



(12) CERERE DE BREVET DE INVENȚIE

(21) Nr. cerere: a 2018 00473

(22) Data de depozit: 28/06/2018

(41) Data publicării cererii:
30/01/2020 BOPi nr. 1/2020

(71) Solicitant:
• INSTITUTUL NAȚIONAL DE
CERCETARE-DEZVOLTARE ÎN
INFORMATICĂ, BD.MAREȘAL AVERESCU
NR.8-10, SECTOR 1, BUCUREȘTI, B, RO

(72) Inventatori:
• POPESCU THEODOR DAN,
ȘOS.COLENTINA NR.55, BL.83, SC.1, ET.6,
AP.29, SECTOR 2, BUCUREȘTI, B, RO;

• ALEXANDRU ADRIANA, STR.BELGRAD,
NR.6, ET.3, AP.7, SECTOR 1, BUCUREȘTI,
B, RO;
• IANCULESCU MARILENA,
STR.CPT.AVIATOR NICULAE DOSSU,
NR.16, ET.2, AP.5, SECTOR 1,
BUCUREȘTI, B, RO;
• TUDORĂ ELEONORA, STR.SIBIU, NR.4,
BL.OD2, SC.1, ET.6, AP.28, SECTOR 6,
BUCUREȘTI, B, RO

Această publicație include și modificările descrierii,
revendicărilor și desenelor depuse conform art. 35
alin. (20) din HG nr. 547/2008.

(54) PROCEDURA PENTRU PREDICȚIA SERIILOR DE TIMP
MULTIVARIABILE UTILIZÂND SEPARAREA OARBĂ
A SURSELOR INDEPENDENTE

(57) Rezumat:

Invenția se referă la un procedeu de predicție a seriilor de timp multivariabile, cu aplicabilitate în domeniile ingineriei, științelor fizice, biologiei, medicinei, sociologiei, hidrologiei, geofizicii, economiei care se confruntă cu analiza unor date de măsură și observație. Procedeul conform invenției realizează transferul problemei, care face obiectul analizei, din spațiul datelor originale în spațiul surselor independente, de dimensiune mult redusă, și unde se dispune de un arsenal bogat de metode și tehnici de modelare și predicție a seriilor de timp monovariabile, rezultatele predicției surselor independente fiind transferate apoi în spațiul original al seriei de timp multivariabile, utilizând modelul de mixare al surselor independente, rezultat în urma separării "oarbe" a acestora.

Revendicări inițiale: 4
Revendicări amendate: 1
Figuri: 11

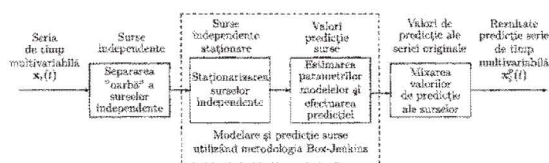
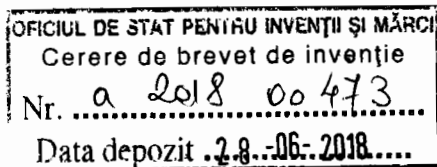


Fig. 1



54



PROCEDURĂ PENTRU PREDICȚIA SERIILOR DE TIMP MULTIVARIABILE UTILIZÂND SEPARAREA "OARBĂ" A SURSELOR INDEPENDENTE

DESCRIEREA INVENȚIEI

Titlul invenției

Procedură pentru predicția seriilor de timp multivariabile utilizând separarea "oarbă" a surselor independente.

Domeniul tehnic la care se referă invenția

Invenția se referă la o procedură neconvențională de predicție a seriilor de timp multivariabile, ce face uz de separarea "oarbă" a surselor independente.

Stadiul tehnicii în domeniu

Specialiștii din domeniul ingineriei, științelor fizice, biologiei, medicinei, sociologiei, hidrologiei, geofizicii, economiei se confruntă cu analiza unor date de măsură și de observație. Astfel de secvențe de date, sau secvențe de măsurători, efectuate la momente de timp ordonate (de obicei, la intervale egale de timp), sunt cunoscute ca serii de timp, sau serii dinamice, [1].

Scopul analizei cantitative a acestei categorii de date, în cele mai multe dintre cazuri, este caracterizarea succintă, dar de calitate a sistemului care a generat aceste date, printr-un model matematic. Acest model poate fi utilizat pentru analiza sistemului și predicția evoluției sale viitoare, în cazul apariției unor schimbări în mediul în care acesta operează. Informația obținută în urma unei astfel de analize poate fi utilizată pentru a modifica anumiți factori și variabile din sistem, în scopul atingerii unor performanțe optime într-un anumit sens, [2].

Domeniul analizei, modelării și predicției seriilor timp se bucură de o atenție deosebită pe plan internațional, iar cercetarea teoretică și aplicațiile practice reprezintă o provocare pentru viitor. Părerea generală a comunității științifice, cu preocupări în domeniu, este că există o discrepanță între rezultatele teoretice și practice. Aceasta implică eforturi susținute pentru elaborarea de noi metode și tehnici capabile să depășească aceste probleme, ținând seama de posibilitățile largi de utilizare a acestora, în diferite domenii tehnice și non-tehnice. Dezvoltarea acestui domeniu s-a datorat în mare măsură și interacțiunii puternice dintre domeniile prelucrării semnalelor și recunoașterii formelor, teoriei controlului, statisticii aplicate și controlului de calitate.

Metodele statistice aplicate datelor seriilor de timp au fost utilizate inițial, în principal, în economie și apoi în multe alte domenii, cum ar fi ecologia, fizica, medicina și ingineria. În

primele aplicații, accentul era pus pe problemele de predicție, scopul fiind acela de a se realiza o prognoza cât mai exactă a valorilor viitoare ale seriei ce face obiectul analizei.

Utilizarea metodelor de analiză și predicție a seriilor de timp în cazul seriilor multivariabile, unde problemele sunt mult mai complexe decât în cazul seriilor monovariabile, s-a bucurat de un interes deosebit în ultimii ani. O trecere în revistă a ultimilor 25 de ani de cercetare în acest domeniu a fost publicată de J.G.De Gooijer, R.J. Hyndman, "25 years of time series forecasting", *International Journal of Forecasting*, 2006, [3]. Pe durata acestei perioade se poate constata un progres remarcabil, dar așa cum autorii acestui articol remarcă, există încă un mare număr de subiecte și probleme ce necesită a fi abordate și rezolvate. Mai multe probleme practice privind analiza, modelarea și predicția seriilor de timp sunt prezentate sub forma a 139 de principii în lucrarea lui J.Scott Armstrong, "Standards and practices for forecasting", din *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Kluwer Academic Publishers, 2001, [4].

Aceste lucrări furnizează informații privind formularea problemei de predicție, rezolvarea unei astfel de probleme, selectarea și aplicarea metodelor de predicție, evaluarea acestora și folosirea rezultatelor predicției. Fiecare din principiile enumerate este descris împreună cu scopul său, precum și cu condițiile în care acesta este relevant, în scopul asistării celui interesat în realizarea unei predicții sau prognoze.

În prezent, în literatură sunt raportate abordări ce folosesc diferite tipuri de modele stohastice, ce vor fi menționate succint în continuare. Modelele vectoriale integrate autoregresive și de medie alunecătoare (VARIMA) se utilizează frecvent în predicția seriilor de timp multivariabile [5], și reprezintă o generalizare a modelelor stohastice monovariabile integrate autoregresive și de medie alunecătoare (ARIMA). Deoarece modelele VARIMA acceptă ipoteze privind prezența variabilelor exogene și interacțiunea dintre variabile, acestea oferă noi soluții pentru realizarea de prognoze și luarea deciziilor.

Modele vectoriale autoregresive (VAR), [5], reprezintă un caz special al clasei mai generale de modele VARMA. În esență, un model VAR este o aproximare flexibilă a unei varietăți largi de modele econometrice dinamice. Modelele VAR pot fi specificate în mai multe moduri. În general, aceste modele pot fi afectate de o "supraparametrizare", ceea ce implică prezența în model a unui număr relativ mare de parametri nesemnificativi. În consecință, aceste modele pot să furnizeze prognoze nesatisfăcătoare pentru date din afara eșantionului utilizat pentru determinarea modelului.

Modelele autoregresive vectoriale Bayesiene (BVAR), [6], au fost utilizate pentru prognoze macroeconomice, [7], prognoza pieței muncii, [8], prognoze de afaceri, [9], sau prognoze economice locale, [10], etc.

La începutul anilor '80, modelele în spațiul stărilor au început să fie utilizate de statisticieni pentru predicția seriilor de timp, aceste idei fiind deja frecvent utilizate în literatura specifică ingineriei, [11]. Modelele în spațiul stărilor furnizează un cadrul unitar în care poate fi reprezentat orice model liniar al unei serii de timp. Contribuția esențială în domeniul predicției, datorată lui Kalman, a constat în furnizarea unui algoritm recursiv (cunoscut ca filtrul Kalman) pentru calculul valorilor de predicție. Filtrul Kalman furnizează un algoritm eficient pentru calculul erorilor de predicție pe un pas și a valorilor de dispersie asociate acestora, necesare evaluării funcției de verosimilitate. Combinarea algoritmului de maximizare a valorii așteptate (EM) cu filtrul Kalman, [12], furnizează o procedură generală de predicție a seriilor de timp folosind modele în spațiul stărilor, inclusiv în cazul unor observații lipsă.

Problema tehnică rezolvată de invenție

Problema tehnică rezolvată în cadrul invenției se referă la o procedură neconvențională de predicție a seriilor de timp multivariabile, bazată pe separarea "oarbă" a surselor independente.

Analiza directă a seriilor de timp multivariabile, ridică multe probleme în practică, în special datorită dificultății reprezentării canonice a acestora, fapt pentru care se preferă transferul problemei, care face obiectul analizei, din spațiul datelor originale în spațiul surselor independente, de dimensiune mult redusă și unde se dispune de un arsenal bogat de metode și tehnici de modelare și predicție a seriilor de timp monovariabile, dată fiind independența statistică a surselor independente, care au generat seria de timp multivariabilă. În final, rezultatele analizei pot fi transferate în spațiul original al măsurătorilor, ținând seama că tehnicile de separare a surselor independente furnizează și modelul de mixare al datelor sursă, de tip instantaneu sau convolutiv.

Așa cum s-a menționat, abordarea propusă face uz de separarea "oarbă" a surselor independente (Blind Source Separation - BSS), [13], ca un instrument de preprelucrare a datelor seriei multivariabile, înaintea efectuării predicției propriu-zise. Această tehnică de calcul statistic este strâns legată de Analiza Componentelor Independente (ICA), poate cel mai frecvent utilizată pentru separarea "oarbă" a surselor independente, și care poate fi privită ca o extensie a Analizei Componentelor Principale (PCA) și Analizei Factoriale (FA). ICA este cea mai puternică tehnică, capabilă a determina factorii sau sursele independente care stau la baza generării datelor ce fac obiectul analizei și predicției.

Procedura ce face obiectul invenției se prezintă în Figura 1 și implică următoarele etape:

1. Eliminarea valorilor medii ale componentelor seriei de timp multivariabile (centrarea datelor seriei).
2. Estimarea surselor independente $s_j(t)$ și a matricei de mixare a acestora \mathbf{A} ; procedura este reprezentată în Figura 2.
3. Obținerea staționarității componentelor independente, în special în valoarea medie și în dispersie, cazurile cel mai frecvent întâlnite în practică. În acest scop, pentru fiecare componentă independentă $s_j(t)$, se realizează, funcție de specificul acesteia, estimarea și eliminarea tendinței, diferențierea nesezonieră sau sezonieră de diferite ordine, etc. Aceste tipuri de transformări urmează a fi reținute pentru a fi aplicate valorilor de predicție ale acestor componentelor, și obținerea în final a valorilor de predicție ale seriei originale.
4. Fiecare componentă rezultată este modelată, de regulă utilizând modele parametriche de tip autoregresiv (AR), de medie alunecătoare (MA), autoregresive și de medie alunecătoare (ARMA), utilizând metodologia Box-Jenkins, [1]; în această etapă pot fi utilizate toate metodele și tehnicile de modelare a seriilor de timp monovariabile. Modelele rezultate urmează a fi validate.
5. Utilizând modelele rezultate, fiecare sursă independentă este predictată în mod separat, pentru un anumit orizont de predicție. Tot în această etapă, se aplică acestor predicții transformările inverse utilizate la obținerea staționarității surselor independente, în Etapa 2.
6. Valorile de predicție rezultate și limitele de încredere sunt mixate utilizând coeficienții de mixare, a_{ij} , ai matricei \mathbf{A} , pentru a obține valorile de predicție, $x_i^p(t)$ ale componentelor seriei de timp multivariabile originale $x_i(t)$.
7. Corectarea rezultatelor de predicție cu valorile medii ale componentelor seriei inițiale, estimate în Etapa 1.

Expunerea invenției

Principalele etape ale procedurii ce face obiectul invenției, sunt reprezentate în Figura 1, și se prezintă în cele ce urmează.

Borderou figuri

Figura 1 Procedura de predicție a seriilor de timp multivariabile

Figura 2 Mixarea și separarea componentelor seriei de timp multivariabile

Figura 3 Sursele originale

Figura 4 Componentele seriei de timp multivariabile

Figura 5 Sursele estimate

Figura 6 Predicția surselor estimate

Figura 7 Predicția componentelor seriei de timp originale

Figura 8 Seriile de timp financiare

Figura 9 Sursele estimate ale seriei de timp financiare

Figura 10 Predicția surselor independente ale seriei de timp financiare

Figura 11 Predicția componentelor seriei de timp financiare originale

1. Separarea ”oarbă” a surselor

Separarea ”oarbă” a surselor (BSS) constă în estimarea surselor independente care generează seria de timp multivariabilă originală, utilizând Analiza Componentelor Independente (ICA), [13]. Problema este de mare actualitate în domeniul științelor fizice și ingineresti, biologiei, științelor socio-economice, etc., unde prezintă interes estimarea surselor endogene, necunoscute, din măsurătorile exogene și oferă noi soluții în analiza acestor categorii de date.

Modelul utilizat pentru BSS presupune existența a n semnale independente $s_1(t), \dots, s_n(t)$ și măsurarea mixărilor liniare și instantanee ale acestora: $x_1(t), \dots, x_n(t)$:

$$x_i(t) = \sum_{j=1}^n a_{ij}s_j(t) + n_i(t) \quad (1)$$

pentru fiecare $i = 1, n$. Modelul poate fi reprezentat compact prin următoarea ecuație de mixare:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \quad (2)$$

unde $\mathbf{s}(t) = [s_1(t), \dots, s_n(t)]^T$ este un vector coloană, $n \times 1$, ce conține semnalele sursă, în timp ce vectorul $\mathbf{x}(t)$ conține n semnale observate, iar ”matricea de mixare” \mathbf{A} , $n \times n$, conține coeficienții de mixare.

BSS constă în estimarea vectorului sursă $\mathbf{s}(t)$ folosind numai datele observate $\mathbf{x}(t)$, ipoteza de independență a componentelor vectorului $\mathbf{s}(t)$ și, posibil, unele informații a priori privind distribuția de probabilitate a intrărilor, $\mathbf{s}(t)$. Aceasta poate fi formulată ca estimarea unei ”matrici de separare”, \mathbf{W} , $n \times n$, ale cărei ieșiri, vectorul $\hat{\mathbf{s}}(t)$, reprezintă o estimare a vectorului semnalelor sursă, în cazul unei mixări instantanee, având forma (vezi Figura 2):

$$\hat{\mathbf{s}}(t) = \mathbf{W}\mathbf{x}(t) \quad (3)$$

Elementul cheie în succesul aplicării unei metode de separare, constă în alegerea criteriului statistic în raport cu care se realizează separarea. Astfel, dacă semnalele sunt temporal coerente, este posibil a rezolva problema BSS folosind numai statistici de ordinul doi. În acest context a fost dezvoltat algoritmul, SOBI (Second Order Blind Identification), care estimează sursele originale, pe baza autocorelațiilor pentru mai multe valori ale întârzierii și diagonalizarea aproximativă asociată acestora, [14]. Această metodă prezintă avantajul că este ușor de implementat și conduce, în multe cazuri, la soluții liniare care necesită un efort de calcul redus, și folosește tehnici numerice standard.

Numărul surselor independente rezultă în urma descompunerii după valori proprii a matricii de covarianță eșantion, $\hat{\mathbf{R}}_x(0) = \frac{1}{N} \sum_{t=1}^N \mathbf{x}(t)\mathbf{x}(t)^T$, a vectorului măsurătorilor $\mathbf{x}(t)$.

$$\hat{\mathbf{R}}_x(0) = \mathbf{H}\Delta\mathbf{H}^T \quad (4)$$

unde

$$\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_m] \quad (5)$$

și

$$\Delta = \text{diag}[\lambda_1, \dots, \lambda_m] \quad (6)$$

cu $\lambda_i \geq \lambda_j$ for $i < j$. Numărul surselor independente poate fi estimat pe baza spectrului rezultat Δ , [15], [16].

Aplicarea procedurii de separare "oarbă" a surselor independente de aplică după centrarea componentelor originale ale seriei originale (eliminarea valorilor medii ale acestora).

2. Staționizarea surselor independente rezultate în urma aplicării BSS

Sursele independente rezultate în urma aplicării BSS pot fi caracterizate printr-o comportare nestaționară, în care tendințele și alte caracteristici cvasi-staționare se pot modifica în timp. În [2] se propun soluții pentru obținerea staționarității acestor componente, în special în valoarea medie și în dispersie, cazurile cel mai frecvent întâlnite în practică. Transformările care se aplică acestor componente, în scopul obținerii staționarității (estimarea și eliminarea tendinței, diferențierea nesezonieră sau sezonieră a componentelor de diferite ordine, etc.), urmează a fi reținute pentru a fi aplicate valorilor de predicție ale acestor componente, rezultate în urma aplicării metodologiei Box-Jenkins de modelare și predicție.

3. Modelarea și predicția surselor independente ale seriei de timp multivariabile utilizând metodologia Box-Jenkins

În cadrul acestei secțiuni vom prezenta succint procedura de modelare și predicție a seriilor de timp ce utilizează metodologia Box-Jenkins, [1]. Principalele tipuri de modele dinamice, unele dintre ele urmând a fi utilizate în rezolvarea problemei ce face obiectul prezentei Cereri de Brevet, respectiv analiza, modelarea și predicția seriilor de timp, sunt discutate în [2].

Metodologia care se utilizează în acest scop furnizează modele sub forma unor ecuații cu diferențe stohastice, plecând de la datele de observație.

Prezentăm în continuare principalele etape ale metodologiei Box-Jenkins de analiză și modelare a seriilor de timp:

A. Identificarea modelelor

Această etapă de analiză are ca obiect determinarea structurii modelelor seriilor de timp monovariabile menționate anterior. Modelul rezultat, pentru fiecare din serii, structura și estimațiile preliminare ale parametrilor acestuia, constituie un punct de plecare pentru obținerea unor estimații eficiente statistic ale parametrilor modelului final, în etapa de estimare.

Așa cum s-a menționat anterior, pentru modelarea surselor independente, $s_i(t)$, rezultate în urma separării "oarbe" a acestora, după staționarizare, se pot utiliza următoarele tipuri modele, [1], unde $\epsilon(t)$ reprezintă inovația modelului:

- Modelul autoregresiv, AR(p):

$$s_i(t) = -\phi_1 s_i(t-1) - \phi_2 s_i(t-2) - \dots - \phi_p s_i(t-p) + \epsilon(t) \quad (7)$$

- Modelul de medie alunecătoare, MA(q):

$$s_i(t) = \epsilon(t) + \theta_1 \epsilon(t-1) + \theta_2 \epsilon(t-2) + \dots + \theta_q \epsilon(t-q) \quad (8)$$

- Modelul autoregresiv și de medie alunecătoare, ARMA(p,q):

$$s_i(t) = -\phi_1 s_i(t-1) - \phi_2 s_i(t-2) - \dots - \phi_p s_i(t-p) + \epsilon(t) + \theta_1 \epsilon(t-1) + \theta_2 \epsilon(t-2) + \dots + \theta_q \epsilon(t-q) \quad (9)$$

Etapa de determinare a structurii modelului și cea de estimare interacționează în mare măsură. Astfel, se pot estima parametrii unui model cu o structură mai complexă decât este necesar, urmând ca apoi să se decidă, pe baza rezultatelor estimării, asupra simplificărilor care se impun a fi efectuate în structura modelului. În această situație, rezultatele care se obțin în etapa de estimare vor fi utilizate pentru determinarea unei noi structuri. Trebuie subliniat faptul că structura modelului va fi una aproximativă, deoarece determinarea acestuia nu poate fi realizată numai cu argumente pur matematice.

B. Estimarea parametrilor modelelor

Așa cum s-a menționat anterior în etapa de specificare a structurii modelului (identificare) este selectată o clasă de modele care permite reprezentarea adecvată din punct de vedere statistic, cu număr minim de parametri, a datelor analizate. În cadrul etapei de estimare, făcând uz în mod mai eficient de datele disponibile, se obțin estimații precise ale parametrilor modelului a cărui structură și parametri preliminari au fost determinați în etapa anterioară.

Estimarea parametrilor modelului presupune că datele analizate verifică un model probabilistic a cărui structură este cunoscută, dar care include parametri necunoscuți. Valorile acestor parametri se determină în raport cu un criteriu de *optimalitate*, astfel încât estimațiile acestora să se obțină cu cea mai bună precizie posibilă. Unul din obiectivele principale ale estimării parametrilor modelului îl constituie *robustetea*, adică determinarea unor modele care să nu fie puternic sensibile la proprietățile realizării analizate și care să poată reprezenta astfel, în mod adecvat, largi clase de probleme.

C. Validarea-diagnoza modelelor

Odată ce structura modelului a fost determinată, iar parametrii acestuia au fost estimați, urmează să se decidă dacă modelul este adecvat și, eventual, poate fi îmbunătățit. În cazul în care se constată că acesta nu este reprezentativ pentru datele analizate se va reveni la etapa de specificare a structurii (identificare), pentru selectarea uneia sau mai multor structuri de

modele cu posibilități mai mari de reprezentare a datelor care fac obiectul analizei. Testele de validare-diagnoză pot furniza informații în legătură cu eventualele modificări care se impun a fi efectuate în structura modelului, pentru îmbunătățirea calității acestuia.

D. Predicția seriilor de timp

În cazul în care modelul seriei monovariabile, care face obiectul analizei, este complet cunoscut, deci structura și coeficienții acestuia, precum și toate valorile variabilelor care intervin în acesta, anterioare momentului predicției, sunt precizate, se poate realiza predicția valorilor viitoare ale componentei analizate. Astfel, la momentul curent t *originea* predicției, se presupune că dorim să realizăm predicția seriei înainte peste p intervale de timp sau pași. Intervalul de timp p este denumit *orizont* de predicție. În această etapă, se pot determina și intervalele de încredere ale valorilor de predicție, pentru diferite grade de încredere.

După estimarea valorilor de predicție ale componentelor independente ale seriei originale se impune efectuarea operațiilor inverse ce au fost efectuate pentru obținerea staționarității acestor componente, înaintea modelării lor cu metodologia Box-Jenkins.

4. Revenirea în spațiul seriei de timp multivariabile originale

Valorile de predicție rezultate și limitele de încredere sunt mixate utilizând coeficienții de mixare, a_{ij} , ai matricei \mathbf{A} , pentru a obține valorile de predicție, $x_i^p(t)$ ale componentelor seriei de timp multivariabile originale $x_i(t)$. În această etapă a procedurii, ce face obiectul Cererii de Brevet, se realizează corectarea rezultatelor de predicție cu valorile medii ale componentelor seriei originale.

Avantajele invenției în raport cu soluțiile actuale

Avantajele pe care soluția propusă le prezintă, comparativ cu soluțiile cunoscute în prezent, sunt:

- Analiza directă a seriilor de timp multivariabile, atât modelarea cât și predicția acestora, de tipul celei menționate în secțiunea dedicată Stadiului tehnicii în domeniu, ridică multe probleme în practică, datorită dificultății obținerii unei reprezentări canonice a modelului seriilor de timp multivariabile, citeACA; soluția ce face obiectul Cererii de Invenție evită astfel de probleme.
- Soluția propusă de predicție a seriilor de timp multivariabile, este una necovențională, originală, care transferă problema ce face obiectul analizei din spațiul datelor originale, în spațiul surselor independente, de dimensiune mult redusă și unde se dispune de un arsenal bogat de metode și tehnici de analiză, dată fiind independența statistică a surselor independente, generatoare ale seriei de timp multivariabile.
- Modelarea și predicția sursele independente într-un nou spațiu de dimensiune redusă, față de cel al seriei de timp multivariabile originale, simplifică problema de predicție. În acest caz, numărul seriilor de timp independente rezultate în urma aplicării separării "oarbe" a componentelor seriei multivariabile este, de regulă, mai mic, și mai mult, modelarea și predicția se va realiza pentru serii de timp monovariabile.
- Rezultatele modelării și predicției pot fi ușor transferate, în spațiul original al datelor seriei de timp multivariabile, utilizând matricea de mixare a surselor independente, \mathbf{A} , determinată în etapa de separare "oarbă" a surselor.
- Procedura ce face obiectul invenției va permite realizarea cu efort minim a predicției seriilor de timp multivariabile în diferite domenii de aplicație.

Modul de realizare și aplicare a invenției

Realizarea și aplicarea invenției constă în prelucrarea datelor reprezentând componentele seriei de timp multivariabile ce face obiectul analizei, conform procedurii discutate și prezentate în Figura 1. Realizarea și aplicarea invenției sunt exemplificate în 2 studii de caz, unul realizat în simulare, pentru o serie de timp multivariabilă cu 5 componente, iar celalalt pentru o serie de timp financiară multivariabilă cu 4 componente.

A. Rezultate obținute în simulare

Seria de timp multivariabilă artificială, cu 5 componente, a fost generată folosind 3 surse independente aleatoare și o matrice de mixare, \mathbf{A} , generată aleator.

Sursele originale utilizate sunt reprezentate în Figura 3. Semnalele mixate, reprezentând componentele seriei de timp multivariabile, obținute prin mixarea surselor independente din Figura 3 cu matricea de mixare \mathbf{A} , generată aleator, sunt reprezentate în Figura 4.

Sursele independente au fost estimate din datele seriei multivariabile folosind algoritmul de separare SOBI, [14]. Sursele estimate sunt reprezentate în Figura 5; numărul surselor independente, estimat pe baza spectrului rezultat Δ (6) a rezultat egal cu cel real utilizat în simulare, 3.

Acestea au fost asimilate cu 3 serii de timp monovariabile, pentru care s-a aplicat metodologia Box-Jenkins, în scopul modelării și predicției acestora. Au rezultat următoarele modele pentru cele 3 serii: $s_1(t)$, $s_2(t)$ și $s_3(t)$, respectiv, unde ϵ_t reprezintă inovația modelului, iar B este operatorul de întârziere pe un pas, $Bx_t = x_t - x_{t-1}$:

$$(1 - B)s_1(t) = (1 - 0.402B - 0.239B^2)\epsilon_t, \text{ cu } \sigma_\epsilon^2 = 0.330 \quad (10)$$

$$s_2(t) = -9.066 + (1 + 0.467B - 0.188B^4)\epsilon_t, \text{ cu } \sigma_\epsilon^2 = 0.552 \quad (11)$$

$$(1 + 0.181B^3)s_3(t) = 1.076 + (1 + 0.472B)\epsilon_t, \text{ cu } \sigma_\epsilon^2 = 0.571 \quad (12)$$

Pe baza modelelor rezultate s-a realizat predicția celor 3 surse independente pentru un orizont de timp de 12 pași. Intervalul de încredere pentru valorile de predicție a fost ales 95%. Rezultatele predicției sunt reprezentate în Figura 6. Pe baza acestor rezultate și cunoscând matricea de mixare a surselor $\hat{\mathbf{A}}$, s-au determinat valorile de predicție pentru seria de timp multivariabilă originală. Aceste valori și limitele de încredere sunt reprezentate în Figura 7.

B. Rezultate obținute cu date reale

Seria de timp multivariabilă care a făcut obiectul analizei conține 4 componente, ce reprezintă randamentul obligațiunilor, înregistrat zilnic, pentru 4 țări:

- Randamentul obligațiunilor americane (US bond).
- Randamentul obligațiunilor britanice (UK bond).
- Randamentul obligațiunilor vest germane (West Germany bond).
- Randamentul obligațiunilor japoneze (Japan bond).

Datele provin din [18] unde se regăsesc înregistrări zilnice ai celor 4 indicatori, în perioada 1.04.1986-29.12.1989. Prezentul studiu de caz utilizează numai 460 din acestea, înregistrate în ultima perioadă de timp.

Componentele seriei de timp multivariabile sunt reprezentate în Figura 8. Separarea surselor independente s-a făcut cu algoritmul SOBI, [14], rezultând 4 surse independente reprezentate

în Figura 9. Aplicarea metodologiei Box-Jenkins a condus la următoarele modele pentru cele 4 surse estimate:

$$(1 - B)(1 + 0.105B^5)s_1(t) = \epsilon_t, \text{ cu } \sigma_\epsilon^2 = 0.000036$$

$$(1 - B)(1 + 0.077B^8)s_2(t) = \epsilon_t, \text{ cu } \sigma_\epsilon^2 = 0.001936$$

$$(1 - B)(1 - 0.111B)s_3(t) = (1 + 0.117B^6 + 0.116B^8)\epsilon_t, \text{ cu } \sigma_\epsilon^2 = 0.001225$$

$$(1 - B)s_4(t) = \epsilon_t \text{ cu } \sigma_\epsilon^2 = 0.001936$$

Valorile de predicție, rezultate pentru cele 4 surse, cu un grad de încredere de 95%, pentru un orizont de predicție de 12 pași, sunt reprezentate în Figura 10. Valorile de predicție ale seriei multivariabile originale, rezultate în urma mixării rezultatelor de predicție ale surselor independente sunt prezentate în Figura 11.

Aplicarea în practică a invenției

În ceea ce privește aplicarea practică a invenției, considerăm ca posibil următorul mod de utilizare:

- Disponând de o serie de timp multivariabilă, $\mathbf{x}(t)$, cu n componente și N eșantioane, pentru care se dorește a se efectua predicția, se procedează la centrarea valorilor componentelor acesteia prin eliminarea și reținerea valorilor medii ale celor n componente.
- Seria de timp multivariabilă rezultată reprezintă date de intrare pentru procedura de calcul ce realizează estimarea surselor independente $s_j(t)$ și a matricei de mixare a acestora \mathbf{A} , conform procedurii reprezentate în Figura 2.
- În scopul obținerii staționarității componentelor independente, $s_j(t)$, în special în valoarea medie și în dispersie, pentru fiecare din acestea se aplică o procedură adecvată de diferențiere simplă sau sezonieră, de ordin corespunzător, pentru eliminarea componentelor tendință sau a componentelor sezoniere.
- Componentele staționare rezultate sunt modelate, de regulă cu modele parametrice de tip autoregresiv (AR), de medie alunecătoare (MA), autoregresive și de medie alunecătoare (ARMA), utilizând metodologia Box-Jenkins; în această etapă pot fi utilizate toate metodele și tehnicile de modelare a seriilor de timp monovariabile. Modelele rezultate urmează a fi validate.
- Utilizând modelele rezultate, fiecare sursă independentă este predictată în mod separat, pentru un anumit orizont de predicție și grad de încredere. În această etapă, după calculul valorilor de predicție, se impune efectuarea operațiilor inverse ce au fost efectuate pentru obținerea staționarității acestor componente, înaintea modelării lor cu metodologia Box-Jenkins.
- Valorile de predicție rezultate și limitele de încredere sunt mixate utilizând coeficienții de mixare, a_{ij} , ai matricei \mathbf{A} , pentru a obține valorile de predicție, $x_i^p(t)$ ale componentelor seriei de timp multivariabile originale $x_i(t)$.
- În ultima etapă se realizează corectarea rezultatelor de predicție cu valorile medii ale componentelor seriei, eliminate în prima etapă a procedurii.

Procedura propusă spre brevetare, este una neconvențională, originală, de predicție a seriilor de timp cu mai multe componente, cu interacțiuni între acestea, care se reduce în final la predicția unor serii de timp monovariabile, într-un spațiu de dimensiune redusă. Aceasta simplifică rezolvarea problemei inițiale, cunoscută ca una ce ridică multe probleme în practică. Procedura poate fi utilizată în domeniile ingineriei, științelor fizice, biologiei, medicinei, sociologiei, hidrologiei, geofizicii, economiei, domenii care se confruntă cu analiza unor date de măsură și de observație.

Bibliografie

- [1] G. E. P. Box, G. M. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, 2-nd Edition, Holden Day, San Francisco.
- [2] Th. D. Popescu (2000), *Serii de timp. Aplicații în analiza sistemelor*, Editura Tehnică, București.
- [3] J. G. De Gooijer, R. J. Hyndman (2006), 25 years of time series forecasting, *International Journal of Forecasting*, 443-473.
- [4] J. Scott Armstrong (2001), Standards and practices for forecasting, *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Kluwer Academic Publishers, 1-46.
- [5] H. Lütkepohl (1993), *Introduction to Multiple Time Series Analysis*, Springer-Verlag Berlin Heidelberg.
- [6] H. Lütkepohl (2005), *New Introduction to Multiple Time Series Analysis*, Springer-Verlag Berlin Heidelberg.
- [7] M. J. Artis, W. Zhang (1990), BVAR forecasts for the G-7. *International Journal of Forecasting* 349-362.
- [8] J. P. LeSage, M. Magura (1991), Using interindustry input output relations as a Bayesian prior in employment forecasting models, *International Journal of Forecasting*, 231- 238.
- [9] D. E. Spencer (1993), Developing a Bayesian vector autoregressive forecasting model, *International Journal of Forecasting*, 407-421.
- [10] J. P. LeSage (1989), Incorporating regional wage relations in local forecasting models with a Bayesian prior, *International Journal of Forecasting*, 37-47.
- [11] R. E. Kalman (1960), A new approach to linear filtering and prediction problems, *Transactions of the ASME Journal of Basic Engineering*, 35-45.
- [12] R. H. Shumway, D. S. Stoffer (1982), An approach to time series smoothing and forecasting using the EM algorithm, *Journal of Time Series Analysis*, 253-264.
- [13] A. Hyvärinen, J. Karhunen, E. Oja (2001), *Independent Component Analysis*, John Wiley.
- [14] A. Belouchrani, K. Abed Meraim, J. F. Cardoso, E. Moulines (1997), A blind source separation technique using second - order statistics, *IEEE Trans. Signal Processing* 45, 434-444.
- [15] M. Wax, T. Kailath (1983), Determining the number of signals by information theoretic criteria, *Workshop on Spectral Estimation II, Florida, USA*, 192-196.
- [16] Y. Yin, P. Krishnaiah (1987), Methods for detection of the number of signals, *IEEE Trans. on ASSP* 35, 1533-1538.

h4

- [17] M. Tertisco, P. Stoica, Th. D. Popescu (1985), Modelarea și predicția seriilor de timp, Editura Academiei, București.
- [18] T. C. Mills, R. N. Markellos (1993), The Econometrics Modelling of Financial Time Series, Cambridge University Press.

PROCEDURĂ PENTRU PREDICȚIA SERIILOR DE TIMP MULTIVARIABILE UTILIZÂND SEPARAREA "OARBĂ" A SURSELOR INDEPENDENTE

REVENDICĂRI

1. Procedură pentru predicția seriilor de timp multivariabile, cu aplicabilitate în analiza datelor de măsură și de observație, bazată pe separarea "oarbă" a surselor independente care au generat seria de timp multivariabilă originală, **caracterizată prin aceea că** predicția se realizează pe sursele independente, într-un nou spațiu de dimensiune redusă, față de cel al seriei originale.
2. Procedura pentru predicția seriilor de timp multivariabile, conform revendicării 1, **caracterizată prin aceea că**, simplifică problema de predicție, dat fiind faptul că numărul componentelor pentru care se face predicția este mai mic, față de cel al componentelor seriei de timp multivariabile originale, iar procedurile de modelare și predicție se aplică pe serii monovariabile, pentru care se dispune de un arsenal bogat de metode și tehnici.
3. Procedură pentru predicția seriilor de timp multivariabile, conform revendicării 1, **caracterizată prin aceea că** permite, în mod simplu, transferul rezultatele predicției în spațiul original al datelor seriei de timp multivariabile, utilizând matricea de mixare a surselor independente, **A**, determinată în etapa de separare "oarbă" a acestora.
4. Procedură pentru predicția seriilor de timp multivariabile, conform revendicării 1, **caracterizată prin aceea că** permite realizarea cu efort minim a predicției seriilor de timp multivariabile în domeniile ingineriei, științelor fizice, biologiei, medicinei, sociologiei, hidrologiei, geofizicii, economiei, și evită problemele ce apar în aplicarea metodelor conventionale de modelare și predicția a seriilor de timp multivariabile.

Li

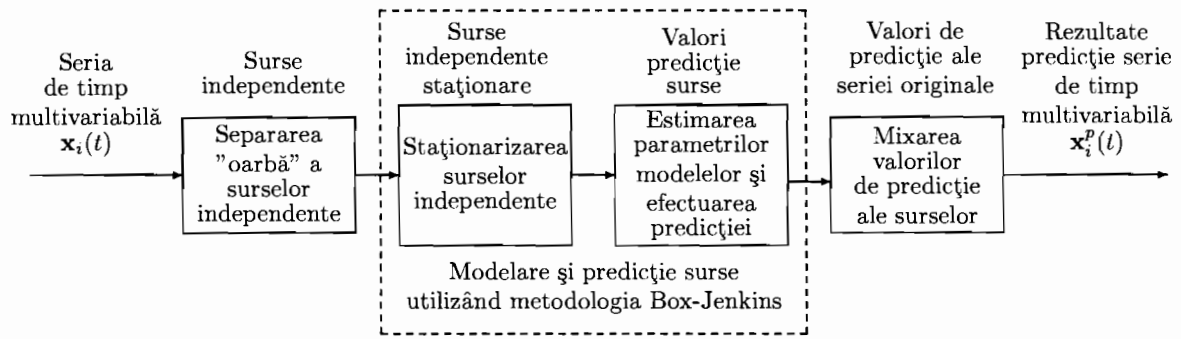


Figure 1: Procedura de predicție a seriilor de timp multivariabile

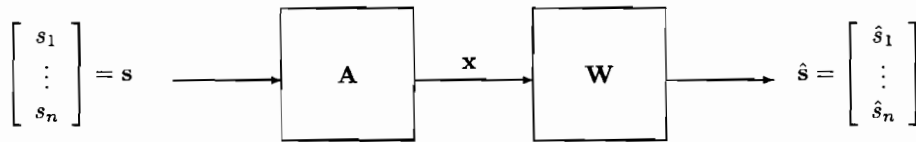


Figure 2: Mixarea și separarea componentelor seriei de timp multivariabile.

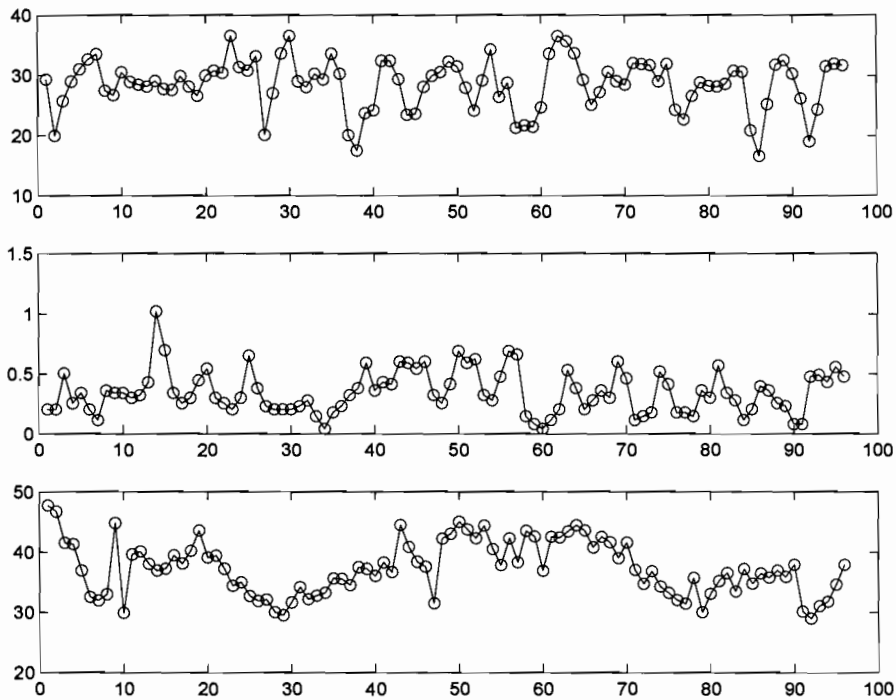


Figure 3: Sursele originale

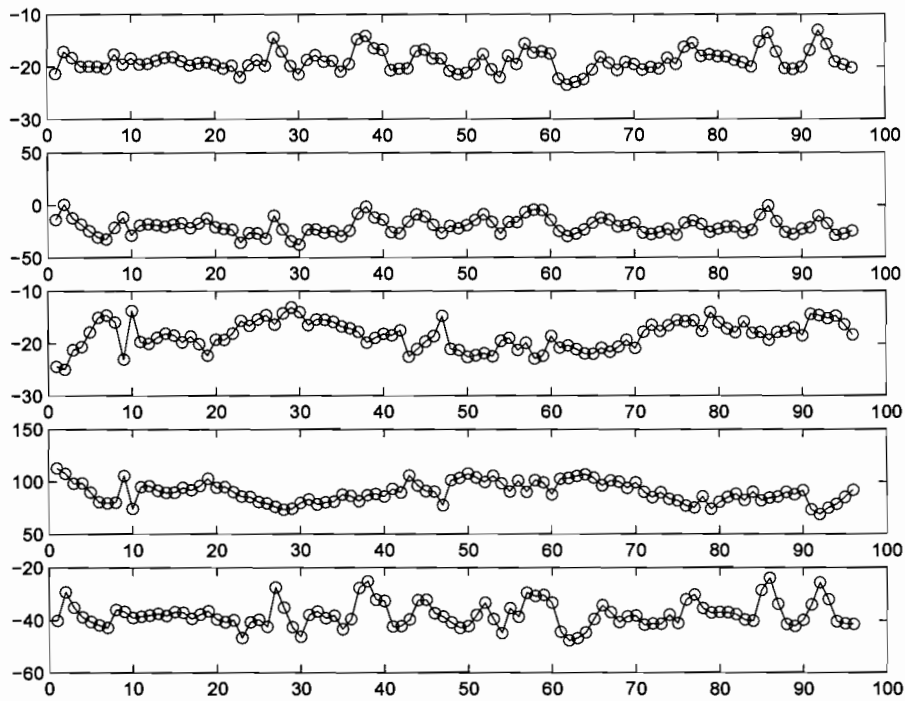


Figure 4: Componentele seriei de timp multivariabile

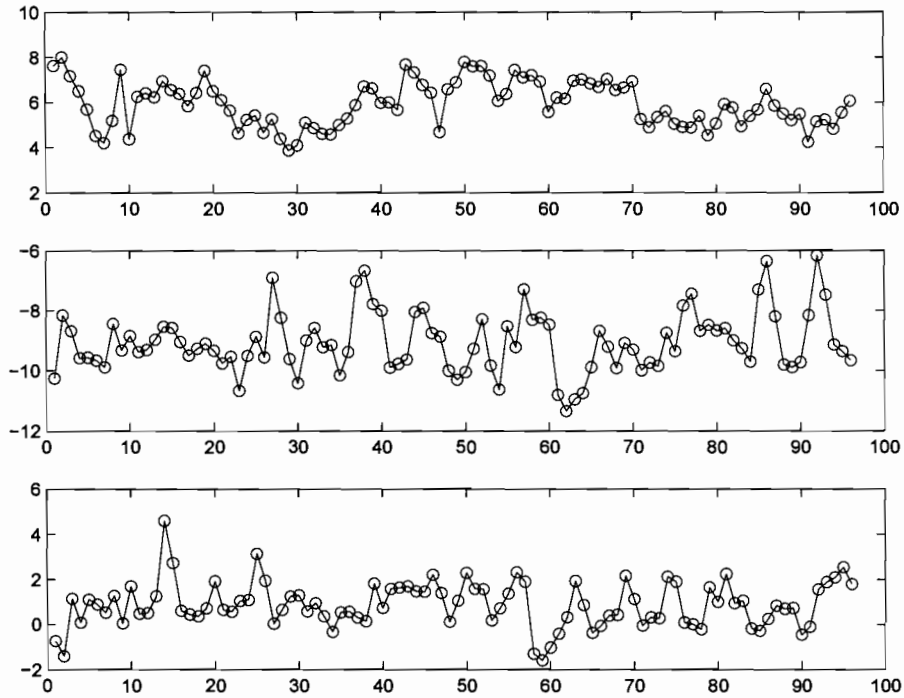


Figure 5: Sursele estimate

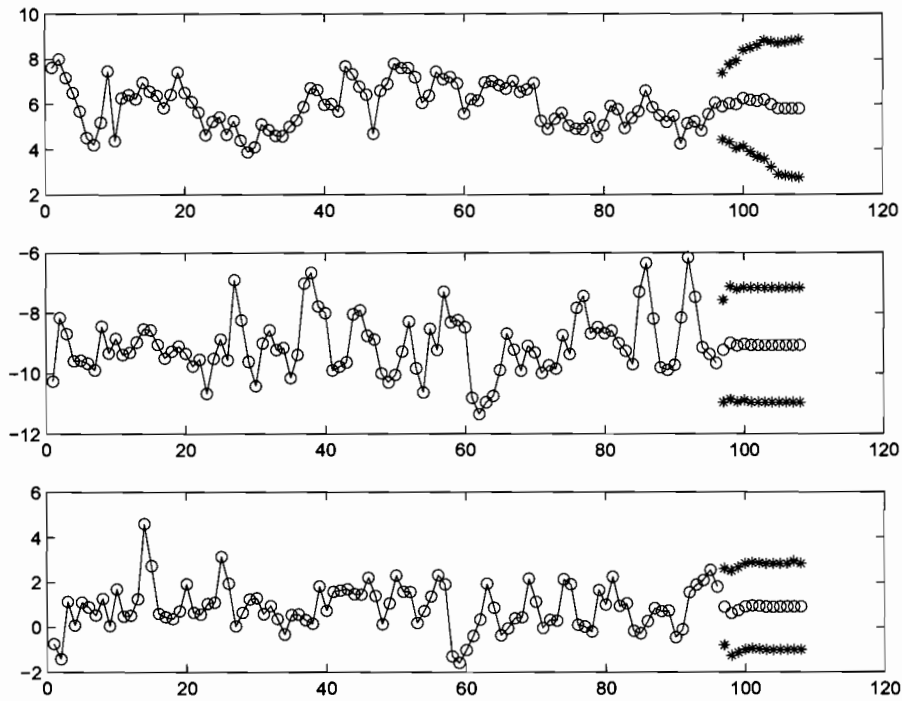


Figure 6: Predicția surselor estimate.

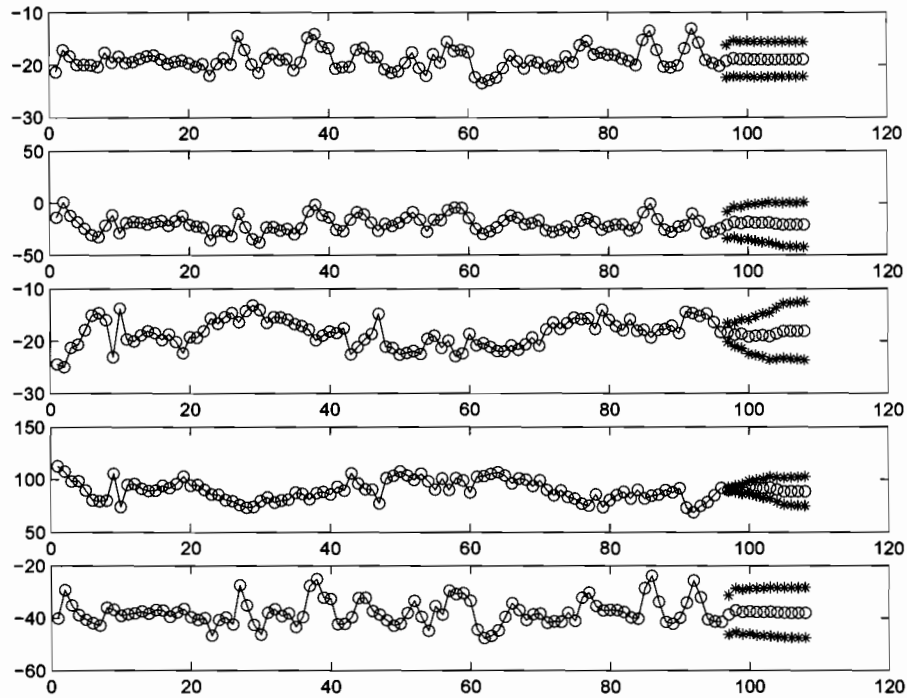


Figure 7: Predicția componentelor seriei de timp originale.

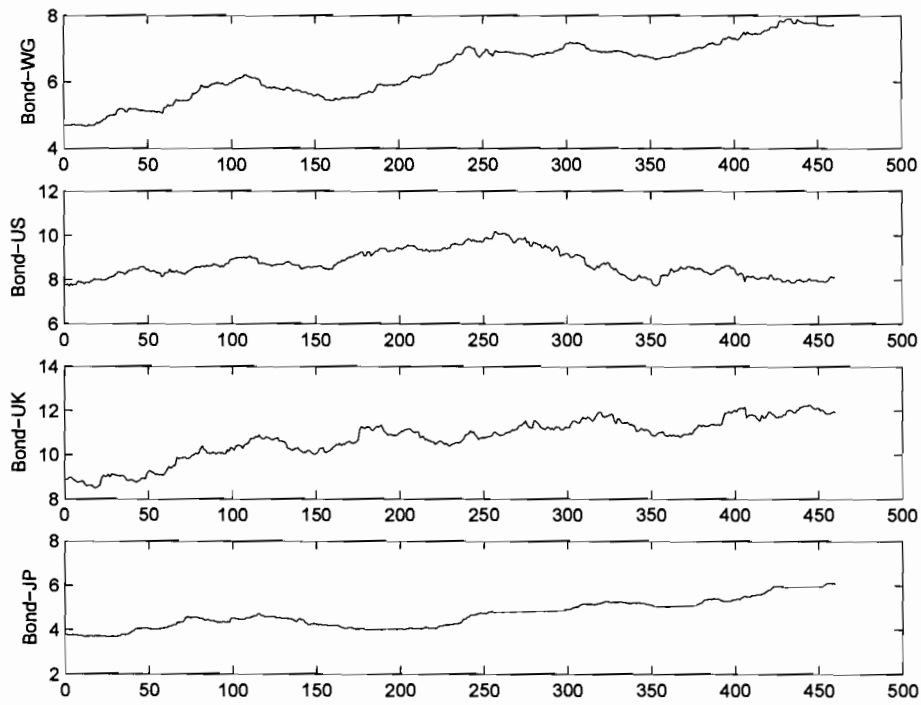


Figure 8: Seriile de timp financiare

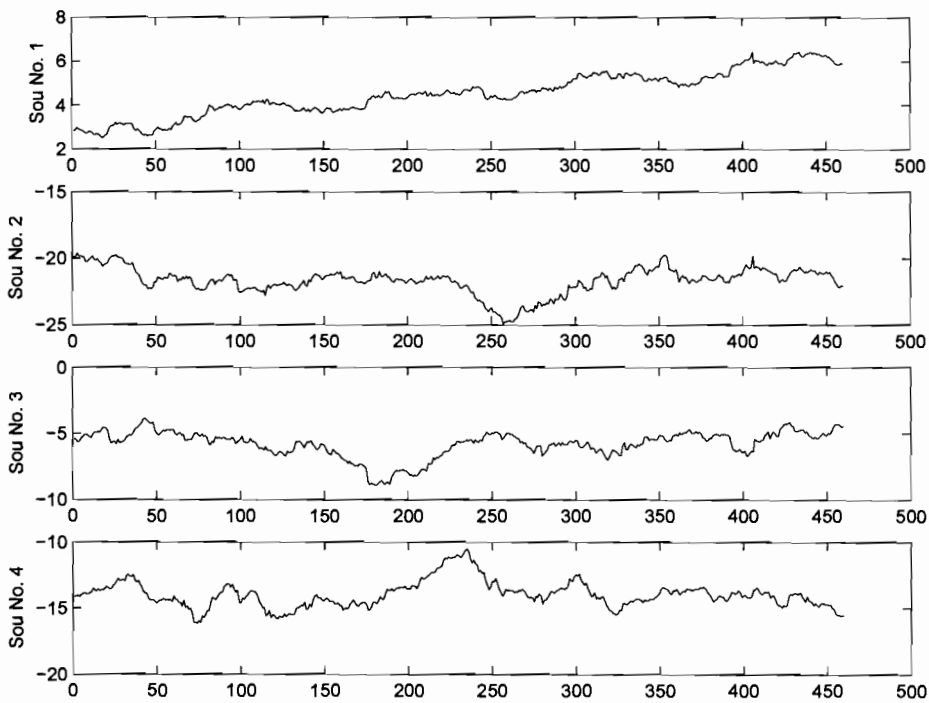


Figure 9: Sursele estimate ale seriei de timp financiare

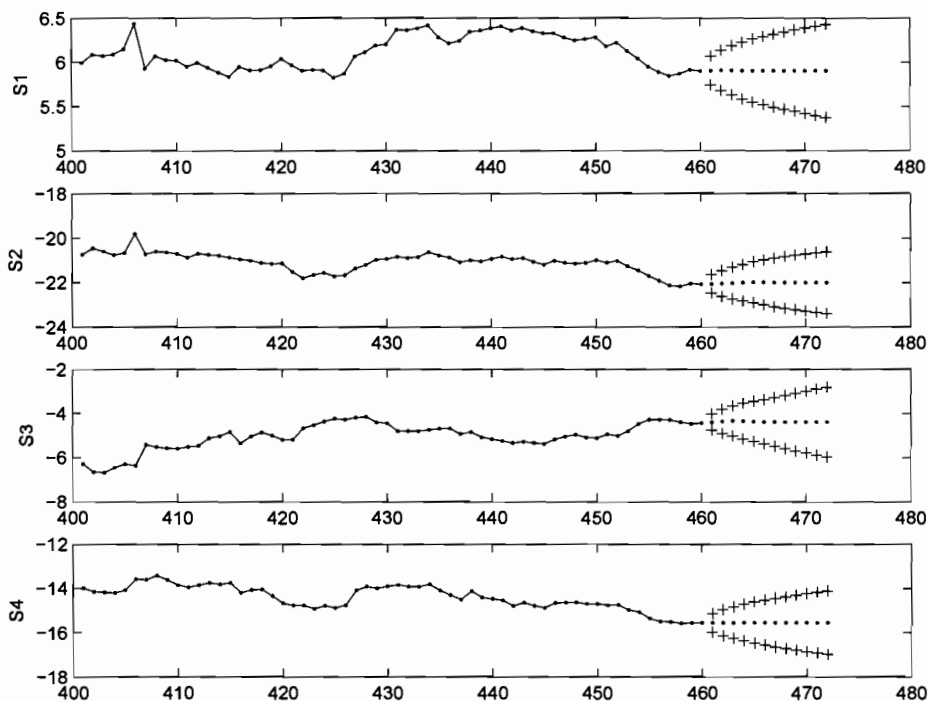


Figure 10: Predicția surselor independente ale seriei de timp financiare

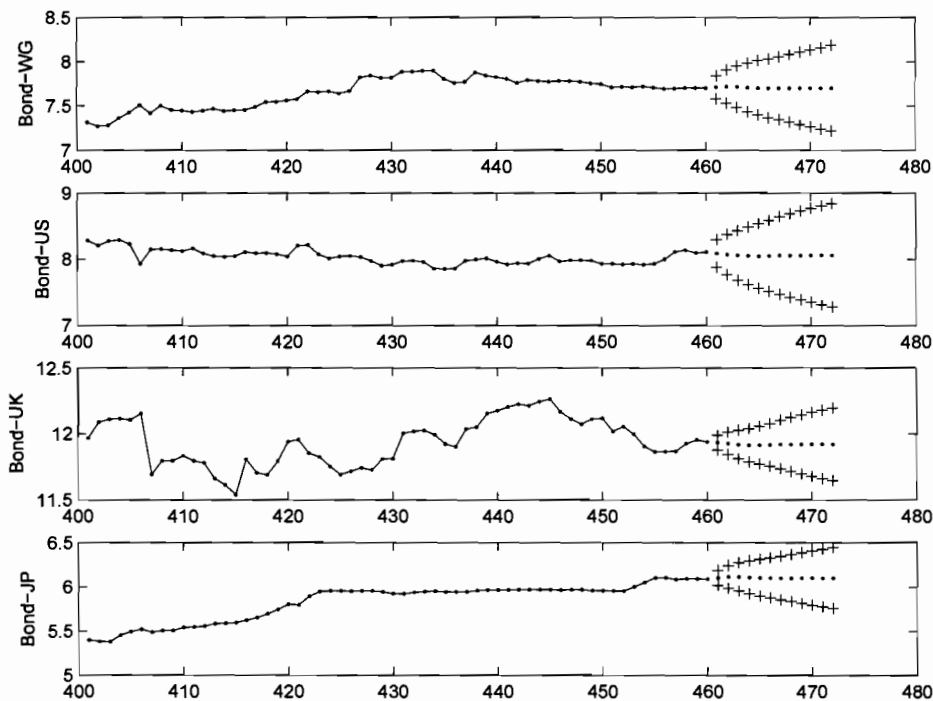


Figure 11: Predicția componentelor seriei de timp financiare originale

PROCEDURĂ PENTRU PREDICȚIA SERIILOR DE TIMP MULTIVARIABILE UTILIZÂND SEPARAREA "OARBĂ" A SURSELOR INDEPENDENTE

REVENDICĂRI

Procedură pentru predicția seriilor de timp multivariabile, **caracterizată prin aceea că** face uz de estimarea "oarbă" a surselor independente, care au generat seria de timp multivariabilă originală, precum și de estimarea matricei de mixare a acestora, într-un nou spațiu de dimensiune redusă, față de cel al datelor inițiale, în care se realizează predicția surselor independente, utilizând metodologia Box-Jenkins, rezultatele de predicție obținute, aplicate matricei de mixare, furnizând valorile de predicție ale seriei de timp originale, așa cum se prezintă în Figura 1.