ANALELE ŞTIINŢIFICE ALE UNIVERSITĂŢII "ALEXANDRU IOAN CUZA" DIN IAŞI
Tomul L/LI Ştiinţe Economice 2004/2005

IDENTIFICAREA TRANSFORMĂRILOR NECESARE STAŢIONARIZĂRII UNEI SERII DE TIMP

MARIANA GAGEA*

L'identification des transformations nécessaires à stationnariser une série temporelle

Résumé

Cette étude a comme objectif la présentation des modalités d'identification des transformations que nous pouvons appliquer à une série temporelle non-stationnaire pour la rendre stationnaire.

La présentation théorique sommaire des instruments utilisés est complétée par l'exemple pratique de la série temporelle définie par le nombre des arrivées dans les pensions touristiques rurales de Roumanie, entre janvier 2002 et décembre 2004. L'analyse des donnés a été réalisée avec le logiciel SPSS.

Mots clés: série temporelle, stationnarité, tests statistiques, différenciation, transformation Box – Cox.

Introducere

Analiza modernă a seriilor de timp, realizată cu metodologia Box-Jenkins, presupune o succesiune de etape care se finalizează cu previziunea nivelului fenomenului studiat. Obiectivul final al analizei impune găsirea procesului stochastic generator al datelor serie de timp, care să permită inferenţa comportamentului observat din trecut în viitor. În aceste condiţii, seria de timp observată trebuie să satisfacă, în primul rând, condiţia de staţionaritate.

În general, seriile de timp sunt nestaţionare, fiind caracterizate de tendinţe de evoluţie şi de heteroscedasticitate. Seriile de timp economice prezintă o nestaţionaritate omogenă, adică sunt serii care în urma unor transformări devin staţionare.

Instrumentele statistice folosite în identificarea transformărilor necesare pentru staţionarizarea unei serii de timp sunt: cronograma liniară, funcţia de autocorelaţie, diagrama medie – dispersie, testul Levene, testul Dickey-Fuller.

Prezentarea succintă a instrumentelor şi modalităţilor de staţionarizare a unei serii de timp este completată de exemplul practic al seriei definită de „numărul de sosiri în pensiunile turistice rurale din România, în perioada ianuarie 2002 – decembrie 2004". Prelucrarea datelor s-a realizat în programele SPSS şi EViews.

* Asistent la Universitatea „Alexandru Ioan Cuza" Iaşi, Facultatea de Economie şi Administrarea Afacerilor

Analiza staţionarităţii

Staţionaritatea unui proces stochastic se defineşte în sens strict şi în sens larg. În continuare vom face referiri la procese stochastice staţionare în sens larg.

Un proces stochastic Y_t, $t \in T$, este staţionar în sens larg [Bourbonnais, R., 2000, pp. 221-222] dacă satisface condiţiile:

- $M(Y_t) = \mu \ \forall t, t \in T$, *media este constantă şi independentă de timp;*
- $V(Y_t) = \sigma^2, \forall t, t \in T$, *varianţa este constantă în timp;*
- $\mathrm{cov}(Y_t, Y_{t+k}) = \gamma(k)$, *covarianţa depinde doar de decalajul temporal dintre variabile.*

Analiza staţionarităţii unei serii de timp, considerată realizare a unui proces stochastic, presupune următoarea succesiune de operaţii [Alonso, J., H., 2000, p. 123]:

a) se studiază dispersia seriei şi, dacă este cazul, se aplică transformările Box-Cox adecvate pentru a obţine o dispersie constantă în timp;

b) se testează şi se elimină variaţiile sezoniere;

c) se studiază existenţa şi natura tendinţei de evoluţie a observaţiilor şi se alege metoda potrivită pentru staţionarizarea ei.

Ipoteze privind staţionaritatea unei serii de timp sunt formulate prin analiza în paralel a cronogramei liniare şi a funcţiei de autocorelaţie, simplă şi parţială.

Construirea *cronogramei liniare* [Jaba, E., 2002, p.77] presupune respectarea principiului proporţionalităţii, adică a armoniei dintre scara absciselor, pe care se reprezintă timpul, şi scara ordonatelor, pe care se reprezintă nivelul atins în fiecare moment de către fenomenul studiat.

Cronograma (figura 1.a) seriei de timp definită de numărul de sosiri în pensiunile turistice rurale din România permite reţinerea următoarelor observaţii: heteroscedasticitate, întrucât valorile înregistrate prezintă fluctuaţii de amplitudini diferite în timp; tendinţă uşor ascendentă de evoluţie a fenomenului; sezonalitate lunară, pusă în evidenţă de nivelul maxim al fenomenului înregistrat, an de an, în lunile august.

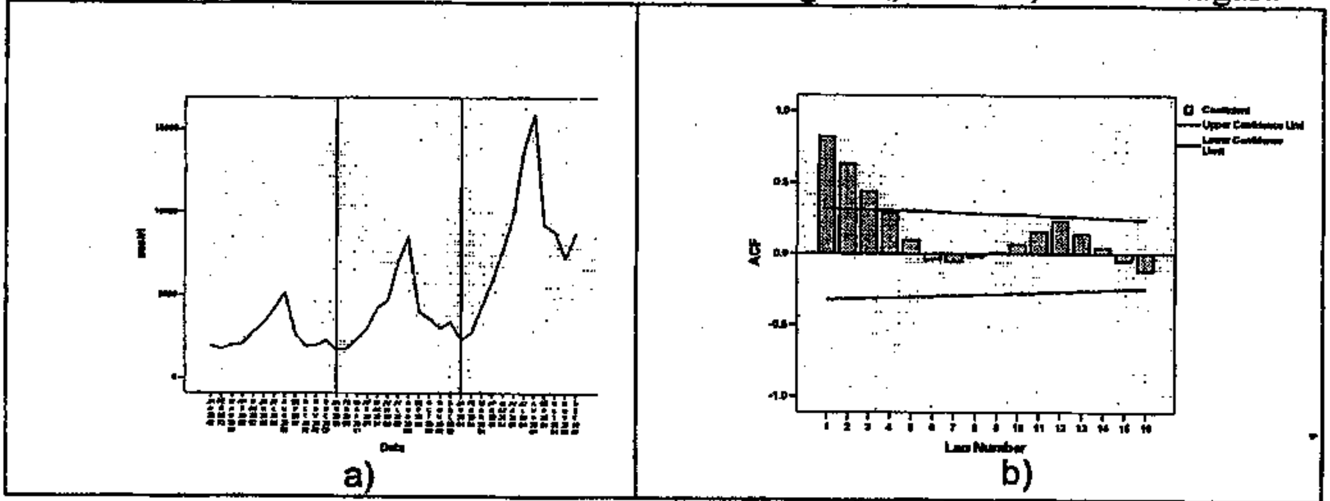

Fig. 1 Dinamica (a) şi autocorelaţiile (b) numărului de sosiri în pensiunile turistice rurale din România, în perioada ianuarie 2002 – decembrie 2004

Funcţia de autocorelaţie (ρ_k) măsoară corelaţia seriei cu ea însăşi decalată cu k perioade. Coeficientul de autocorelaţie [Berdot, J.P., 2002, p. 21] se defineşte prin relaţia: $\rho_k = \dfrac{\sum (Y_t - \mu)(Y_{t+k} - \mu)}{\sum (Y_t - \mu)^2}$

şi tinde către 0 dacă variabila Y care defineşte seria de timp este staţionară.

Reprezentarea grafică a funcţiei de autocorelaţie, simbolizată FAC (ACF în programul SPSS), poartă denumirea de *corelogramă*. Forma corelogramei trădează componentele seriei de timp (tendinţa, componenta sezonieră) sau procesul stochastic care a generat seria respectivă, cu implicaţii asupra analizei staţionarităţii.

Corelograma (figura 1.b) seriei sosirilor în pensiunile turistice rurale din România prezintă o descreştere a coeficienţilor de autocorelaţie, care, în condiţiile unei serii de timp cu număr mic de observaţii, poate fi explicat prin tendinţă. În plus, coeficienţii cu valori semnificative evidenţiază existenţa unei sezonalităţi lunare.

În concluzie, urmare a analizei cronogramei liniare şi a corelogramei pentru seria de timp studiată, se reţin ipotezele care urmează a fi verificate: dispersie variabilă în timp, tendinţă de evoluţie uşor crescătoare a fenomenului şi sezonalitate lunară.

Staţionarizarea unei serii de timp în dispersie

Corectarea dispersiei se face cu transformări Box-Cox.

Transformările Box-Cox au următoarea expresie [Alonso, J., H., 2000, p. 42]:

$$W_t = \begin{cases} Y_t^P, \ P \in R, \ -1 \le P \le 1 \ daca \ P \ne o \\ \ln Y_t, \ daca \ P = 0 \end{cases}$$

Cronograma liniară (figura 1.a) a variabilei studiate evidenţiază variaţii de amplitudini diferite specifice unei serii nestaţionare în dispersie, însă nu oferă informaţii care să conducă la alegerea celei mai bune transformări.

Diagrama medie – rang sau *medie – dispersie* înfăţişează grafic relaţia dintre nivelul mediu şi varianţa unei variabile. În funcţie de modul de distribuire a punctelor, se poate aproxima valoarea parametrului P al transformărilor Box-Cox potrivite în staţionarizarea dispersiei unei serii de timp.

Programul SPSS reprezintă pe abscisa diagramei mediana, care redă nivelul mediu al fenomenului, şi pe ordonată, rangul interquartilic, care exprimă dispersia.

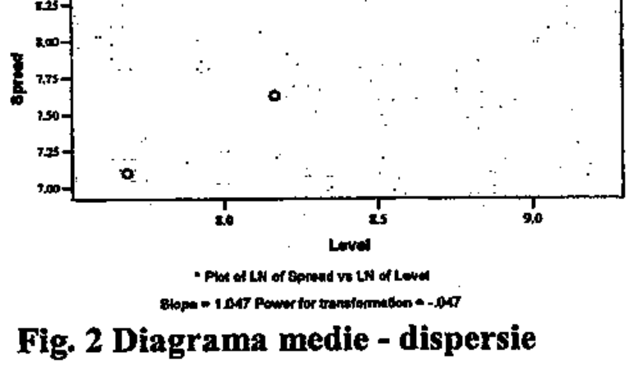

Fig. 2 Diagrama medie - dispersie

Diagrama medie – dispersie (figura 2) construită în SPSS pentru seria numărului de sosiri în pensiunile turistice rurale din România prezintă o distribuire a punctelor deasupra diagonalei principale, iar valoarea recomandată pentru parametrul P este $P = -0,047$. Dilema care trebuie soluționată în acest caz este alegerea celei mai bune valori pentru parametrul P dintre următoarele: $P = -0,5$, valoare mai apropiată de cea recomandată în diagrama medie – dispersie, și $P = 0$, valoare potrivită atunci când punctele sunt distribuite în jurul diagonalei principale a diagramei.

Alegerea se poate face prin două metode. Se compară diagramele medie – dispersie corespunzătoare celor două valori ale parametrului P, fiind considerată mai bună transformarea prin care se obține cea mai mică pantă a dreptei. O altă metodă are la bază testul Levene de homoscedasticitate, transformarea optimă fiind indicată de valoarea cea mai mică a testului.

Tabel 1 Testarea homoscedasticității pentru seria sosirilor în pensiunile turistice rurale

Parametrul P	Levene Statistic	df1	df2	Sig.
1	5.272	2	33	.010
-.5	.159	2	33	.854
0	1.101	2	33	.344

Pentru seria considerată ca exemplu în lucrare, care prezintă inconvenientul unui număr mic de observații, este mai potrivit testul Levene.

Testul Levene verifică ipoteza nulă de existență a homoscedasticității (dispersie constantă în timp). Ipoteza nulă se respinge când nivelul de semnificație *sig* este inferior valorii de 0,05: $sig < 0,05$.

Testul Levene aplicat la datele seriei inițiale ($P = 1$) confirmă ipoteza de heteroscedasticitate, formulată pe baza reprezentărilor grafice.

Se compară valorile testului Levene pentru cele două transformări, $P = -0.5$, respectiv $P = 0$, și se optează pentru transformarea seriei prin inversa rădăcinii pătrate.

Staționarizarea în medie a unei serii de timp

Testul Dickey-Fuller (DF), propus în 1979, completează insuficiențele metodelor folosite până atunci, permițând stabilirea cu exactitate a modelului de proces stochastic, caracterul stochastic sau determinist al tendinței și, implicit, alegerea metodelor adecvate de staționarizare. Se aplică doar pentru serii de timp homoscedastice.

Testul este construit pe următoarele 3 modele [Bourbonnais, R., 2000, p. 231]:

(1) $Y_t = \varphi_1 Y_{t-1} + \varepsilon_t$ - model autoregresiv fără constantă;

(2) $Y_t = \mu + \varphi_1 Y_{t-1} + \varepsilon_t$ - model autoregresiv cu constantă;

(3) $Y_t = \mu + \gamma t + \varphi_1 Y_{t-1} + \varepsilon_t$ - model autoregresiv cu tendință,

unde ε_t reprezintă un proces „zgomot alb”.

Ipoteza testată este $H_0 : \phi_1 = 1$, contra ipotezei alternative $H_1 : \phi_1 < 1$. Ipoteza nulă poate fi formulată în mod echivalent $H_0 : \phi_1 - 1 = 0$.

Autorii au tabelat valorile critice ale testului DF pentru eşantioane de mărimi diferite. Tabelele obţinute sunt corespondente tabelelor distribuţiei t Student.

Regula de decizie: dacă valoarea calculată a testului este mai mare decât valoarea teoretică, se acceptă ipoteza nulă H_0, deci există o rădăcină unitate.

Principiul testului este simplu. Dacă se acceptă ipoteza nulă H_0 în oricare din cele 3 modele, procesul este nestaţionar. Staţionarizarea [Bourbonnais, R., 2000, pp. 229-231] se va face în funcţie de natura sa:

- dacă testul stabileşte că procesul este un TS (Trend Stationary), se procedează la extragerea din seria de timp a trendului estimat prin MCMMP; reziduurile obţinute vor defini un zgomot alb;
- dacă procesul este un DS (Differency Stationary) se va aplica operatorul de diferenţe simple de ordinul d, stabilit prin testări succesive.

Dickey şi Fuller au propus în 1981 o formă a testului extinsă pentru un model autoregresiv de ordin p, AR(p): $Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$.

Testul Augmented Dickey-Fuller (ADF) este construit pe modelele:

$$(4)\ \Delta Y_t = bY_{t-1} + \sum_{i=2}^{p} \beta_i \Delta Y_{t-i+1} + \varepsilon_t$$

$$(5)\ \Delta Y_t = \mu + bY_{t-1} + \sum_{i=2}^{p} \beta_i \Delta Y_{t-i+1} + \varepsilon_t$$

$$(6)\ \Delta Y_t = \mu + \gamma t + bY_{t-1} + \sum_{i=2}^{p} \beta_i \Delta Y_{t-i+1} + \varepsilon_t,$$

unde $\beta_i = \sum_{i=1}^{p} \phi_i$ şi $b = -\left(1 - \sum_{i=1}^{p} \phi_i\right)$.

Testul ADF se derulează similar testului DF, doar valorile critice tabelate diferă.

Se aplică testul ADF pentru seria numărului de sosiri în pensiunile turistice rurale din România. Prezentăm rezultatele testului doar pentru modelul 6 (tabelul 2). Rezultatele obţinute în urma testării modelelor 5 şi 4 confirmă concluzia existenţei unui proces autoregresiv de ordinul 1 fără constantă.

Tabel 2 Rezultatele testului ADF aplicat la seria iniţială

Modelul testat	Parametrul testat	Ipoteze	Valori calculate	Valori teoretice	Decizia
6	b	$H_0 : b = 0$	$t = -2{,}3371$	$t_{0,01} = -4{,}2412$ $t_{0,05} = -3{,}5426$ $t_{0,10} = -3{,}2032$	Se acceptă H_0, există o rădăcină unitate
	γ	$H_0^\gamma : \gamma = 0$	$sig = 0{,}0956$	$\alpha = 0{,}05$	Se acceptă H_0^γ
	μ	$H_0^\mu : \mu = 0$	$sig = 0{,}6565$	$\alpha = 0{,}05$	Se acceptă H_0^μ

Decizia: Coeficientul de regresie nu este semnificativ diferit de 0, seria este generată de un proces DS şi se staţionarizează prin diferenţe simple de ordin 1.

Tendinţa manifestată de sosirile în pensiunile turistice rurale din România, potrivit testului ADF este stochastică şi se staţionarizează printr-o diferenţă simplă de ordinul 1.

Operatorul de diferenţe simple se notează $\Delta^d Y_t$, unde d reprezintă ordinul diferenţelor şi $\Delta Y_t = Y_t - Y_{t-1}$, diferenţa simplă de ordinul 1. În practică, se folosesc frecvent diferenţele de ordinul 1 sau 2.

Eliminarea variaţiilor sezoniere

Corectarea variaţiilor sezoniere presupune aplicarea *diferenţelor sezoniere* de ordinul D: $\Delta_k^D Y_t$, unde k reprezintă periodicitatea componentei sezoniere, D - ordinul diferenţelor sezoniere, $\Delta_k = Y_t - Y_{t-k}$ diferenţa sezonieră de ordinul 1.

Cronograma (figura 3.a) şi corelograma (figura 3.b) prezintă seria sosirilor în pensiunile turistice rurale staţionarizată şi corectată de variaţiile sezoniere. Pentru eliminarea variaţiilor sezoniere s-au aplicat diferenţele sezoniere de ordinul 1.

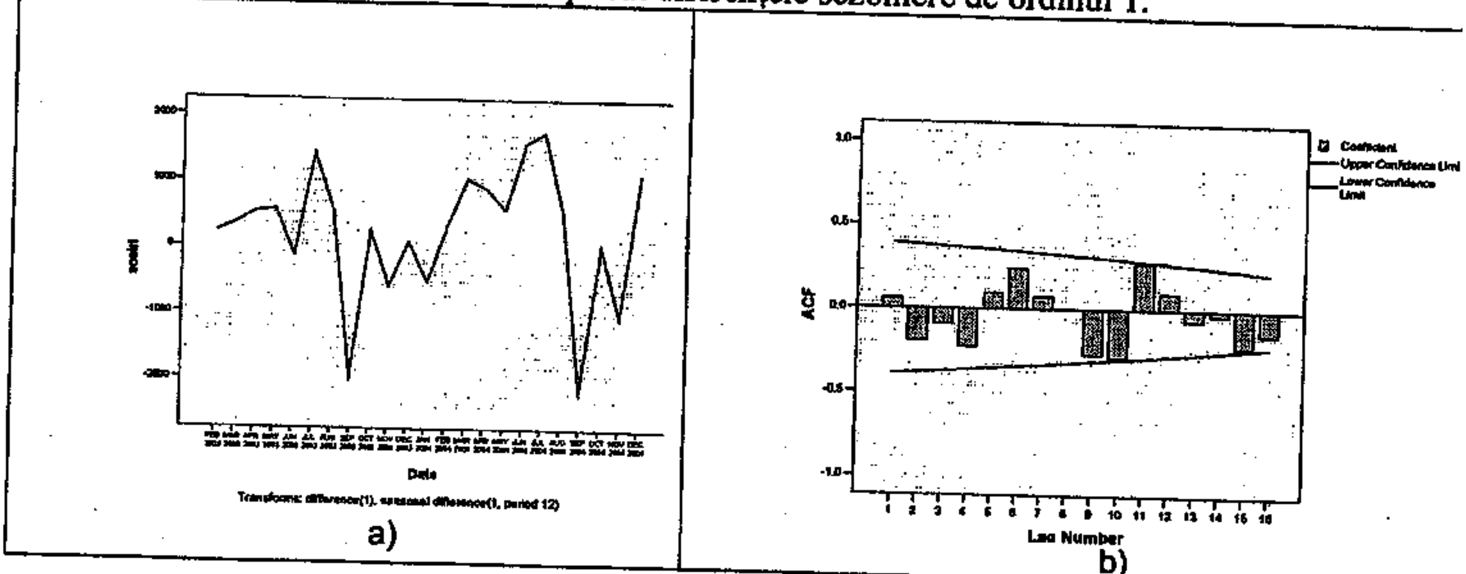

Fig. 3 Dinamica şi autocorelaţiile seriei iniţiale, diferenţiată simplu şi sezonier

Concluzii

Importanţa alegerii cu cât mai mare precizie a transformărilor prin care se staţionarizează o seriei de timp este evidentă. Menţinerea în comportamentul seriei temporale a heteroscedasticităţii sau a unui nivel mediu care variază în timp, după o presupusă staţionarizare, pot degenera imaginea procesului stochastic căutat, cu implicaţii directe asupra prognozei valorilor seriei considerate.

Identificarea transformărilor se bazează, în principal, pe metode grafice a căror eficienţă este limitată de subiectivismul analistului. Metodele grafice sunt completate de metode numerice, cum sunt: testul Levene, testul Dickey - Fuller.

Seria de timp definită de sosirile în pensiunile turistice rurale din România, exemplificată în lucrare, s-a staţionarizat prin diferenţe simple şi sezoniere. Modalitatea potrivită de staţionarizare s-a stabilit, în final, cu ajutorul testului ADF şi s-a renunţat la transformarea Box-Cox care ar fi introdus manifestări artificiale ale datelor seriei.

Bibliografie

1. Alonso, J., H., Morales, M., *Econometria de series temporales*, Editorial Universitas, S.A., Madrid, 2000
2. Berdot, J.P., *Econometrie*, Universite de Poitiers, 2001-2002
3. Bourbonnais, R. - Terraza, M., *Analyse des series temporelles en économie*, Presses Universitaires de France, Paris, 1998
4. Bourbonnais, R., *Économétrie*, 3e edition, DUNOD, Paris, 2000
5. Gourieroux, Ch., Monfort, A., *Series temporelles et modeles dynamiques*, ediţia a doua, Ed. Economica, Paris, 1995
6. Jaba, E., *Statistica*, ediţia a treia, Ed. Economică, Bucureşti, 2002
7. Vaté, M. - *Statistique chronologique et prevision*, Ed. Economica, Paris, 1993