TÉNYI GYÖRGY

Az adatsorozatok spektrálelemzésének általánosítása és alkalmazása

Nem újkeletű az a felismerés, hogy a nyers statisztikai adatsorozatok gyakran kiigazításra szorulnak, mert az adatsorozatokban rejlő többféle tendenciát, (nem mindig szigorúan) periodikus ingadozást el kell különítenünk egymástól, hogy jobban értékelhessük a szóban forgó adatsorozat jövőbeli alakulására gyakorolt hatásukat. Például a szezonális ingadozások olyan erősek lehetnek, hogy „elfedik" az összes többi tendenciát. Minthogy a szezonális ingadozások a legszembetűnőbbek, és ezek okozzák a legtöbb nehézséget, a „hagyományos", empirikus kiigazítási eljárások mind a szezonális ingadozásoknak a kiküszöbölésére szolgáltak. Kezdetben a spektrálelemzést is elsősorban úgy fogták fel, mint a szezonális ingadozások vizsgálatának egy sztochasztikusan megalapozott módszerét.

A spektrálelemzés hatósugara azonban — legalábbis elméleti erejét tekintve — jóval nagyobb ennél. Cikkünk célja, hogy nagy vonalakban ismertesse a spektrálanalízis néhány olyan alkalmazási területét, amelyről — úgy tűnik — nem esett még szó a magyar irodalomban. A problémakör egzakt matematikai tárgyalása iránt érdeklődő olvasónak GRENANDER és ROSENBLATT [1] könyvét ajánljuk figyelmébe. Könnyen érthető, bár kevésbé pontos — és a nehezebb, hosszadalmasabb levezetéseket mellőző — tárgyalás található GRANGER és HATANAKA [2] monográfiájában, ebben viszont a szerzők nagy teret szentelnek a módszer közgazdasági alkalmazásaival kapcsolatos problémáknak is.

A spektrálelemzés alapjai ugyan hozzáférhetők magyarul [3], [4], [5], mégis a teljesség kedvéért itt is röviden összefoglaljuk őket.

Sztochasztikus folyamatnak nevezzük valószínűségi változók egy $\{x(t, w)\}$ összességét, ahol a t időparaméter valamely (t_1, t_2) intervallumban veszi föl értékeit, w pedig az eseménytéren fut végig. Egy ilyen folyamatnak lényeges jellemzője az $M(t)$ várható értéke és a $\sigma(t)$ szórása, továbbá az, hogy a különböző időpontokban felvett értékei között milyen szoros kapcsolat van. Ennek a kapcsolatnak a szorossága az alábbi ún. autokovariancia-függvénnyel definiálható

$$r(t, \tau) = M\left[\big(x(t, w) - M(t)\big)\big(x(t+\tau, w) - M(t+\tau)\big)\right].$$

Ha mármost két időpontra, t-re és $t+\tau$-ra vonatkozóan rendelkezésünkre áll egy-egy N elemű, független megfigyelésekből álló minta (legyenek ezek $x_i(t)$, ill. $x_i(t+\tau)$; $i = 1, \ldots, N$) akkor $r(t, \tau)$-ra a következő becslést írhatjuk föl:

$$\hat{r}(t,\tau) = \frac{1}{N}\sum_{i=1}^{N}[x_i(t) - M(t)]\,[x_i(t+\tau) - M(t+\tau)]\,.$$

Vagyis az átlagolást úgy kellene elvégezni, hogy t és τ értékét rögzítve, a kívánatos számú mintára átlagolnánk. Csakhogy a gazdasági adatsorozatok elemzésénél éppen az okozza az egyik fő problémát, hogy adott t értékhez igen sokszor csak egyetlen adatot tudunk megállapítani, mert a folyamatnak csak egyetlen megvalósulása létezik, vagy legalábbis csak egyet ismerünk. Pl. egy bizonyos évben adott országra vonatkozóan csak egy adatunk van a nemzeti jövedelemre. Viszont hosszabb-rövidebb *időszakra* vonatkozóan vannak adataink. Az ilyen adatsorozatot szokták — időbeli jellegére utalva — idősornak nevezni. Az idősort a mögötte rejlő sztochasztikus folyamat egyik realizációjának, magát a folyamatot pedig az adatsorozat generálójának nevezzük. Az $\{x(t, w)\}$ sztochasztikus folyamat generálta adatsorozatot $\{x(t);\, t = 1, \ldots, n\}$-nel jelöljük.

Az elmondottak miatt egyelőre csak azokkal az adatsorozatokkal fogunk foglalkozni, amelyeknél az *idő szerinti* átlagból tudunk következtetni az $r\,(t, \tau)$ értékére.

Ergodikusnak nevezünk egy $\{x(t, w)\}$ sztochasztikus folyamatot, ha négyzetes átlagban konvergens, azaz ha $t_1, t_2, \ldots \to t$, akkor

$$\lim_{n\to\infty} M[(x(t,w) - x(t_n, w))^2] = 0\,.$$

Ismeretes (lásd pl. [6]-ot), hogy az idő szerinti és a mintasokaságra vonatkozó átlagolás felcserélhető az olyan ergodikus folyamatokra, amelyek kielégítik a következő feltételeket:

$$M[x(t,w)] = m,$$

$$M[(x(t,w) - m)^2] = \sigma^2$$

és

$$M[(x(t,w) - m)\,(x(t+\tau, w) - m)] = r(\tau),$$

tehát a várható érték, a szórás és az autokovariancia *független* az időtől. Az ilyen folyamatot gyengén vagy másodrendben stacionáriusnak szokás nevezni. Ekkor tehát $r(\tau)$ becslésére az alábbi formulát alkalmazhatjuk:

(1) $$\hat{r}(\tau) = \frac{1}{n-\tau}\sum_{t=1}^{n-\tau}[x(t) - m]\,[x(t+\tau) - m]\,.$$

Az eddigiekben előlegeztük, hogy az autokovariancia-függvénynek fontos szerepe van az adatsorozatok elemzésében. Ahhoz azonban, hogy rátérhessünk arra, miben is áll ez a szerep, be kell vezetnünk néhány új fogalmat.

Mielőtt valóságos, tapasztalati adatsorozatokkal kezdenénk foglalkozni, vegyük a következő „mesterséges" sztochasztikus folyamatot:

$$x(t,w) = \sum_{j=1}^{k} w_j \cos\omega_j t$$

ahol ω_j-k valós számok, w_j-k pedig független valószínűségi változók ($j = 1, 2, \ldots k$), továbbá

$$M[w_j] = 0, \qquad M[w_j^2] = \sigma_j^2 \qquad j = 1, 2, \ldots, k$$

és

$$M[w_i w_j] = 0, \qquad \text{ha } j \neq i$$

Az $x(t, w)$ sztochasztikus folyamat minden egyes $w_j \cos\omega_j t$ alakú tagja egy $\frac{\omega_j}{2\pi}$ frekvenciájú, azaz $\frac{2\pi}{\omega_j}$ periódusú és w_j – tehát véletlen – amplitudójú periodikus függvény.

Látható, hogy

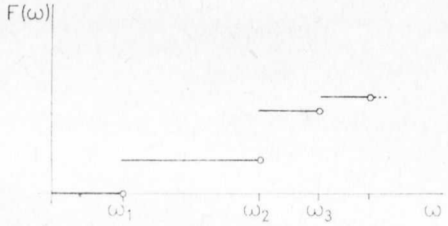
$$M[x(t, w)] = 0,$$

$$r(\tau) = M[x(t, w)\, x(t+\tau, w)] = \sum_{j=1}^{k} \cos \omega_j t \cdot \sigma_j^2$$

és ez utóbbi így is írható:

$$r(\tau) = \int_0^\pi \cos \tau\omega \, dF(\omega), \tag{2}$$

ahol $F(0) = 0$ és innen kezdve $F(\omega)$ lépcsőzetesen növekszik az előre rögzített ω_j helyeken σ_j^2 értékkel, amíg el nem éri a $\sum_{j=1}^{k} \sigma_j^2$ értéket. (Lásd az 1. ábrát.)



1. ábra.

Az $F(\omega)$ függvényt a folyamat ,,spektrális eloszlásfüggvényének" nevezzük. Vegyük észre, hogy a (2) egyenlet összefüggést állít fel a szóban forgó speciális folyamat esetében az $r(\tau)$ és az $F(\omega)$ függvény között.

Mármost, ha sem k értékét, sem pedig az w_j valószínűségi változókat nem rögzítjük előre, akkor igen sokféle adatsorozatot generálhatunk, sőt [7] alapján tudjuk, hogy ha az $F(\omega)$ függvényről nem kötjük ki, hogy lépcsős függvény legyen, hanem csak annyit kívánunk meg, hogy $\frac{F(\omega)}{\Sigma \sigma_j^2}$ eloszlásfüggvény legyen – azaz monoton nem-csökkenő, $F(0) = 0$, $F(\pi) = \Sigma \sigma_j^2$ – akkor igaz, hogy *minden* $x(t, w)$ gyengén stacionárius sztochasztikus folyamathoz található olyan – a fenti osztályba tartozó – $F(\omega)$ függvény, hogy

$$r(\tau) = 2 \int_0^\pi \cos \omega \, dF(\omega) \tag{3}$$

továbbá találhatók olyan $u(\omega)$ és $v(\omega)$ valószínűségi változók, hogy

$$x(t, w) = 2 \int_0^\pi \cos t\,\omega\, du(\omega) + 2 \int_0^\pi \sin t\,\omega\, dv(\omega)\,; \tag{4}$$

ahol

$$\begin{aligned} & M[(u(\omega_1 + \Delta\omega) - u(\omega_1))(u(\omega_2 + \Delta\omega) - u(\omega_2))] = \\ & = M[(v(\omega_1 + \Delta\omega) - v(\omega_1))(v(\omega_2 + \Delta\omega) - v(\omega_2))] = \\ & = \begin{cases} 0, \text{ ha } (\omega_1 + \Delta\omega, \omega_1) \text{ és } (\omega_2 + \Delta\omega, \omega_2) \text{ diszjunkt;} \\ F(\omega + \Delta\omega) - F(\omega), \text{ ha } \omega_1 = \omega_2 = \omega\,; \end{cases} \end{aligned} \tag{5a}$$

illetve

$$M[(u(\omega_1 + \Delta\omega) - u(\omega_1))(v(\omega_2 + \Delta\omega) - v(\omega_2))] = 0 \tag{5b}$$

$$0 \leq \omega_1, \omega_2 \leq \pi.$$

Vagyis minden gyengén stacionárius folyamat végtelen sok frekvenciából összetevődöttnek fogható fel, amelyeknek amplitúdója a (5a) és (5b) tulajdonságokkal rendelkező — úgynevezett ortogonális — valószínűségi változó és minden egyes frekvencia két — egymáshoz képest $\pi/2$ fáziskülönbségű — komponensre bontható.

A továbbiakban olyan sztochasztikus folyamatokkal foglalkozunk, amelyeknek létezik a

$$\frac{dF(\omega)}{d\omega} = f(\omega)$$

egyenlettel definiált ún. *spektrális sűrűségfüggvényük* (röviden spektrumuk), ahol $f(\omega)$ abszolút folytonos függvény.

Az ilyen folyamatokra a (3) egyenlet a következő alakot ölti:

$$r(\tau) = 2 \int_0^\pi \cos \tau\omega\, f(\omega)\, d\omega. \tag{3'}$$

Mármost az $f(\omega)$ függvényt jól felhasználhatjuk a sztochasztikus folyamatok vizsgálatára. Ha ugyanis egy adatsorozatban periodikus ingadozások vannak, akkor az őt generáló sztochasztikus folyamat $f(\omega)$ spektruma a megfelelő $\omega_1, \omega_2, \ldots$ frekvenciáknál a közvetlen környezethez képest többé-kevésbé kiugró csúcsértékeket vesz föl, az ingadozás erősségétől függően.

A Fourier-transzformáció elmélete szerint (3') formális inverziójával a következőt kapjuk:

$$f(\omega) = \frac{1}{2\pi}\left[r(0) + 2 \sum_{j=1}^{\infty} r(j) \cos j\omega\right] \tag{6}$$

Ha most az adatsorozat rendelkezésünkre álló elemei $x_1, x_2, \ldots, x_n$, akkor ezekből (1) alapján elkészítve az $\hat{r}(\tau)$ becsléseket és (6)-ba behelyettesítve, az

$$\hat{f}(\omega) = \frac{1}{2\pi}\left[\hat{r}(0) + 2 \sum_{j=1}^{n-1} \hat{r}(j) \cos j\omega\right] \tag{7}$$

becslés adódik. Világos, hogy véges sok adat alapján nem lehet ω minden értékére becslést adni, hanem $f(\omega)$-nak csak alkalmasan megválasztott intervallumokra vonatkozó átlagait becsülhetjük. A (7) formulával azonban van egy további nehézség is. Ismeretes ugyanis (lásd pl. [8] 52. és kk.) hogy [7] torzítatlan becslése ugyan [6]-nak, de nem konzisztens (tehát várható értéke $f(\omega)$-val egyenlő, de szórása nem tart zérushoz, ha $n \to \infty$).

Ez a nehézség vezetett ahhoz, hogy [7] helyett a következő alakú becsléseket kezdjék vizsgálni:

$$\hat{f}(\omega) = \frac{1}{2\pi}\left[\hat{r}(0)\,\lambda_0(\omega) + 2\sum_{j=1}^{m-1} \lambda_j(\omega)\, r(j) \cos \omega j\right], \tag{7'}$$

ahol $m < n$.

A $\{\lambda_i(\omega),\ i = 0, 1, \ldots, m-1\}$ súlyfüggvények alkalmas megválasztásán múlik a becslés jósága. Leggyakrabban a Tukey—Hanning-féle súlyfüggvényeket $\left\{\lambda_i(\omega) = 1 + \cos\frac{\pi i}{m}\right\}$ alkalmazzák. Ha ezt a súlyfüggvényt írjuk (7')-be, a kapott becslés nem lesz ugyan torzítatlan, de a torzítás megbecsülhető, és a legtöbb esetben elhanyagolható.

Ha mármost a spektrumot valamilyen (ω_1, ω_2) intervallumban akarjuk behatóbban tanulmányozni, vagy ellenkezőleg, ki akarunk küszöbölni valamely frekvenciát vagy frekvenciasávot, amely zavarja a lappangó tendenciák felismerését, ezt az ún. szűrők segítségével tehetjük meg.

Szűrőnek nevezzük az adatsorozatok következő alakú lineáris transzformációit:

$$y(t) = L[x(t)] = \sum_{-m}^{m} a_j x_{t+j}, \text{ ahol } a_j \text{ valós szám.}$$

Bebizonyítható, hogy az $y(t)$ adatsorozat spektruma

$$f_y(\omega) = f_x(\omega)\, g^2(\omega),$$

ahol

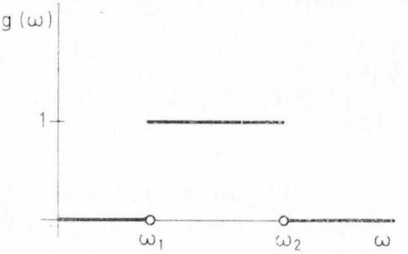
$$g(\omega) = a_0 + 2\sum_{j=1}^{m} a_j \cos \omega j\,,$$

$g(\omega)$-t nevezzük transzferfüggvénynek. Ha mármost valamilyen (ω_1, ω_2) intervallumot akarunk részletesen megvizsgálni, olyan szűrőre volna szükségünk, amelynek transzferfüggvénye

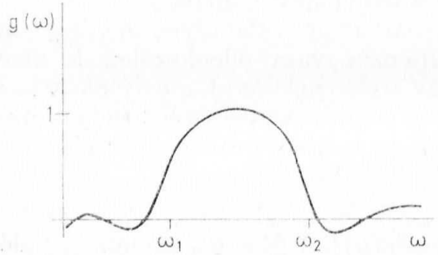
$$g(\omega) = \begin{cases} 1, & \text{ha } \omega_1 \leq \omega \leq \omega_2 \\ 0 & \text{egyébként} \end{cases}$$

(lásd a 2. ábrát). Mivel azonban ilyen transzferfüggvénnyel bíró rövid szűrő nincs, olyan szűrőkkel kell megelégednünk, amelyeknek transzferfüggvénye a 3. ábrán látható. Ekkor az okoz problémát, hogy ha az (ω_1, ω_2) intervallum közelében a spektrumnak kiugró csúcsa van, a szűrő ezt nem küszöböli ki teljesen, hiszen transzferfüggvényének értéke ott nem azonosan nulla. Az ebből eredő „szivárgást" figyelembe kell venni.

Mivel célunk itt csupán az eljárás elvi vázának az ismertetése, nem térünk ki a becslésekkel kapcsolatos többi kérdésre (konfidencia-intervallumok stb.). Annyi azonban az eddigiek alapján is világos, hogy megbízható becsléseket csak olyan nagyszámú adat birtokában készíthetünk, amennyi a gazdasági adatosorzatok esetében nem mindig áll rendelkezésünkre. Ha pl. éves adatokat



2. ábra.



3. ábra.

kell vizsgálnunk, kevés olyan adatsorozat létezik, amely 60–100 éven keresztül birtokunkban van, és jó közelítéssel stacionáriusnak nevezhető. Elsősorban tehát a sűrűbb adatsorokra lehet ebből a szempontból számítani.

Az adatsorozatok közötti összefüggések

A spektrálelemzés nemcsak egyetlen adatsorozat vizsgálatára alkalmas, hanem arra is, hogy két adatsorozat összefüggésének „finomszerkezetét" felderítsük vele. A spektrálelemzés fogalmainak ehhez szükséges általánosítását Granger [2] tárgyalásmódját követve a következőkben adjuk meg:

Legyen $x(t, w)$ és $y(t, z)$ két gyengén stacionárius sztochasztikus folyamat, ahol t az időparaméter, w és z pedig valószínűségi változó, és legyen ezenkívül az is igaz, hogy ha m_x, ill. m_y jelöli az $x(t, w)$, ill. az $y(t, z)$ folyamat várható értékét, akkor

$$M\left[\left(x(t, w) - m_x\right)\left(y(t + \tau, z) - m_y\right)\right] = r_{xy}(\tau)\,, \tag{8}$$

tehát a kovariancia szintén csak a szóban forgó időpontok *különbségétől* függjön.

Mint említettük, a „gyengén stacionárius" tulajdonság voltaképpen azt jelenti, hogy az adatsorozatok generáló folyamata mögött rejlő szabályszerűségek nem változnak a vizsgált időtartam folyamán; a (8) feltétel ehhez lényegében azt teszi hozzá, hogy a két idősor közötti összefüggés is változatlan.

Fölírva $x(t, w)$ és $y(t, z)$ Cramér-féle előállítását:

$$x(t, w) = \int_0^\pi \cos t\omega \, du_x(\omega) + \int_0^\pi \sin t\omega \, dv_x(\omega)$$

$$y(t, w) = \int_0^\pi \cos t\omega \, du_y(\omega) + \int_0^\pi \sin t\omega \, dv_y(\omega) ;$$

ebben az esetben érvényes a következő:

$$r_{xy}(\tau) = \int_0^\pi \cos(\tau\omega)\, c(\omega)\, d\omega + \int_0^\pi \sin(\tau\omega)\, q(\omega)\, d\omega , \tag{9}$$

ahol

$$M\,[du_x(\omega_1)\, du_y(\omega_2)] = M\,[dv_x(\omega_1)\, dv_y(\omega_1)] =$$

$$= \begin{cases} 0, \text{ ha } \omega_1 \neq \omega_2 \\ 2c(\omega)\, d\omega, \text{ ha } \omega_1 = \omega_2 = \omega , \end{cases} \tag{10a}$$

$$M\,[du_x(\omega_1)\, dv_y(\omega_2)] = -\,M\,[du_y(\omega_1)\, dv_x(\omega_2)] =$$

$$= \begin{cases} 0, \text{ ha } \omega_1 \neq \omega_2 \\ 2q(\omega)\, d\omega, \text{ ha } \omega_1 = \omega_2 = \omega . \end{cases} \tag{10b}$$

A (9), (10a) és (10b) egyenletek a következőképpen értelmezhetők: a $c(\omega)$ függvény a két folyamat azonos fázisban levő, ω frekvenciájú komponensének a kapcsolatára jellemző, $q(\omega)$ a $\pi/2$ fáziskülönbségű komponensek összefüggését méri; a különböző frekvenciájú komponensek között pedig nincs korreláció. Tehát pl. ha valamilyen adott $x(w, t)$ és $y(z, t)$ esetében $c(\omega_0) \neq 0$ és $q(\omega_0) = 0$, ez azt jelenti, hogy a két folyamat ω_0 frekvenciájú komponense között van korreláció, és hogy pontosan azonos fázisban vannak. Ha $q(\omega_0)$ sem zérus, akkor a két folyamat ω_0 frekvenciájú komponense között valamekkora fáziskülönbség is van. Vagyis két stacionárius folyamat összefüggése *kimerítően* leírható úgy is, hogy minden frekvenciára megadjuk, az azonos frekvenciájú komponensek közötti összefüggést.

Az $x(t, w)$ és $y(t, z)$ sztochasztikus folyamat közötti összefüggést a

$$C(\omega) = \frac{c^2(\omega) + q^2(\omega)}{f_x(\omega) + f_y(\omega)}$$

ún. koherencia-függvénnyel fejezzük ki. Bebizonyítható, hogy

$$0 \leq C(\omega) \leq 1 .$$

Minél közelebb van $C(\omega)$ értéke 1-hez, annál erősebb a korreláció az $x(t, w)$ és $y(t, z)$ folyamat ω-frekvenciájú komponense között.

Ha a két komponens közötti fáziskülönbséget is meg akarjuk állapítani, a

$$\varphi(\omega) = \text{arc tg} \left(\frac{q(\omega)}{c(\omega)} \right)$$

függvényt kell meghatároznunk. Természetesen $\varphi(\omega)$-nak csak olyan ω-értékek esetében van jelentősége, amelyekre $C(\omega)$ elég nagy, hiszen ha két azonos frekvenciájú komponens között nincs számottevő korreláció, a fáziskülönbségük nem érdekes.

Amennyiben tehát módunkban áll $C(\omega)$-t és $\varphi(\omega)$-t becsülni valamely $x(t, w)$ és $y(t, z)$ folyamatra vonatkozólag, akkor — ha nem is minden egyes frekvenciára, hiszen csak véges sok adatunk van — de legalábbis bizonyos frekvenciasávokra külön-külön meg tudjuk becsülni a két adatsor közötti korrelációt. Ez azért fontos, mert semmilyen biztosíték nincs arra, hogy pl. két gazdasági adatsorozat esetében akár a korreláció, akár pedig a fáziskülönbség ugyanakkora legyen a két adatsorozat különböző periódusú komponensei között. Pl. a sajtgyárak kibocsájtásának hosszútávú ingadozása valószínűleg szorosabb kapcsolatban áll a tejtermelés hasonló ingadozásaival, mint ugyancsak a sajtgyártás heti ingadozása a tejtermelés heti ingadozásával.

Röviden érintjük még a koherencia és bázisfüggvény becslésének problémáját. Ha a rendelkezésre álló adatok:

$$\{x(t),\quad y(t);\quad t=0,1,\ldots,n\},\quad \text{akkor az}$$

$\omega_j = \dfrac{\pi j}{m}$ (ahol $m < n$ és $j = 1, \ldots, m$) pontokban a következőképpen kaphatunk becslést:

$$\hat{C}(\omega_j) = \frac{\hat{c}^2(\omega_j) + \hat{q}^2(\omega_j)}{\hat{f}_x(\omega_j) + \hat{f}_y(\omega_j)}, \tag{11}$$

ahol

$$\hat{c}(\omega_j) = \frac{1}{\pi}\left[\frac{a_0\lambda_0(\omega_j)}{2} + \sum_{k=1}^{m-1} \lambda_k(\omega_j)\, a_j \cos k\,\omega_j\right], \tag{12}$$

és

$$\hat{q}(\omega_j) = \frac{1}{\pi}\left[\sum_{k=1}^{m-1} \lambda_k(\omega_j)\, a_k \sin k\,\omega_j\right], \qquad \text{ahol} \tag{13}$$

$$a_k = \frac{1}{n-k}\sum_{j=1}^{n-k}[x(j)\,y(j-k) + y(j)\,x(j-k)] \qquad j, k = 1, \ldots, m$$

Mindaz, amit ebben a részben elmondtunk, értelemszerűen alkalmazható kettőnél több adatsorozat esetében is.

Nem-stacionárius adatsorozatok

Idáig stacionárius adatsorozatokkal foglalkoztunk. Sajnos azonban a legtöbb gazdasági adatsorozatról nem tételezhetjük fel, hogy stacionárius. Még a legegyszerűbb eset az, ha ez a nem-stacionárius jelleg csupán abban nyilvánul meg, hogy az idősorban valamilyen trend van, és ha a trendet az adatsorozatból kiszűrve, már stacionárius sorozatot kapunk. Ekkor ugyanis több módszer áll rendelkezésünkre (lásd [2]-t) ennek a trendnek a becslésére.

Vegyük viszont a következő folyamat-sorozatot:

$$x_0(t, z) = \int_0^\pi \cos t\omega \, du(\omega) + \int_0^\pi \sin t\omega \, dv(\omega) ,$$

ahol $u(\omega)$ és $v(\omega)$ olyan valószínűségi változó, amelyre érvényes az (5a), ill. (5b) összefüggés, valamint

$$F(\omega + \Delta\omega) - F(\omega) = C$$

továbbá

$$x_k(t, z) = \int_0^\pi \cos t\omega \, a(k, \omega) \, du(\omega) + \int_0^\pi \sin t\omega \, a(k, \omega) \, dv(\omega) , \tag{14}$$

$k = 1, 2, \ldots,$; ahol $a(k, \omega)$ minden k-ra ω-ban folytonos függvény.

Belátható, hogy az $x_k(t, z)$ folyamat spektruma

$C.a^2(k, \omega)$-tel egyenlő, és hogy ha

$M\,[x_0(t, z)] = 0$, akkor $M\,[x_k(t, z)] = 0$.

Legyen most

$$y(t, z) = x_t(t, z) . \tag{15}$$

Az $y(t, z)$ folyamat nyilván nem stacionárius, hiszen

$$M\,[y(t, z)\, y(t + \tau, z)] = C \int_0^\pi \cos \tau\omega \, a(t, \omega) \, a(t + \tau, \omega) \, d\omega ,$$

ez pedig általában függ t-től. Mármost mivel a spektrum fogalmát stacionárius stochasztikus folyamatokra értelmeztük, az

$$f_y(t, \omega) = C \cdot a^2(t, \omega)$$

függvény nem tekinthető spektrumnak, de megőriz néhányat a spektrum tulajdonságaiból. Pl. ha F olyan szűrő, amelynek transzferfüggvénye

$$g(\omega) = \begin{cases} 0, & \text{ha} \quad \omega \neq \omega_0 \\ 1, & \text{ha} \quad \omega = \omega_0 \end{cases}$$

akkor belátható, hogy F-et $y(t, z)$-re alkalmazva, az

$$Y(t, z) = F\,[y(t, z)]$$

folyamat csak az ω_0 frekvenciát fogja tartalmazni, és az $f_y(t, \omega_0)$ függvény azt mutatja, hogyan változik az ω_0 frekvenciához tartozó amplitúdó az idő függvényében, ami pedig a stacionárius esettel analóg. Fontosabb azonban a következő tulajdonság:

Legyen

$$\hat{r}(\tau) = \frac{1}{n - \tau} \sum_{t=1}^{n-\tau} y(t)\, y(t + \tau) ;$$

belátható, hogy ha $n \gg \tau$, $a(t, \omega)$ pedig lassan változik az időben, akkor

$$M[\hat{r}(\tau)] \approx \int_0^\pi \cos\omega\tau \left[\frac{1}{n} \sum_{t=1}^n a^2(t, \omega)\right] d\omega\,;$$

azaz, ha az adatosorzatból úgy becsüljük az autokorrelációt, mintha a sorozat stacionárius volna, és ezt az $\hat{r}(\tau)$-t írjuk be a (7') becslésbe, akkor valójában az

$$\frac{1}{n} \sum_{t=1}^n f(t, \omega)$$

átlagot becsüljük. Ami a becslés megbízhatóságához szükséges feltételek pontosabb megfogalmazását illeti, Granger [2] könyvének 9. fejezetére utalunk. Durván annyit mondhatunk, hogy némi információ nyerhető az itt leírt eljárással nem-stacionárius adatsorozatokra is, mégpedig annál több, minél lassúbb a spektrum változása a vizsgált időszakban.

Ha mármost két nem-stacionárius adatsorozat összefüggését vizsgáljuk, figyelembe kell vennünk azt is, hogy maga ez az összefüggés sem lesz általában már független az időtől. Bizonyos összefüggések felírhatók abban az általános esetben is, amikor mind a koherenciafüggvény, mind pedig a fázisfüggvény függ az időtől; gyakorlatilag azonban nem látszanak alkalmazhatónak. Ha azonban legalább az a feltevés jogosult, hogy a fázisfüggvény nem függ az időtől, akkor némileg a koherenciafüggvény is egyszerűsödik. Ugyanis, ha $y_1(t)$ és $y_2(t)$ a (14) és (15) egyenlet szerinti nem-stacionárius adatsorozat, akkor a (12), illetve (13) becslés formális alkalmazásával valójában az

$$\frac{1}{n} \sum_{t=1}^n c(\omega, t)$$

és

$$\frac{1}{n} \sum_{t=1}^n q(\omega, t)$$

átlagokra kapunk becslést, ebből pedig — a fázisfüggvényre vonatkozó előbbi feltevésünket kihasználva — a következőt kaphatjuk (a levezetést mellőzve):

$$\frac{1}{n} \sum_{t=1}^n C(\omega, t) \approx \frac{\left[\sum\limits_{t=1}^n r(\omega, t)\right]^2}{\left[\sum\limits_{t=1}^n f_{y1}(\omega, t)\right]\left[\sum\limits_{t=1}^n f_{y2}(\omega, t)\right]}$$

vagyis a koherenciafüggvények átlagára adódik becslés; a fázisfüggvényre a stacionárius esetben megadott becslés pedig most is a fázisfüggvény konzisztens, torzítatlan becslése lesz. Így tehát nem-stacionárius adatsorozatok összefüggéséről is kaphatunk információkat, ha a spektrumok változása lassú.

Alkalmazási lehetőségek

Az irodalomban bemutatott alkalmazások közül kettőt emelünk ki, mint amelyek jól illusztrálják a spektrálelemzés lehetőségeit.

GODFREY és KARREMAN [9] azt vizsgálva, milyen tényleges hatást gyakorolnak a szokásos szezonális kiigazítási eljárások az adatsorozatokra, a következő kísérletet végezték: Elektronikus számítógépen mesterségesen generáltak másodrendű autoregresszív sémával egy alapadat-sorozatot, továbbá olyan sorozatokat, amelyek változó — mégpedig más és másképpen változó — szezonális ingadozást képviseltek. Az alapadat-sorozat tagjaihoz rendre hozzáadták valamelyik ingadozássorozat megfelelő tagját, az így kapott adatsorozatot pedig valamelyik vizsgált kiigazítási eljárásnak vetették alá.

Mármost több sorozat állt rendelkezésükre: a nyers adatsorozat és két komponense — tudniillik az alapadat-sorozat és az ingadozás-sorozat — ezen felül a szóban forgó eljárással kiigazított sorozat és az elkülönített szezonális ingadozás. Ezután a koherencia-, illetve fázisfüggvény becslésével meghatározták a tényleges és a számított szezonális ingadozás stb. között az összefüggést. Így bizonyos következtetésekre lehetett jutni arra vonatkozóan, milyen hatékonyan szűrik ki az egyes kiigazítási eljárások a különbözőképpen változó szezonalitást az adatsorozatból, illetve nem okoznak-e torzításokat.

Egy másik érdekes lehetőségre — tudniillik a sztochasztikus ökonometriai modellek vizsgálatára — Granger hívta fel a figyelmet [10]-ben.

Vegyük a következő egyszerű ökonometriai modellt:

$$C_t = cY_{t-1} + \varepsilon'_t$$

$$I_t = v(Y_t - Y_{t-1}) + \eta'_t$$

$$Y_t = C_t + I_t,$$

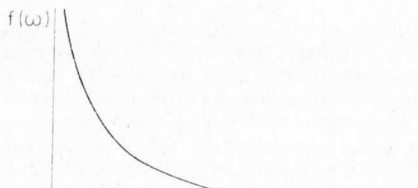
ahol ε'_t és η'_t zérus várható értékű, konstans spektrumú sztochasztikus folyamat („fehér zaj") C_t a nemzeti jövedelemből fogyasztásra fordítható rész, I_t a felhalmozás, Y_t a teljes nemzeti jövedelem. Átrendezve:

$$Y_t - \alpha Y_{t-1} = \varepsilon_t$$

$$C_t - \alpha C_{t-1} = \eta_t$$

$$I_t - \alpha I_{t-1} = \varepsilon_t - \eta_t,$$

ahol $\alpha = \dfrac{v - c}{v - 1}$, ε_t-ről és η_t-ről pedig belátható, hogy mindkettő szintén „fehér zaj". Mármost tehát mindhárom adatsorozat (Y_t, C_t, I_t) felfogható, mint ugyanannak a szűrőnek a kimenő jele, csak a bemenő jel más és más. Belátható, hogy ebben az esetben az egyes adatsorozatok spektruma a következő:



4. ábra.

$$f_Y(\omega) = \frac{f_\varepsilon(\omega)}{2(1 - \alpha \cos \omega)},$$

$$f_C(\omega) = \frac{f_\eta(\omega)}{2(1 - \alpha \cos \omega)},$$

$$f_I(\omega) = \frac{f_\varepsilon(\omega)}{2(1 - \alpha \cos \omega)},$$

Ha α közel van 1-hez, akkor a nevezők igen kicsinyek lesznek, ahogy $\omega \to 0$, ezért az elméleti spektrumok alakja a 4. ábrán látható görbéhez hasonlít. Amennyiben a tényleges adatok alapján becsült spektrumok lefutása valóban ilyen, ez alátámasztja a felírt modell helyességét.

(Beérkezett: 1969. VII. 7.)

IRODALOM

[1] GRENANDER, U.—ROSENBLATT, M.: Statistical Analysis of Stationary Time Series, New York, 1957.

[2] GRANGER, C. W. J.—HATANAKA, M.: Spectral Analysis of Economic Time Series, Princeton University Press, 1964.

[3] ÁCS, M.: Idősorelemzés és tervezés, OT. Tervgazdasági Intézet, 1968. (Sokszorosított anyag.)

[4] HRUBOS, I., PAJZS, J.: A szezonális kiigazítási eljárások összehasonlítása, Ökonometriai füzetek, 9. sz., 1968.

[5] THEISS, E.: Idényszerű változások mérése spektrálelemzés segítségével, 1966. (Kézirat.)

[6] WIENER, N.: The Extrapolation, Interpolation and Smoothing of Stationary Times Series, New York, 1949.

[7] KOLMOGOROFF, S.: Sur l'interpolation et extrapolation des suites stationnaires, C. R. Acad. Sci. Paris, 1939, 2043—2045. p.

[8] HANNAN, E. J.: Time Series Analysis, London, 1960.

[9] GODFREY, M. D.—KARREMAN, H. F.: A Spectrum Analysis of Seasonal Adjustement, az „Essays in Mathematical Economics in Honor of Oskar Morgenstern" c. kötetben. (Ed. M. Shubik), 367—421. p.

[10] GRANGER, C. W. J.: Spectral Shape of an Economic Variable, Econometrica, 1966, 34. 150—161. p.

THE GENERALIZATION AND APPLICATION OF THE SPECTRAL ANALYSIS OF DATA SERIES

The purpose of the analysis of statistical data series is the forecasting of the future development of the data series on the basis of the investigation and separation of trends, long- and short-term periodical fluctuations, seasonality and random fluctuations.

The article presents the theoretically most effective analytical method, that of spectral analysis. The first part deals with definitions and theorems concerning the breaking down of stationary data series into infinitely many periodical components with chance-depending amplitudes. The second part introduces the concept of coherence function which lends itself to the investigation of the relationships between several data series. The third part contains the generalization of spectral analysis for certain types of non-stationary data series, and in the fourth part the author presents two examples relating to the application of the method for checking econometric models.

ОБОБЩЕНИЕ И ПРИМЕНЕНИЕ СПЕКТРАЛЬНОГО АНАЛИЗА РЯДОВ СТАТИСТИЧЕСКИХ ДАННЫХ

Цель анализа рядов статистических данных — прогнозирование их будущей динамики на основе изучения и выявления тренда, долгосрочных или краткосрочных периодических колебаний, сезонности и случайных колебаний.

В статье представляется наиболее эффективный метод анализа — спекральный анализ. В первой части статьи даются определения и основные положения относительно разбивки стационарных рядов данных на бесконечное множество периодических колебаний, амплитуда которых зависит от случайных моментов. Во второй части фигурирует понятие функции когеренции, при помощи которой может быть изучена взаимозависимость между несколькими рядами данных. В третьей части статьи спектральный анализ распространяется на определенные типы нестационарных рядов данных, а в четвертой части приводится два примера относительно использования этого метода для проверки эконометричеких моделей.

FOGALMAK ÉS MÓDSZEREK c. rovatunk e számunkból anyagtorlódás miatt maradt ki. Az előző számban elkezdett KONDOR GYÖRGY cikk befejező részét a következő számban (3. évf. 1. szám) közöljük.

(Szerk.)