrarea (curățirea) zgomotului din date, astfel încât pattern-ul ce le guvernează să fie uşor identificabil şi apoi analizat.

După filtrarea zgomotului din date, este necesară analizarea şi descrierea pattern-ului definitoriu seriei temporale studiate.

Cele mai multe pattern-uri corespunzătoare seriilor temporale pot fi descrise prin două componente principale:

- *Tendința* (*trendul*), care se referă la o componentă sistematică, generală, liniară sau, de cele mai multe ori, neliniară, care se schimbă cu trecerea timpului și nu se repetă;

  de exemplu pentru cazul bidimensional, o curbă paralelă cu axa *Ox*, care apoi ia o turnură ascendentă, urmată iar de o porțiune paralelă cu axa *Ox*, după care apare o porțiune descendentă etc..

- *Periodicitate* (*seasonality*) reprezintă componenta care se repetă din timp în timp (sezonier), chiar dacă poate avea câteodată forma trendului (paralel cu *Ox*, crescător...). Cel mai bun exemplu este cel din domeniul vânzărilor sau internărilor în spital în preajma Sărbătorilor de Crăciun şi Paşte.

exemplu

(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.*

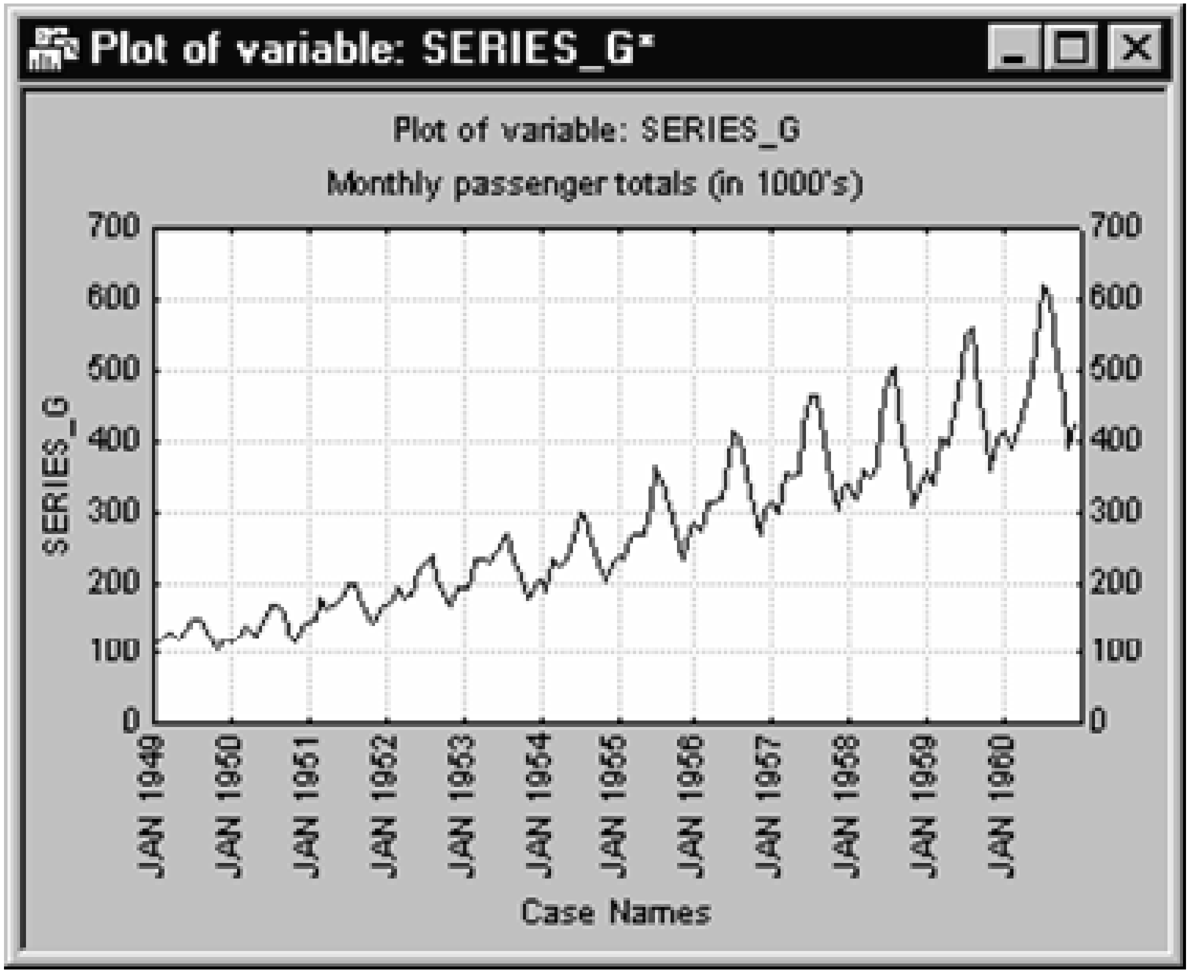
Baza de date corespunzătoare poate fi disponibilă sub forma următorului fișier (*Statistica*):

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



ce se observa din grafic?

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

câteva aspecte importante ale identificării pattern-urilor în seriile temporale.

analiza tendintei

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

Totuși, 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.

În cele mai multe cazuri concrete, prelevate din realitate, cazuri în care „zgomotul" este suficient de mare pentru a perturba identificarea tendinței, singura modalitate de lucru se bazează pe procesul de *ajustare* sau *netezire* (*smoothing*) a seriei temporale.

ajustarea

Ajustarea presupune un proces de mediere locală a datelor, cu scopul eliminării componentelor nesistematice care pot apărea în seria temporală dată.

Cea mai cunoscută tehnică de mediere este aşa-numită „media mobilă" (*moving average*) - tehnică de înlocuire a unei valori prin media simplă sau ponderată a *n* valori vecine acesteia.

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.

O serie *medie mobilă* poate fi calculată pentru orice serie temporală.

O *medie mobilă simplă* (*simple moving average* -SMA) este de fapt media (neponderată) a *n* puncte precedente.

De exemplu, pentru 10 valori precedente, SMA este dată de formula:

$$SMA = \frac{p_M + p_{M-1} + \ldots + p_{M-9}}{10}$$

unde p_{M-i} reprezintă valorile respective.

Când se calculează medii mobile succesive, valoarea nouă va fi introdusă în formulă, renunțându-se în același timp la valoarea cea veche.

exemplu

dacă se mediază prețurile zilnice, avem formula:

$$SMA_{azi} = SMA_{ieri} - \frac{p_{M-n+1}}{n} + \frac{p_{M+1}}{n}$$

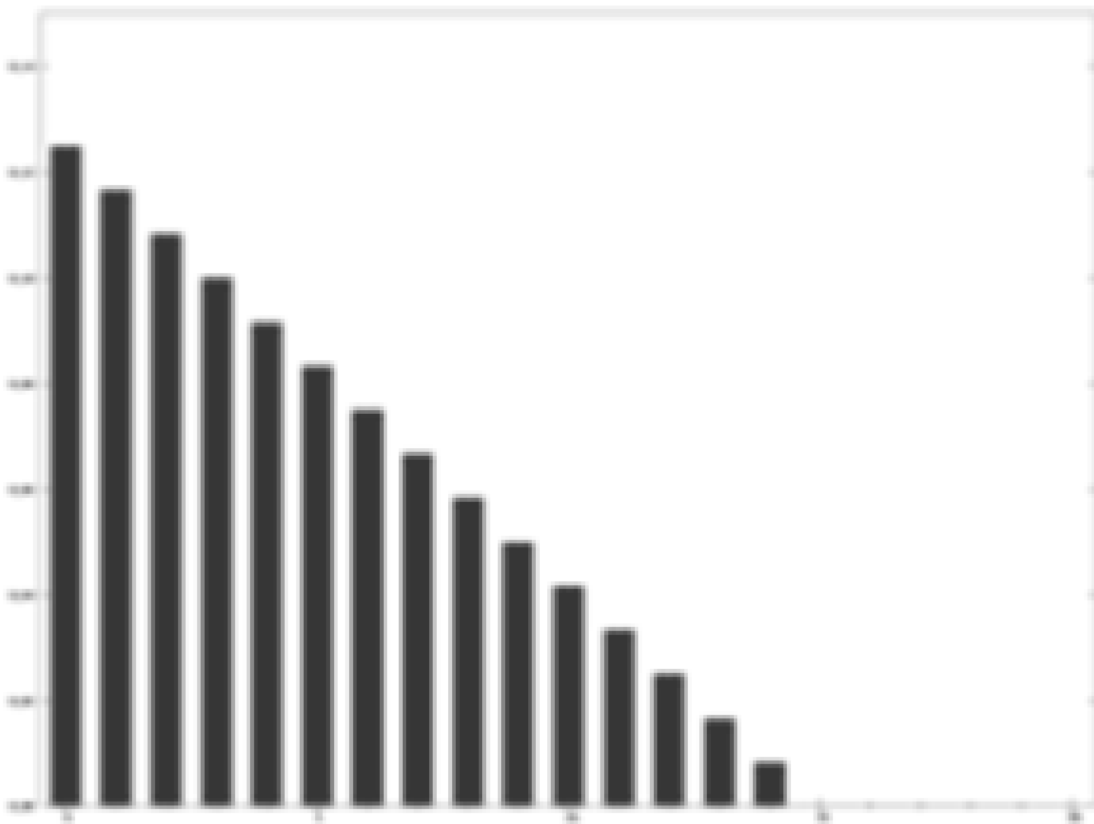
De menționat că perioada pentru care se face medierea,
Adică valoarea lui *n* depinde de context;
există medieri scurte, medii și lungi de exemplu
n = 10, *n* = 50, *n* = 200 etc.

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

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

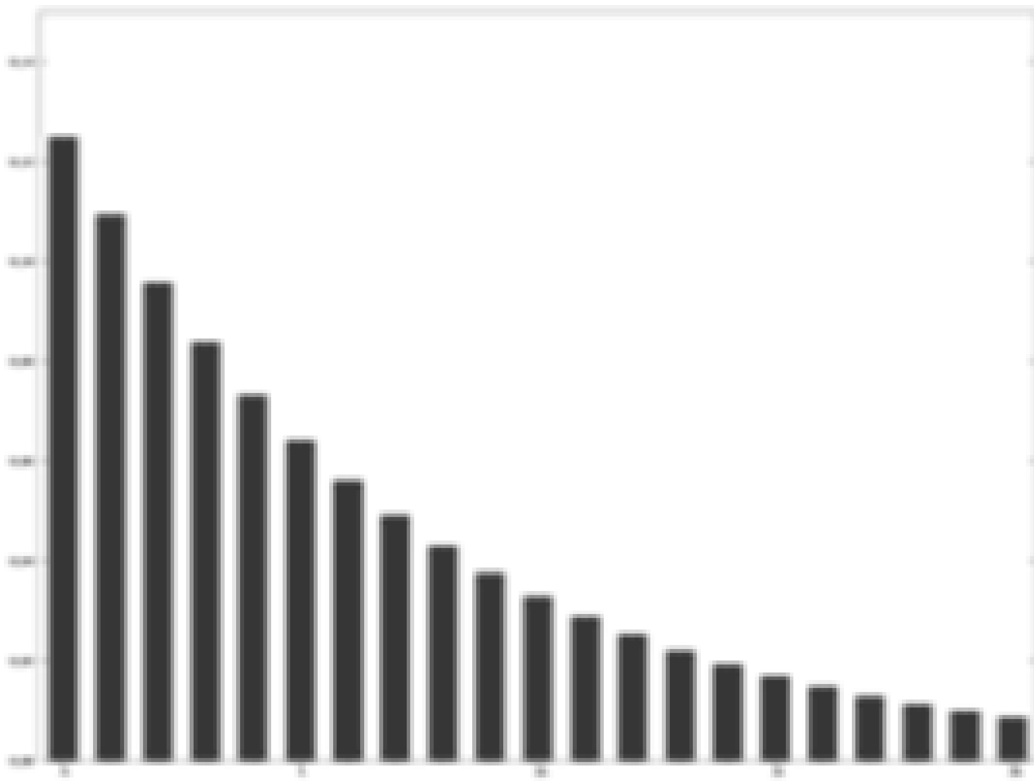
Se observă că valorile ponderilor pentru fiecare dată din serie descreşte odată cu timpul, începând cu n şi terminând cu 0 pentru p_{M-n}.

Ilustrăm descreşterea ponderilor de la cea mai recentă dată până la cea mai veche, $n = 15$.



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 și nu liniar (astfel, datele cele mai recente sunt și cele mai importante).

graficul unei descresteri exponentiale a ponderilor



Adaptarea (*ajustarea*) *unei funcții* reprezintă o altă abordare privind „netezirea” datelor.

Astfel, în cazul în care se analizează o serie temporală care este monotonă, aceasta poate fi, de multe ori, aproximată printr-o funcție liniară de timp.

Se pot utiliza aproximări ale serie temporale prin alte tipuri clasice de funcții ca, de exemplu, funcții polinomiale, exponențiale, logaritmice etc.

2. Analiza periodicitatii

O altă caracteristică a seriilor temporale reflectă dependențele periodice, sezoniere.
În exemplul referitor la numărul pasagerilor pentru cursele aeriene internaționale, prezentat anterior, se observă că, înafara tendinței crescătoare, în fiecare an pattern-ul numărului călătorilor are forma „Λ", cu o creştere repetabilă în perioada lunilor de vară (maxim).

În cazul unei serii temporale care nu prezintă „zgomot” excesiv, această periodicitate poate fi uşor identificată ca o formă repetabilă după fiecare *k* unităţi temporale, de exemplu 12 luni în cazul curselor aeriene.

Din punct de vedere tehnic, computațional, pattern-urile periodice pot fi puse în evidență cu ajutorul așa-numitelor *corelograme* (sau *autocorelograme - autocorrelograms*). Concret, printr-o corelogramă reprezentăm grafic *funcția autocorelație* (autocorrelation function -ACF).

Pe scurt, fiind dată o serie temporală ξ_t, asociată unui proces stochastic notat, pentru simplitate, $\{\xi_t, t > 0\}$, de medie μ şi dispersie σ^2, atunci ACF este definită de:

$$R(t,s) = \frac{E[(\xi_t - \mu)(\xi_s - \mu)]}{\sigma^2}$$

Cu alte cuvinte, autocorelația, aşa cum arată şi numele, reprezintă corelația unei serii temporale cu ea însăşi, translatată cu o anumită defazare (t - s).

De reținut că atunci când este examinată o corelogramă, trebuie avută în vedere ipoteza că autocorelațiile pentru defazări consecutive (t - s) sunt formal dependente, ilustrând un fel de tranzitivitate a relației între elemente.

De exemplu, dacă primul element al seriei este corelat cu al doilea, iar acesta cu al treilea, este normal să presupunem că al treilea va fi corelat cu primul.

Utilizarea sa în identificarea dependențelor periodice în cazul seriilor temporale se bazează tocmai pe ideea de măsură a corelației seriei cu ea însăși, dar după o perioadă (constantă) de timp.

De menționat că funcția *R* ia valori în intervalul [-1, 1], valoarea 1 indicând o corelație maximă, în timp ce -1 ilustrează o corelație minimă.

În acelaşi context, pentru a îndepărta pe cât este posibil dependența periodică (sezonieră), se procedează la o transformare neperiodică diferențiată (*differencing transformation*), a seriei originale, transformare dată de ecuația:

$$X = X - X(\text{defazaj}),$$

unde defazajul (*lag*) are o anumită valoare $k \geq 1$, ce poată fi aleasă în funcție de situația concretă.

După această transformare, seria obținută va avea lungimea (N - defazaj), unde N era lungimea sa inițială.

ARIMA

Model des folosit în domeniul prognozei ce şi-a dovedit eficienţa în cazurile concrete, extrase din realitate.

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.

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 ϕ

- 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:

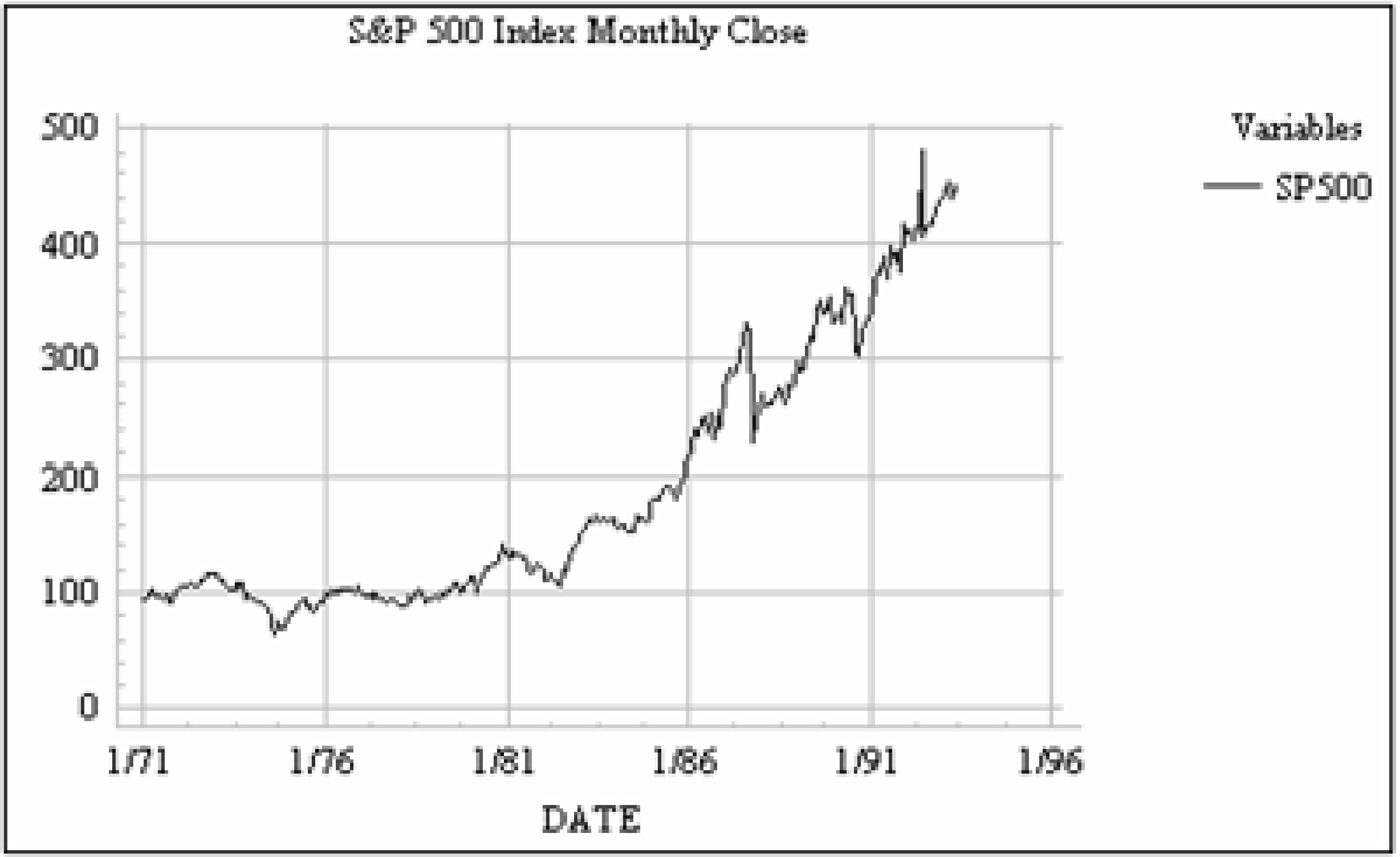
$$x_t = \xi + x_{(t-1)}$$

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

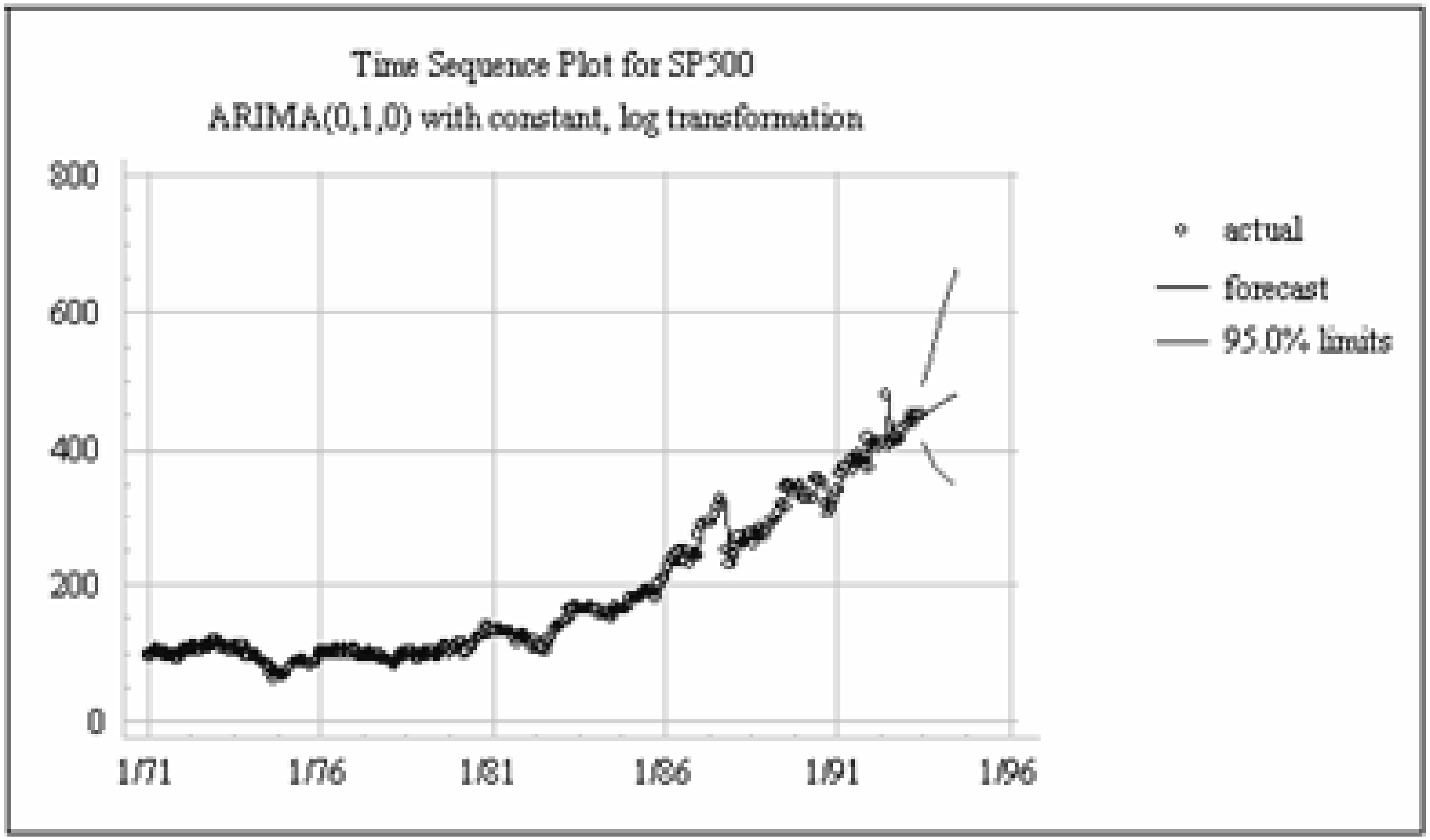
$$DIFF(x) = x_t - x_{(t-1)}$$

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 prezentatăin 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:

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

ecuație ce poate fi scrisă

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

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 $\varepsilon_{(t-1)}$ este eroarea la momentul $t-1$

ARIMA mixt (1,1,1)

Acest model are ecuația de prognoză:

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

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

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

- influenţează în vreun fel (şi dacă da, cum?) o nouă politică fiscală importurile?
- o nouă abordare a Ministerului Sănătăţii privind lista medicamentelor compensate influenţează şi 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*.

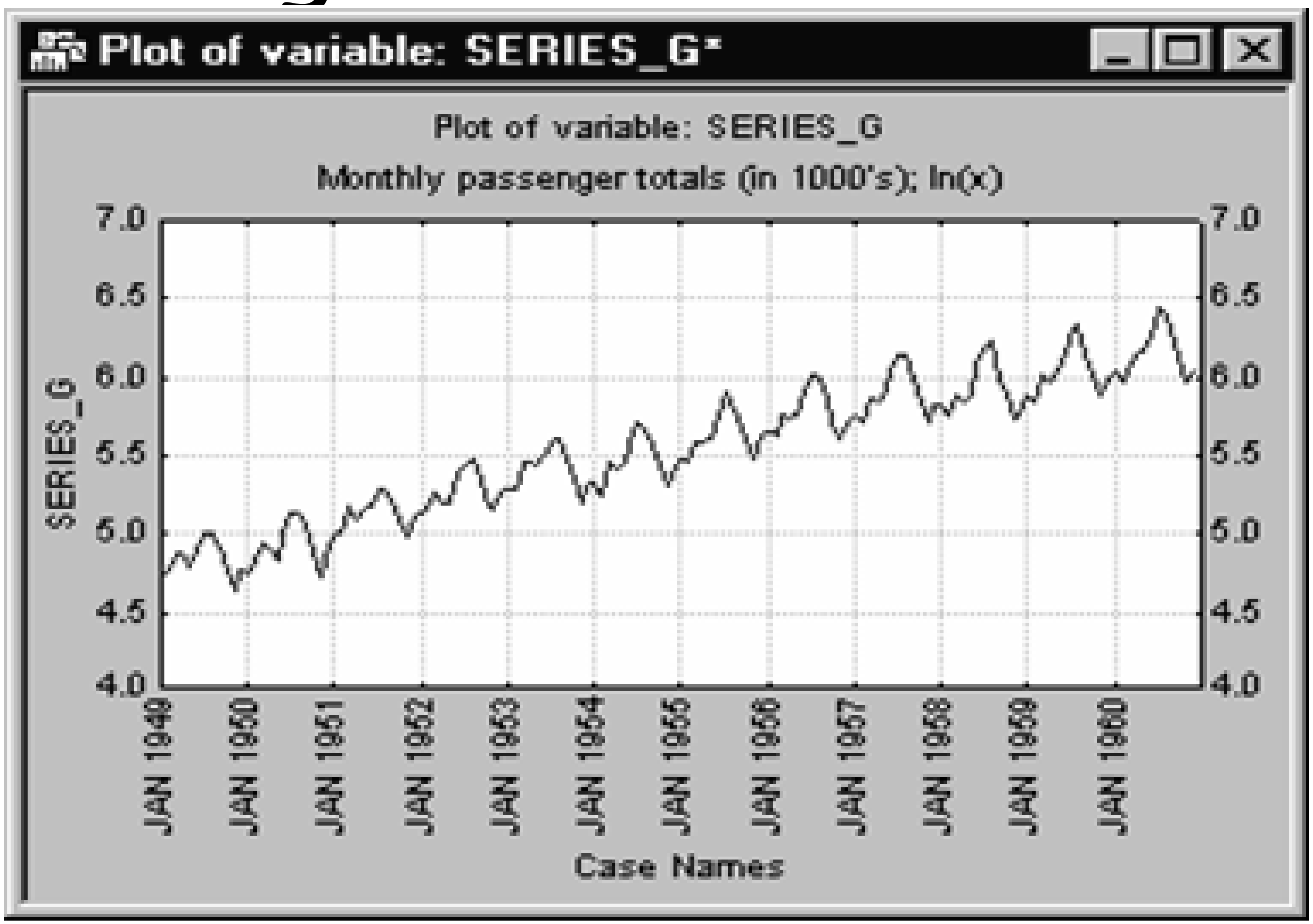
Există în această direcție modele de prognoză, dintre acestea menționând *ARIMA* pentru serii temporale întrerupte

aplicatii

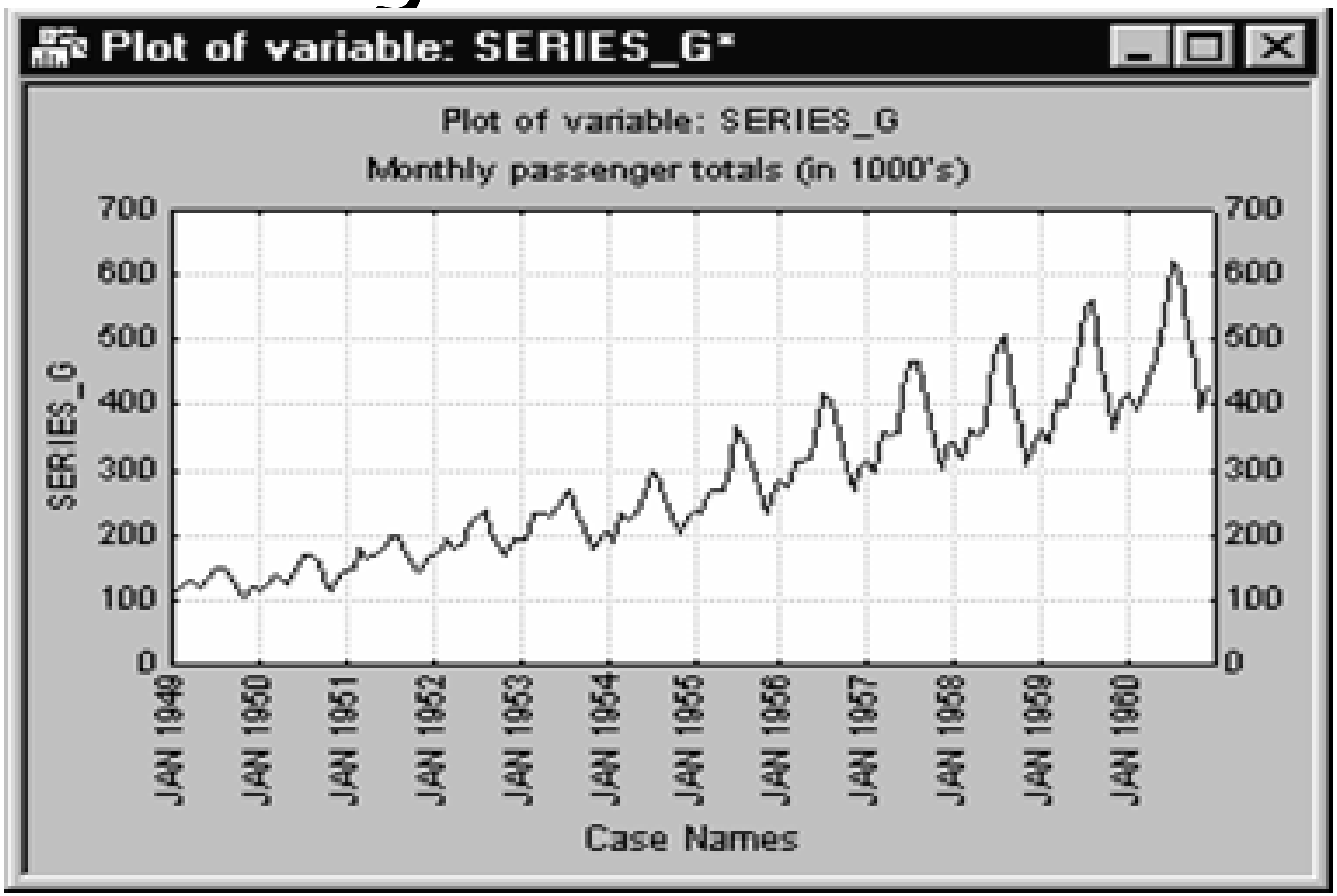
Reluăm cazul referitor la seria temporală reprezentând numărul de pasageri ai curselor aeriene internaționale, înregistrat în perioada 1949-1960

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

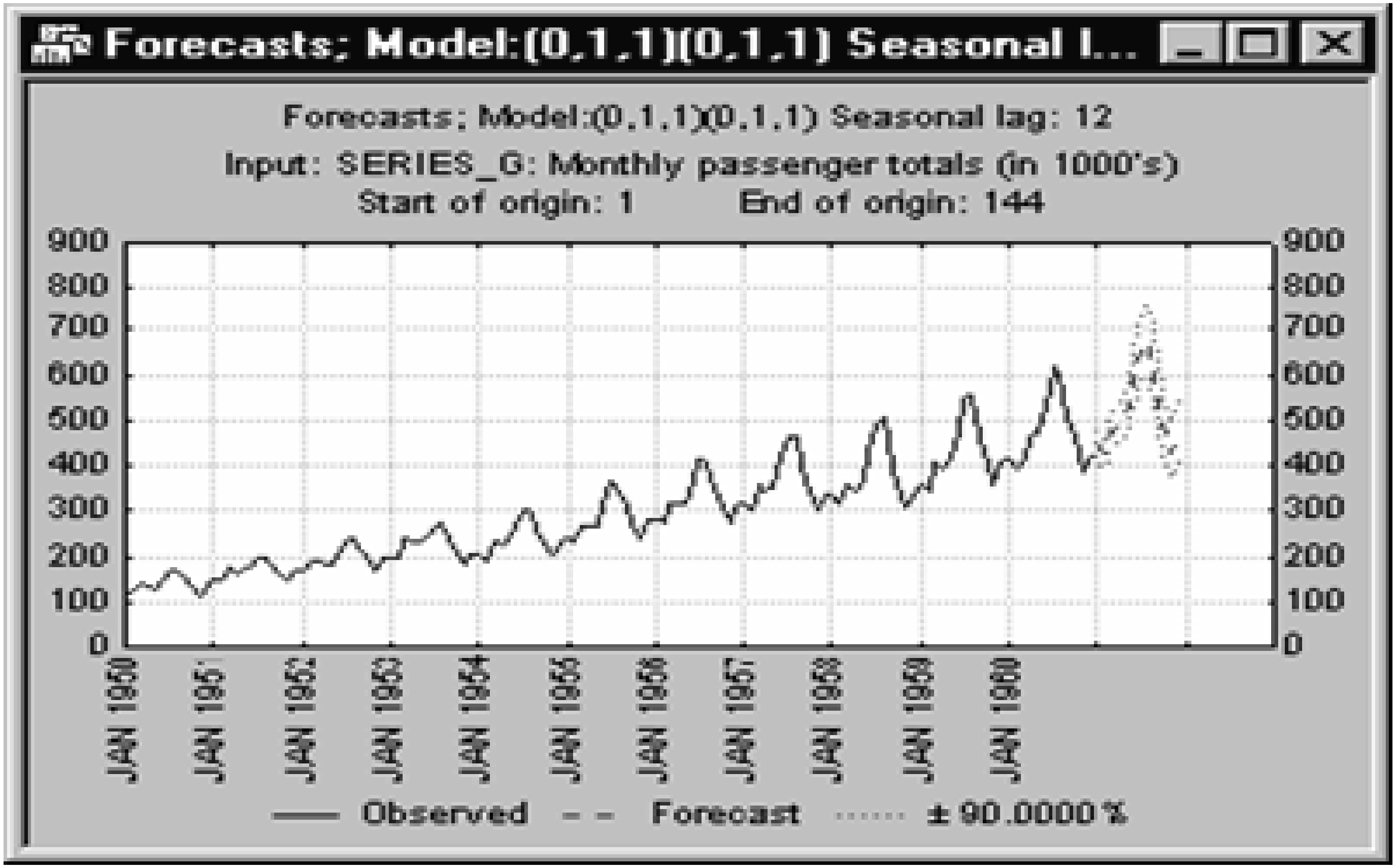


seria originara

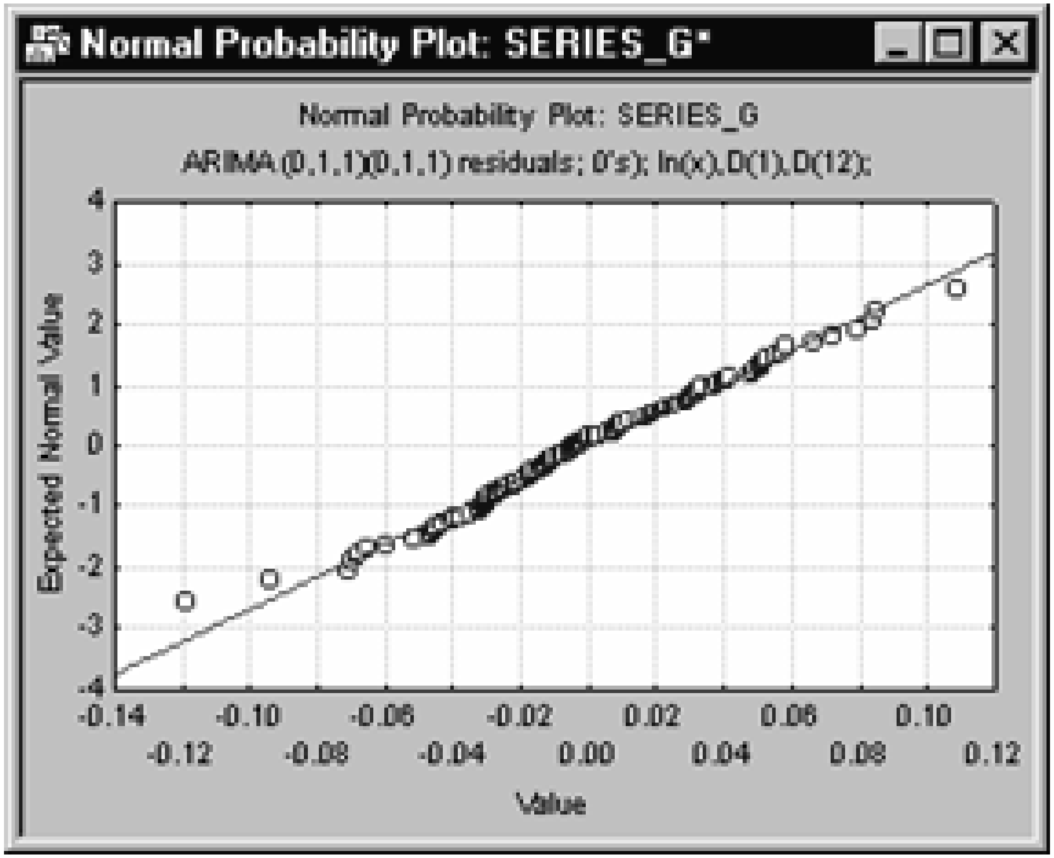


Comparând graficele,al seriei temporale originare cu cel al seriei logaritmate, se observă fenomenul de stabilizare al ei, în sensul că la seria logaritmată amplitudinea creşterilor este stabilă.

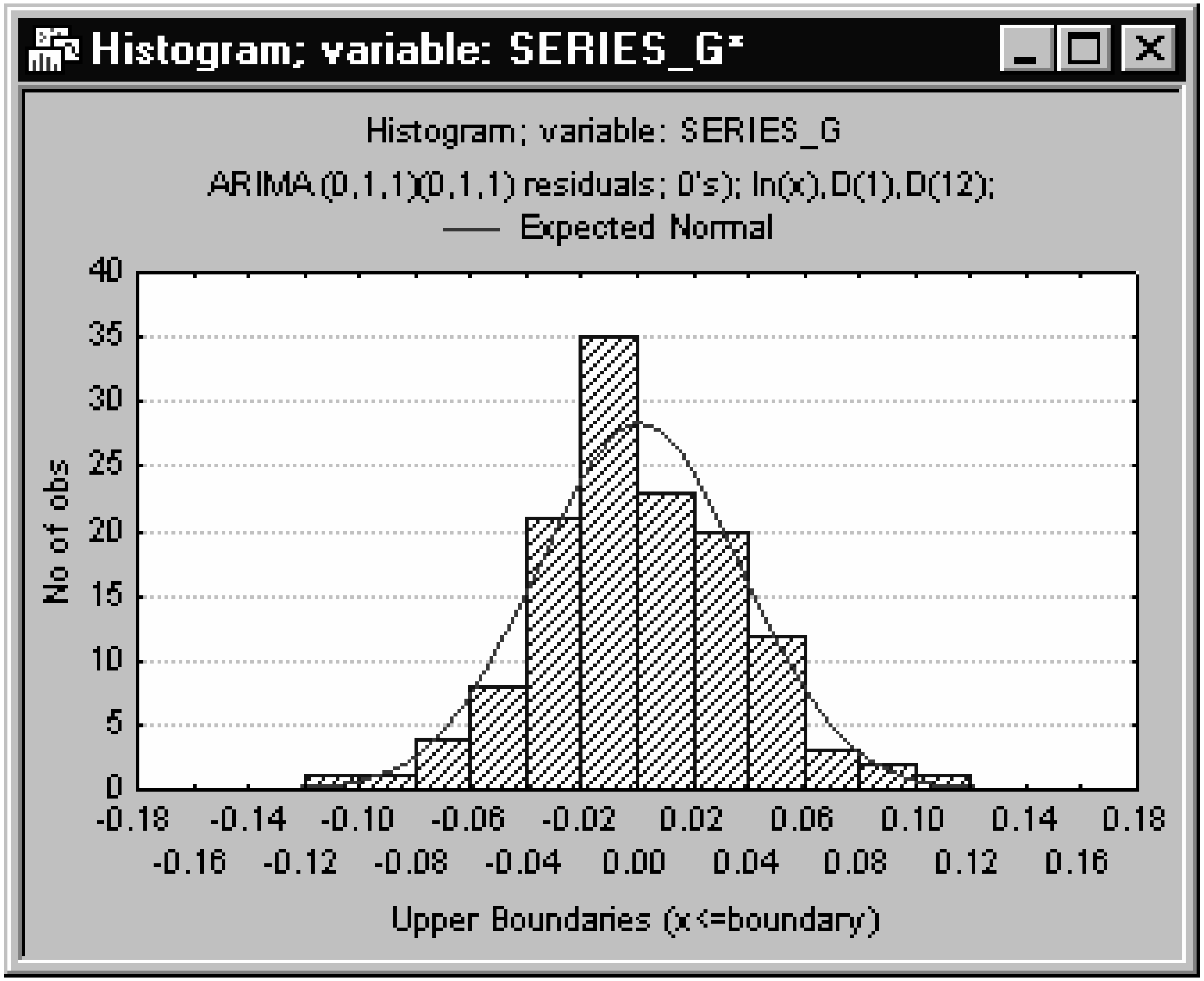
În vederea prognozei am aplicat modelul ARIMA (0, 1, 1) obținând secvența prognozată pentru perioada următoare, reprezentată în figură


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
900
800
700
600
500
400
300
200
100
0
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.
Aşa cum era de aşteptat, 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

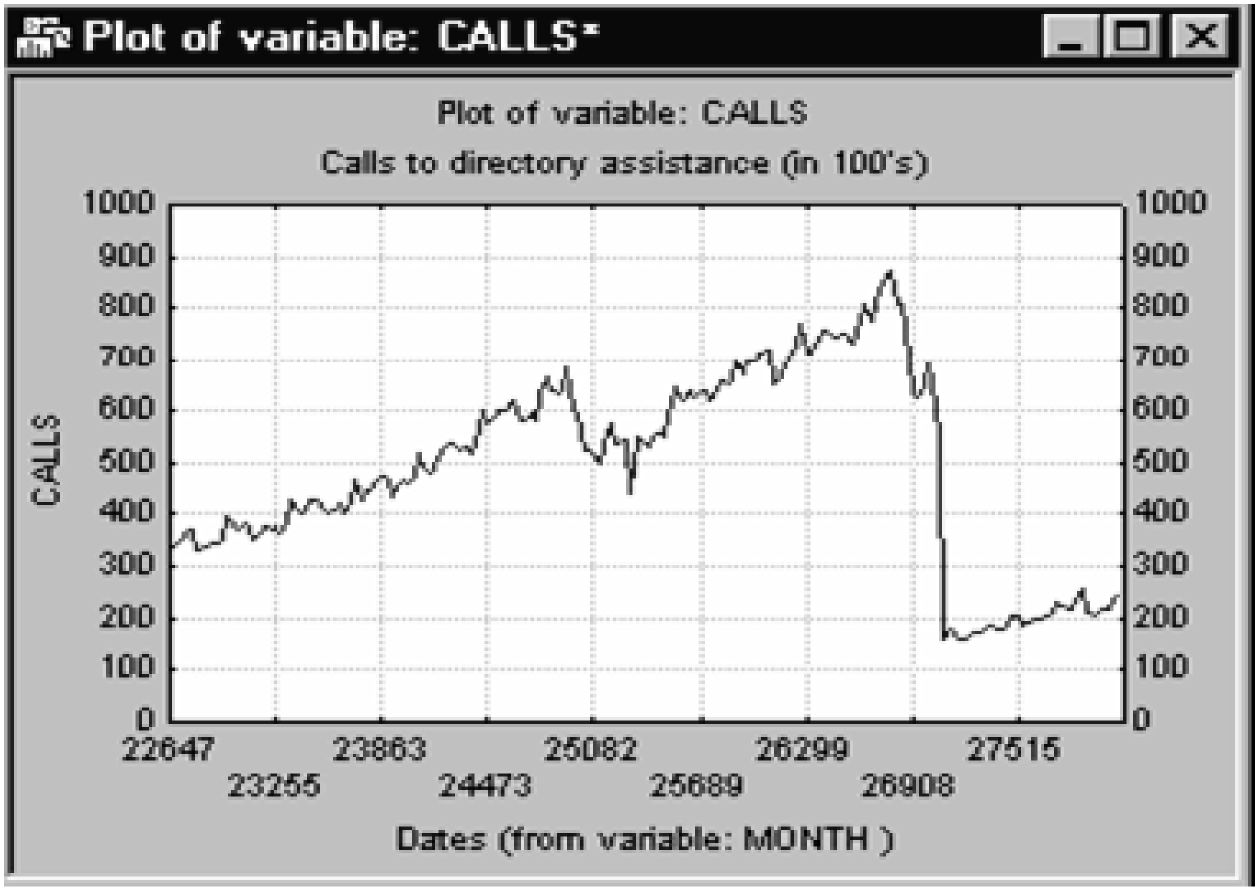
În cele două tabele sunt prezentate cele două eşantioane aparţinând celor două subserii (până la cazul 146 şi după acesta).

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

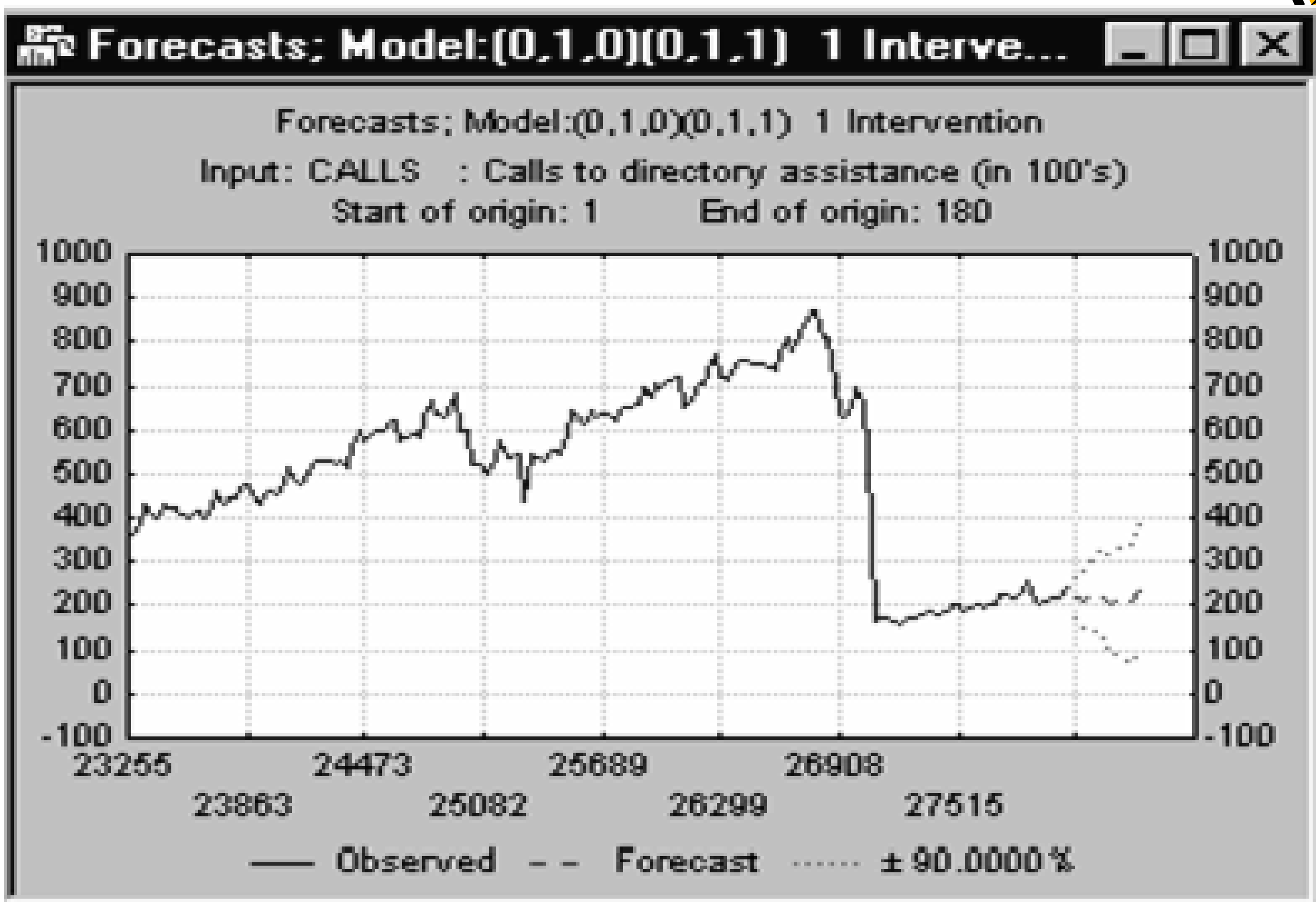


în cadrul modelului ARIMA pentru serii întrerupte există trei abordări:

- 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) și supratazarea care a produs-o a fost permanentă.

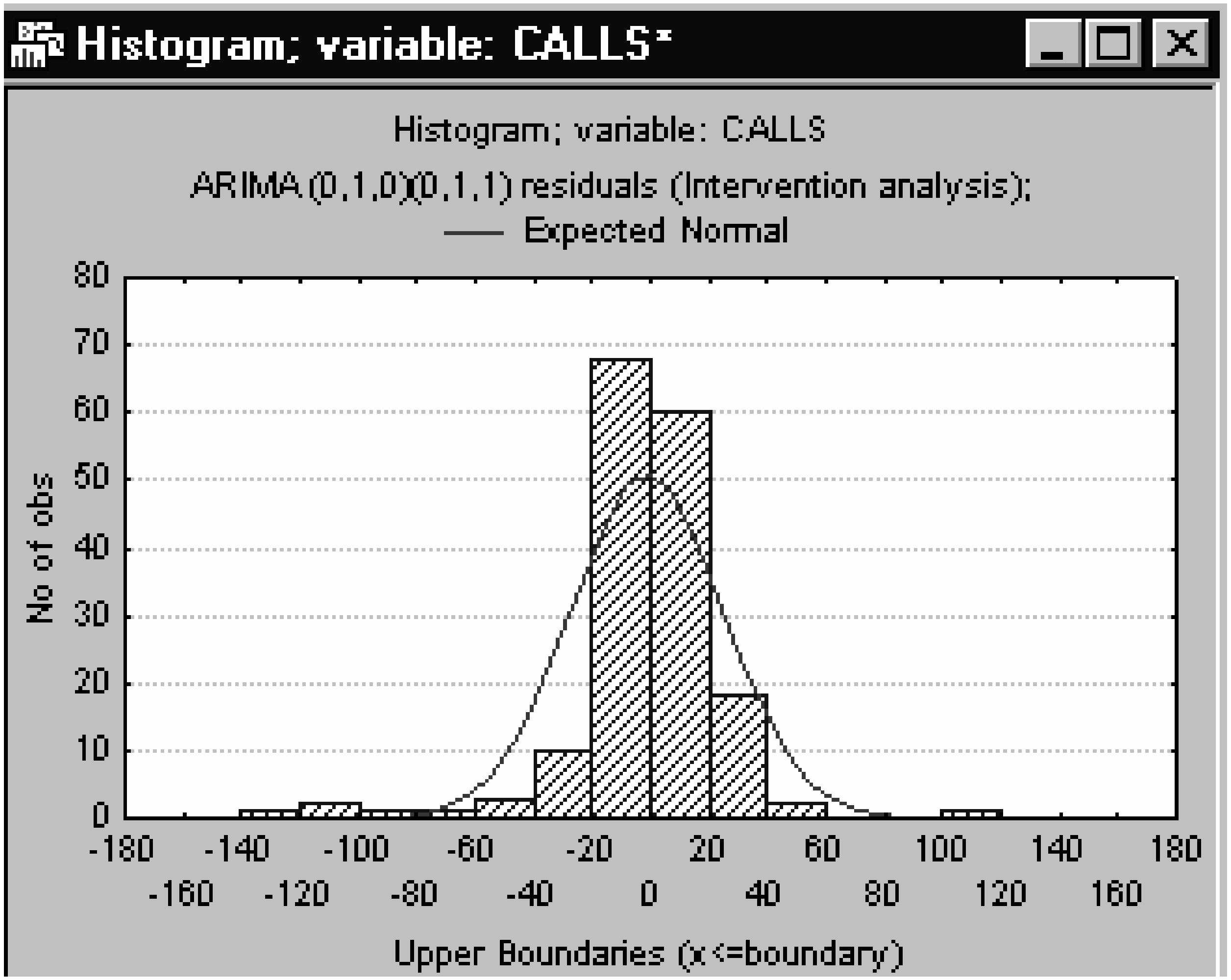
prognoza (un an) obtinuta aplicand modelul ARIMA



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
0
10
20
30
40
50
60
70
80
-180
-160
-140
-120
-100
-80
-60
-40
-20
0
20
40
60
80
100
120
140
160
180
Upper Boundaries (x<=boundary)

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