ESTIMAREA SIMULTANĂ A TRENDULUI ŞI FACTORILOR SEZONIERI

BOSÂNCEANU MARINICĂ
Universitatea „ Ştefan cel Mare“ Suceava
E-mail: boss@eed.usv.ro

The paper presents sequences from ***two techniques of simultaneous estimation of the trend and temporary factors****, based on the same approach, the method of least squares and with identical results.* ***The first technique*** *has the advantage that represents the object of a handed technique since it has* ***used analytical results, relatively simple****. The simplicity is a consequence a restriction over the studied chronics. It has to obligatory behaved the same number of observations for any constitutive duration of the series.*

The importance of ***the second procedure*** *consists in* ***using multiple regression programs****, referring at the variable indicators for measuring the temporality. The major inconvenient for this second procedure is represented by the necessity of using the informatics' programs relatively important, so that generalization is becoming difficult for the routine treatment, in a great number of chronics.*

Referitor la acest aspect sunt disponibile două **tehnici**, care sunt bazate pe acelaşi demers, metoda celor mai mici pătrate, şi dau rezultate identice.

A. ***Prima tehnică*** prezintă avantajul că face obiectul unui tratament manual, întrucât ***utilizează rezultate analitice relativ simple***. Simplitatea este o consecinţă a unei restricţii asupra cronicii studiate: aceasta trebuie a comporta, obligatoriu, acelaşi număr de observaţii pentru fiecare perioadă constitutivă a unui ciclu (dacă avem ciclu anul şi observaţii lunare, ciclul va trebui să aibă pentru fiecare lună acelaşi număr de ani de observare) (de ex., dacă seria temporală debutează în martie, ultima observaţie se impune a corespunde lunii februarie).

Uzanţa impune ca astfel de cronici să fie prezentate în tabele rectangulare, unde lunile sunt reperate în coloane şi anii în linii – tabele Buys-Ballot. Aşadar această primă tehnică o putem defini „ *estimarea simultană a trendului şi factorilor sezonieri, plecând de la metoda celor mai mici pătrate, pe tabelele Buys-Ballot* “.

B. Esenţa acestui traiect: ***folosirea de programe de regresie multiplă***, făcând apel la variabile indicatoare pentru a măsura sezonalitatea. Inconvenientul major al acestei secunde proceduri îl constituie necesitatea folosirii programelor informatice relativ importante, deci este dificilă generalizarea sa în tratamentul de rutină pentru un număr foarte mare de cronici. În replică, putem aminti avantajul conform căruia se poate raţiona, prin intervalul de încredere, parametrii evoluţiei de tendinţă şi factorii sezonieri, plecându-se de la desluşirile clasice furnizate de către programele de regresie, precum şi reda (nu toate programele informatice, însă, oferă această posibilitate) statistica Durbin-Watson, care permite să judecăm dacă ipoteza de independenţă a reziduurilor este sau nu acceptabilă. În acest ultim caz, estimaţiile calculate pentru diferiţii parametri sunt depărtate, ca urmare nu se poate avea încredere în rezultatele obţinute.

Înainte de a prezenta aceste tehnici, introducem **ipotezele** aferente lor.

- Existenţa, implicită, a unicităţii legilor de evoluţie, atât de tendinţă cât şi sezonieră, a fenomenului studiat. Previziunea bazată pe parametrii obţinuţi se clădeşte, de asemenea, pe această ipoteză. În legătură cu acest aspect menţionăm că demersul diferă fundamental de abordările modelului Holt-Winters, care se adaptează, fără probleme, la modificările acestor legi de evoluţie.

- Evoluția de tendință este liniară ($x_t = at + b$), dar în realitate se poate substitui lui x_t toată transformarea monotonă, cea mai folosită fiind transformarea logaritmică, aceasta permițând studierea fenomenelor cu evoluție exponențială.
- Mişcarea sezonieră c_t este riguros periodică şi se va nota c_j, $j = 1 \div m$, **m** fiind numărul de perioade dintr-un ciclu (de ex., m = 12, dacă ciclul este anul şi observațiile lunare).
- Fenomenul studiat suferă, pe lângă aceasta, o perturbație aleatoare ε_t, a cărei natură este de tip *bruit-blanc*.
- Cele trei componente funcționează aditiv, ceea ce conduce la modelul de bază următor:

$$y_t = (at + b) + c_j + \varepsilon_t \quad \text{model aditiv.} \qquad (1)$$

Acest model poate fi echivalent unui model multiplicativ, dar după transformarea logaritmică [, p.298] a unor variabile, deoarece legătura liniară afectează parametrii şi nu variabilele:

$$\ln(y_t) = at + b + c_j + \varepsilon_t, \qquad (2)$$

unde: $a = \ln(1+a')$;

$b = \ln(b')$;

$c_j = \ln(c_j')$, c_j' fiind un coeficient multiplicativ sezonier;

ε_t – perturbația aleatoare de tip *bruit-blanc*, aceasta putându-se scrie şi sub forma $\ln(\eta_t)$.

Ca atare, modelul multiplicativ va fi următorul:

$$y_t = b' \cdot c'_j \cdot \eta_t \, (1+a')^t. \qquad (3)$$

- O ipoteză suplimentară, relativă la factorii sezonieri, care este indispensabilă în abordarea tehnicii celor mai mici pătrate pe tabelele Buys-Ballot, este <u>principiul de conservare a ariilor</u>. Acest principiu nu este necesar în cazul unui demers prin regresie multiplă, fapt ce explică şi diferența de rezultat numeric dintre cele două tehnici. De remarcat totuşi că rezultatele obținute sunt perfect comparabile, aşa cum de altfel vom constata şi în ex. numeric considerat.

Această ipoteză postulează existența compensării mişcărilor sezoniere pe ciclu, respectiv:

$$\sum_{j=1}^{m} c_j = 0, \qquad (4)$$

iar în cadrul unui model multiplicativ:

$$\sum_{j=1}^{m} c_j = \sum_{j=1}^{m} \ln(c_j) = 0 \rightarrow \prod_{j=1}^{m} c_j = e^0 = 1. \qquad (5)$$

Rezultă că media geometrică a coeficienților sezonieri este egală cu **1** la model multiplicativ şi **0** la cel aditiv.

A. Estimarea simultană a trendului şi componentei sezoniere prin metoda celor mai mici pătrate pe tabel Buys-Ballot

Fie:

n – numărul de cicluri existente într-o cronică (numărul de ani, de ex.);

m – numărul de perioade constitutive ale ciclului (m = 12 luni, dacă observațiile sunt lunare şi ciclul anual);

i – indicele de reperare a ciclului;

j – indicele folosit pentru identificarea perioadei în ciclu.

Data **t**, corespunzătoare lunii **j** a anului **i**, poate fi scrisă astfel:

$$t = m(i-1) + j. \qquad (6)$$

Valoarea teoretică aferentă lui x_t:

$$\hat{x}_t = at + b + c_j = a[m(i-1)+j] + b + c_j.$$

Luând $b_j = b + c_j$, avem:

$$\hat{x}_t = a[m(i-1)+j] + b_j. \qquad (7)$$

Aplicarea metodei celor mai mici pătrate constă în căutarea valorilor celor **m** factori sezonieri b_j, precum şi cea a creşterii mediei lunare **a**. Rezultanta aplicării demersului: <u>varianța reziduală este minimă</u>.

$$\sum_t \left(x_t - \hat{x}_t \right)^2 = \sum_i \sum_j \left\{ x_{ij} - a[m(i-1)+j] + b_j \right\}^2 \to \min. \qquad (8)$$

Tabel 1. Buys-Ballot – model aditiv

An	LUNA 1	2	3	4	5	6	7	8	9	10	11	12	Σ	$\bar{X}$
1994		653	713	684	707	714	684	422	629	811	721	803	**8246**	687
1995	831	765	815	830	756	811	746	504	774	762	695	680	**8969**	747
1996	682	684	743	701	641	595	679	439	728	746	688	743	**8069**	672
1997	689	741	815	719	730	764	626	456	757	786	778	762	**8623**	719
1998	719	692	771	692	753	764	652	494	696	742	794	773	**8542**	712
1999	901	815	900	846	883	811	766	584	761	863	773	751	**9654**	805
2000	739	624	640	672	615	638	555	372	605	615	573	674	**7322**	610
Σ	**5266**	**4974**	**5397**	**5144**	**5085**	**5097**	**4708**	**3271**	**4950**	**5325**	**5022**	**5186**	**59425**	
$\bar{x}$	752	711	771	735	726	728	673	467	707	761	717	741		
Me[1]	719	692	771	701	730	764	679	456	728	762	721	751		

După efectuarea calculelor [1, p. 359-363], obținem soluția analitică următoare:

$$a = \frac{12}{mn\left(n^2-1\right)} \left(\sum_{i=1}^{n} i \sum_{j=1}^{m} \frac{x_{ij}}{m} - \frac{n+1}{2m} \sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij} \right) \qquad (9)$$

$$b = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{x_{ij}}{nm} - a\frac{nm+1}{2} \qquad (9')$$

$$c_j = \sum_{i=1}^{n} \frac{x_{ij}}{n} - \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{x_{ij}}{nm} - a\left(j - \frac{m+1}{2} \right) \qquad (9'')$$

Tabel 2. Buys-Ballot – model multiplicativ

An	LUNA 1	2	3	4	5	6	7	8	9	10	11	12	Σ
1994	6,56	6,48	6,57	6,53	6,56	6,57	6,53	6,05	6,44	6,70	6,58	6,69	**78,26**
1995	6,72	6,64	6,70	6,72	6,63	6,70	6,61	6,22	6,65	6,64	6,54	6,52	**79,29**
1996	6,53	6,53	6,61	6,55	6,46	6,39	6,52	6,08	6,59	6,61	6,53	6,61	**78,01**
1997	6,54	6,61	6,70	6,58	6,59	6,64	6,44	6,12	6,63	6,67	6,66	6,64	**78,82**
1998	6,58	6,54	6,65	6,54	6,62	6,64	6,48	6,20	6,55	6,61	6,68	6,65	**78,74**
1999	6,80	6,70	6,80	6,74	6,78	6,70	6,64	6,37	6,63	6,76	6,65	6,62	**80,19**
2000	6,61	6,44	6,46	6,51	6,42	6,46	6,32	5,92	6,41	6,42	6,35	6,51	**76,83**
Σ	**46,34**	**45,94**	**46,49**	**46,17**	**46,06**	**46,10**	**45,54**	**42,96**	**45,90**	**46,41**	**45,99**	**46,24**	**550,14**

B. Estimarea simultană a trendului şi componentei sezoniere prin regresia multiplă

Utilizare unui program de regresie necesită definirea a 11 variabile indicatoare d_j, **j** variind de la 1 la 11, dacă sezonalitatea este anuală pentru o cronică lunară. De o manieră generală vom avea m – 1 variabile indicatoare, atunci când ciclul studiat comportă **m** perioade. Aceste variabile indicatoare nu pot lua decât valoarea **0** sau **1** şi nu vom avea niciodată, în acelaşi timp, pentru o perioadă **t** dată, mai mult de o variabilă indicatoare nenulă.

Amintim că se pot utiliza variabilele indicatoare pentru a elimina perturbațiile excepționale (greve prelungite, de ex.), numărul lor fiind egal cu cel al perturbațiilor care urmează a se elimina. Fiecare variabilă va lua valoarea **1** doar pentru perioadele unde perturbația există şi **0** în celelalte cazuri. Folosirea simultană a variabilelor cantitative şi indicatoare ca variabile explicative ale regresiei multiple este cunoscută sub numele de analiza covarianței.

Revenind la modalitatea de atribuire a variabilelor indicatoare, precizăm că se convine a avea $d_1 = 1$, atunci când perioada **t** corespunde la o lună ianuarie, $d_2 = 1$, în situația în care perioada **t** este aferentă unei luni februarie s.a.m.d. Dacă perioada **t** se referă la o lună decembrie, toți d_j sunt nuli.

Ecuația de regresie este următoarea:

$$y_t = at + b + \sum_{j=1}^{11} c_j d_j. \qquad (10)$$

Având o perioadă **t**, multiplu întreg de 12, toți d_j sunt nuli, deci constanta **b** are semnificație de constantă a trendului clasic modulat prin componenta sezonieră a lunii decembrie. În cazul celorlalte perioade **t**, doar un singur d_j nu este nul, coeficientul c_j venind să moduluze **b**, deci termenul lunii decembrie. După cum se constată, referința implicită a lunii decembrie a redat inutil utilizarea principiului conservării ariilor. Cu toate acestea, dacă s-ar fi introdus în ecuația de regresie un $c_{12}d_{12}$, o astfel de ipoteză ar fi fost redată necondiționat, cu scopul de a avea o constantă de referință în vederea calculării c_j, **b** luând imediat valoarea acestei constante. Introducerea termenului $c_{12}d_{12}$ nu ar duce însă la nici un rezultat, dimpotrivă, nedeterminarea care ar rezulta (problema constantei de referință) s-ar traduce matematic prin legătura liniară $\sum_{j=1}^{12} d_j = 1$.

În vederea verificării plauzibilității ipotezei, conform căreia reziduurile sunt pur aleatoare, se utilizează, în mod clasic, testul Durbin-Watson. Acesta nu testează decât independența reziduurilor succesive, fiind, asimptotic, echivalentul unui test asupra coeficientului de autocorelație de ordin 1.

$$d = \frac{\sum_{t=2}^{n} e_t^2 + \sum_{t=2}^{n} e_{t-1}^2 - 2\sum_{t=1}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2} \cong 2(1-\rho_1). \qquad (11)$$

Testul presupune că variabilele explicative sunt certe, el calculându-se plecând de la **n** reziduuri e_t ale regresiei multiple. Indicatorul statistic, notat adesea DW, va fi:

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}, \qquad (12)$$

urmând a se compara cu valorile d_U şi d_L, în scopul de a accepta sau respinge ipoteza de independență. Raționamentul se foloseşte de fig.1, unde zonele haşurate corespund plajelor de valori care nu permit o concluzie rezonabilă.

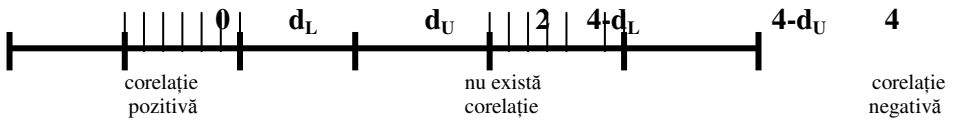

Figura 1

Din păcate, tabele Durbin-Watson există doar pentru $k \leq 5$ (variabile explicative). Aproximări desigur avem, dar punerea lor în operă necesită calcule laborioase, deşi nu la fel de complicate ca în cazul utilizării tabelelor Durbin-Watson. În acest sens pot fi amintite aprox. printr-o lege Student [2, p. 400-402] sau cele cu o lege Beta [1, p.166-169]. În aceste condiții se va utiliza un test pe coeficientul de autocorelație de ordin 1, deşi acest test nu este la fel de edificator ca cel al lui Durbin-Watson.

Bibliografie:

1. Calot, G. – „ Cours de statistique descriptive “, Ediția III, Dunod, 1993.
2. Fuller, A.T. – „ Introduction to statistical time series “, Wiley, 1976.
3. Giard, V. – „ Gestion de la production “, Economica, Ediția III, Paris, 2001.
4. Giard, V. – „ Statistique appliquée à la gestion “, Economica, Ediția V, Paris, 1987.
5. Jaba, E. – „ Statistică “, Ed. Economică, Ediția II, Bucureşti, 2000.