

A SABL-ELJÁRÁS FELHASZNÁLÁSA ELEMZÉSRE ÉS PROGNOZTIZÁLÁSRA*

KISS TIBOR – KRUZSLICZ FERENC
– SIPOS BÉLA – SZENTMIKLÓSI MIKLÓS

A SABL- (Seasonal Adjustment – Bell Laboratories) eljárást a Bell Laboratórium munkatársai 1979-ben publikálták. [2] A SABL-eljárás a szezonális idősorok illesztését, simítását végzi el. Additív komponensekre bontja az eredeti vagy a transzformált adatokat: trend, szezonális (periodikus) és irreguláris (fehér zaj) összetevőket különböztet meg, rezisztens (ellenállóképes) lineáris vagy nemlineáris simítási módszerek alkalmazásával. A rezisztens fogalma ebben az összefüggésben azt a tulajdonságot jelenti, hogy a módszer nem érzékeny néhány adat kiugróan nagy eltérése (az úgynevezett outlier-ek) által okozott erős zavaróhatásra, torzításra. Azokat az értékeket tekintjük szélsőséges, extrém értékeknek, outlier-eknek, amelyek nagyon távol vannak az eloszlás közepétől, jelentősen különböznek a többi értéktől. Elkülönítésük mind elemzési, mind előrejelzési, illetve modellalkotási szempontból fontos feladat. Feltárásukat a grafikus ábrázolás is segíti. [5]

A szezonális, illetve periodikus illesztés célja általában a múltbeli és a jelenbeli adatok szezonális hullámzásának meghatározása előrejelzési célból. Fontos, például árpolitikai és beruházáspolitikai célból, hogy a vállalat tisztában legyen azzal, változik-e az üzleti aktivitás adott időszakban és a hullámzás kisebb, nagyobb vagy szezonális jellegű-e.

A szezonális meghatározásának célja az ilyen fluktuáció eltávolítása az idősből, annak érdekében, hogy az alapul szolgáló trendhatást azonosíthassuk. Számos módszer áll rendelkezésre a szezonális komponens azonosítására, a legtöbb azonban érzékeny az említett extrém értékek torzító hatására. A szezonális illesztés egyik célja olyan szezonális tényező elérése, amely stabil, azaz nem változik az időszakok folyamán.

A SABL néven ismert eljárás [8] lényegét tekintve simítási módszereket alkalmaz, amihez mozgómediánokat, rezisztens mediánokat és átlagokat, valamint duplanégyzetbecsléseket használ.

A módszer alkalmas az úgynevezett naptárhatás kimutatására. *Cleveland–Devlin* ([1], 489–496. old.) a spektrálanalízis eszközével elemezte a naptárhatást (a hónapok hosszának, a hét adott napjának és az ünnepnapoknak a hatását); az idősorok komponensekre

* A tanulmány a *dr. Sipos Béla* által vezetett OTKA T14722 sz. és T23058 sz. kutatások keretében készült. A szerzők több alkalommal dolgozhattak a Middlesex University Business School-jában, Londonban. Ezúton köszönik meg a British Council támogatását, ami hozzájárult a szoftverfejlesztő munka sikeréhez.

bontására a SABL-módszert használta. A SABL-módszer bármely idősorra alkalmazható. A módszer alkalmazásához legalább három periódus megfigyelt adataira van szükség, napi, heti, havi, negyedéves vagy féléves bontásban.

A SABL DEKOMPOZÍCIÓS ELJÁRÁS

Az idősorok SABL dekompozícióján a következő felbontást értjük:

$$Y = T + S + I$$

ahol:

- T – a hosszú távú trend,
- S – a szezonális komponens (értelemszerűen a konjunkturális komponens is, tehát a periodikus hullámmás),
- I – az irreguláris rész, azaz a fehér zaj.

Általában Y -on az eredeti idősor adatait értjük vagy annak transzformált (például \ln -transzformáció) alakját. A SABL-módszer iteratív eljárás. Minden iterációs lépésben újra számítjuk, finomítjuk a T , S , I értékeket. A lépések során kialakult adatsorok kölcsönhatásban vannak egymással. Az eljárás célja, hogy az idősort olyan komponensekre bontsuk, melyek rendelkeznek a következő tulajdonságokkal:

- T tükrözi az eredeti idősor alacsony frekvenciájú (hosszú távú) ciklusait, azaz trendjellegű komponensét;
- S képviseli a viszonylag stabilan ismétlődő szezonális tendenciát;
- I reprezentálja a maradékot, azaz mindent, ami nem trendjellegű és nem szezonális jellegű az adatsorban.

A módszer – lényegéből fakadóan – az alábbi tulajdonságokkal rendelkezik:

- a T , S és I komponensek közötti adatáthatás minimális;
- extrém vagy szokatlan adatok nem befolyásolják a T és S értékek meghatározását, illetve becslését, az ilyen szokatlan adatok hatása csak az I komponensben tükröződik;
- a T és az S becslése, amennyire lehetséges érzékenyen reagál az idősor szerkezetének változására;
- a módszer csak akkor használható, ha legalább 3 periódus adatai ismertek;
- sok helyen a zérussal való osztás elkerülése érdekében és a transzformációs szabály alkalmazhatósága miatt ki kell kötni, hogy az eredeti adatok csak pozitívak lehetnek.

A SABL-eljárás első lépéseként az eredeti adatsort transzformáljuk, azért, hogy:

- minimalizálja a szezonális amplitúdójának függését a trendkomponens szintjétől;
- egyszerűbbé tegye a szezonális illesztést;
- az idősor komponenseit additívvá alakítsa.

Az Y_t adatsor transzformáltját $D[0]$ jelöli, a transzformációs függvényt egy p paraméter értékének megadásával választhatjuk ki. A p paraméter az idősor additivitásának mértékét mutatja, szokásos értékei: -3; -2; -1; -0,5; -0,25; 0; 0,25; 0,5; 1; 2; 3.

A transzformációs függvény nemnegatív értékekre:

$$D[0] = \begin{cases} Y^p & \text{ha } p > 0 \\ \lg Y & \text{ha } p = 0 \\ -Y^p & \text{ha } p < 0 \end{cases} \quad /1/$$

A p paraméter értékének kiválasztásához meg kell határozni az idősor komponensei (az Y , a T és az S) közötti kapcsolat típusát. Additív kapcsolat esetén az $Y=T+S$ összefüggés érvényes, a multiplikatív kapcsolat az $Y=T \cdot S$ formulával jellemezhető. A multiplikatív kapcsolat logaritmus segítségével (azaz $p=0$ választásával) additívvá alakítható. Ekkor ugyanis az $\lg Y = \lg T + \lg S$ formához jutunk. $p=1$ esetén az adatsor nem változik, $p=0,5$ esetén az eredeti sor négyzetgyökét, $p=0,33$ esetén a köbgyökét kapjuk. Tehát ha $p=0,33$, akkor harmadfokú, ha $p=0,5$, akkor másodfokú parabolikus trendet feltételezünk. Ha az eredeti adatok (például relatív növekedési ütem) negatívak, akkor a p is negatív, így a $D[0]$ pozitív lesz.

Az idősor összetevőinek lehetséges kapcsolódási módjait mutatja az 1. ábra. Grafikusan ábrázoljuk a kiinduló adatokat, és vizsgáljuk, hogy az ábrázoltak közül melyik sémához hasonlít leginkább a kapott görbe. Lineáris trendet feltételeztünk additív kapcsolat esetén és exponenciális trendet multiplikatív trend esetén. Ha nincs trend vagy nincs szezonális, akkor a SABL-módszert értelemszerűen nem használjuk (lásd az 1. ábra első sorát és első oszlopát), mivel nem lehet az idősort trend- és szezon-komponensre bontani. Ha multiplikatív a kapcsolat akkor $p=0$ (lásd az 1. ábra harmadik sorát és harmadik oszlopát), így az eredeti változókat logaritmizáljuk, ha additív a kapcsolat és a trend lineáris, akkor a $p=1$. (Lásd az 1. ábra második sorát és második oszlopát.)

1. ábra. Az idősor összetevőinek lehetséges kapcsolódási módjai

Trend	Nincs szezonális	Additív szezonális	Multiplikatív szezonális
Nincs			
Additív			
Multiplikatív			

Forrás: [9] 69. old.

Először simítási eljárást alkalmazunk az első rezisztens simított trendjének meghatározására. A simított trendet kivonjuk az adatokból, és $(D-T)=(S+I)$ alapján így kapjuk az $(S+I)$ sort. Ezután az $(S+I)$ csökkenő súlyozású (tapered) mozgómediánjainak és mozgóátlagainak számítása következik, így nyerjük a kezdeti rezisztens szezonális komponenst. Ezután a kapott kezdeti szezonális komponenst a (már transzformált) eredeti adatsorból kivonva, a második simított trendet határozzuk meg. Ezt követi az irreguláris komponens számítása: $(D-T-S=I)$.

A SABL-dekompozíció számos segédeljárást használ, amelyeket a módszer ismertetése előtt meg kell ismernünk. A SABL-eljárás vázlata:

- a) Előkészítő lépések
 $D[0] = P(\text{Adatsor})$;
 $T[0] = \text{Fél-trend}(D[0])$;

$S[0] = \text{Fél-szezon } (D[0], T[0]);$
 $I[0] = \text{Fél-irreguláris } (D[0], T[0], S[0]).$

b) Iterációs szakasz

iteráció kezdete ($n=1$);

$T[n] = \text{Új-trend } (D[0], T[n-1], S[n-1]);$

$S[n] = \text{Új-szezon } (D[0], T[n], S[n-1]);$

$I[n] = \text{Új-irreguláris } (D[0], T[n], S[n]);$

ha $n < N$, azaz még nincs vége, akkor $S[n] = \text{Finomítás } (S[n]);$

új iteráció végrehajtása ($n=n+1$).

c) Befejezés

$I[N] = \text{Zajcsendesítés } (I[n]);$

$D[N] = \text{Inverz-}P (D[0]);$

$T[N] = \text{Inverz-}P (T[N]);$

$S[N] = \text{Inverz-}P (S[N]);$

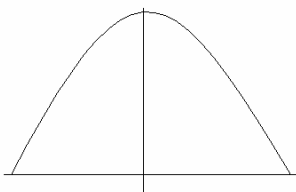
$I[N] = \text{Inverz-}P (I[N]).$

A vázlatban $D[n]$, $T[n]$, $S[n]$ és $I[n]$ jelöli az n -edik iterációs lépés utáni eredeti, trend, szezon és irreguláris adatkomponenseket, valamint N a *niter* értékét, az iterációk számát jelöli; I az adatok transzformációját és Inverz- P a P függvény szerint transzformált adatok visszaalakítását jelenti. A továbbiakban az egyes sorokban szereplő eljárások részletes leírása következik.

A csökkenő súlyozású (tapered) mozgóátlag és mozgómedián számítása során használt súlyok meghatározásához a súlyok olyan sorozatára van szükségünk, ahol a súlyok az adott időszaktól (hónaptól, évtől, naptól stb. attól függően, hogy milyen adatokkal dolgozunk) való távolság arányában csökkennek. Ilyen tulajdonsággal rendelkezik a duplanégyszet- (bisquare) függvény /2/:

$$B(x) = \begin{cases} (1-x^2)^2 & \text{ha } |x| \leq 1 \\ 0 & \text{ha } |x| > 1 \end{cases} \quad /2/$$

amely ábrázolva következő képet mutatja



A szükséges többletadatokat vagy hiányzó adatokat az ndd méretű bázisadatsor alapján becsléssel generáljuk, ha az adatszám n . Legelőször a súlyokat (W) határozzuk meg:

visszafelé becslésnél:
$$W_t = B\left(\frac{t}{ndd+1}\right) \quad /3/$$

előre becslésnél:
$$W_t = B\left(\frac{n+1-t}{ndd+1}\right) \quad /4/$$

ahol: ndd a számítás közben kieső adatok pótlására szolgáló becslés tartományának hossza és B a duplanégyszet függvény.

A súlyozott legkisebb négyzetek módszere a következő. Keressük azt az $y=mx+b$ egyenletű egyenest, amelyre a

$$\sum W_t (Y_t - mX_t - b)^2 \quad /5/$$

összeg minimális. Ahol W_t a B függvénnyel /2/ számított t -edik duplanégyzetes súly /3-/ /4/. Mint ismeretes, m és b értékét úgy kapjuk meg, hogy az előbbi függvényt a két változó szerint parciálisan deriváljuk, és a deriváltakat nullával egyenlővé tesszük, mivel a függvény minimumhelye ott van, ahol az /5/ függvény első deriváltja zérus, és a második derivált ezen a helyen pozitív. Rendezés után a következő megoldás adódik:

$$m = \frac{\sum W_t \cdot \sum W_t \cdot X_t \cdot Y_t - \sum W_t \cdot X_t \cdot \sum W_t \cdot Y_t}{\sum W_t \cdot \sum W_t \cdot X_t^2 - (\sum W_t \cdot X_t)^2} \quad /6/$$

$$b = \frac{\sum W_t \cdot Y_t - m \sum W_t \cdot X_t}{\sum W_t} \quad /7/$$

Az eljárás gyakran használja a mozgómedián- és a mozgóátlag-számítást.

A *mozgómedián- és mozgóátlag-számítások* lényege az, hogy adott méretű „ablakot” mozgatva fentről lefelé, az ablakban szereplő adatok mediánját, illetve átlagát számítjuk ki, és a kapott számokat rendezzük egy új sorozatba. Az ablak mérete, szélessége a mozgóátlag- és mozgómedián-számítás tagszámát jelenti, azaz azt az értéket, ahány szám átlagát vesszük. Mivel az átlagolásnál figyelembe veendő számsor kezdete fentről lefelé mozog, ezért nevezzük az eljárást mozgóátlag-számításnak. A számolás során az ablak méreténél eggyel kevesebb számú adat elvész, a kapott értékeket pedig az adatsoron mozgó ablak közepének feleltetjük meg. Ezt nevezzük centírozásnak. Ha az ablak mérete páros szám, akkor általában újabb ablakmozgató számítás szokott következni, hogy a centírozás pontosan középre essen.

Átlagszámításkor természetesen az ablakban szereplő számok összegét osztjuk az adatok számával, míg a medián kiszámítása során az adatokat sorba rendezzük, és páratlan elemszám esetén a középsőt, páros elemszám esetén pedig a két középső átlagát vesszük.

Súlyozott mozgómedián, súlyozott mozgóátlag. Ezek az eljárások már nem járnak adatvesztéssel úgy, mint a mozgóátlag- és a mozgómedián-számítás. Itt is adott méretű ablakot mozgatunk fentről lefelé, de az adatokat súlyozzuk, és a súlyozott adatokból számolunk átlagot, illetve a „medián” keresésekor kicsit bonyolultabb módon kell kiszámítanunk a középső elemet.

Ha az ablak mérete n , akkor az adatokhoz tartozó súlyok az alábbiak:

$$W_t = B \left(2 \frac{n+1-t}{n+1} \right) \quad t=1 \dots n, \quad B \text{ a duplanégyzetfüggvény.} \quad /8/$$

Mint említettük, az eljárás során nem veszünk adatokat, ami az előző eljárásban abból eredt, hogy az ablak az eljárás kezdetén és végén „kilóg” az adatsorból. Azt mondhatjuk, hogy mindig az ablak közepét helyezzük az aktuális adatra, és a hiányzó adatokat 0-nak

értelmezzük. Az ablakok méretét (nts és nss) éppen azért szokás páratlan számként definiálni, hogy a középső elem egyértelmű legyen. A középső elem súlya ilyenkor mindig 1 lesz, és a többi elem szimmetrikusan csökkenő módon helyezkedik el körülötte.

Páros ablakméret esetén a két középső súly lesz a legnagyobb, és definíció kérdése, hogy melyiket feleltessük meg az aktuális adatnak. A többi súly ez esetben is szimmetrikus lesz. A SABL-program páros ablakméret esetén mindig az első értéket illeszti.

Az iterációs eljárás lényege a következő: először meghatározunk egy kezdeti T , S , I sorozatot, majd az iteráció minden lépésében finomítjuk az adatokat. A *niter* paraméter jelöli a végrehajtandó fokozatos közelítő lépések (a finomítások) számát. Szokásos értéke 5, amely már gyakorlati szempontból megfelelő finomítást ad.

Az egyes lépések végrehajtásához a következő adatok szükségesek:

ndt – az idősor hossza adatokban számolva;

ndp – egy szezon hossza adatszámban mérve, szokásos értékei, ha hónapokkal dolgozunk: 2, 3 (negyedéves), 6 (féléves), 12 (éves), 24 ...; ha hetekkel dolgozunk: 13 (negyedéves), 26 (féléves), 52 (éves);

nss – a szezonális rész simítási faktora, minél nagyobb választunk, annál simább adatokat kapunk a számítás során; szokásos értékei: $[ndt/ndp]+1$ (kis), $2[ndt/ndp]+3$ (közepes), $4[ndt/ndp]+7$ (nagy), ahol a szögletes zárójelek az egészrészfüggvényt jelentik;

nts – a trend rész számításakor használt simítás foka; itt is a nagyobb szám nagyobb simítást eredményez, és szokásos értékei: $[ndp/2]+1$ (kis), $ndp+3$ (közepes), $2ndp+7$ (nagy), a nagyobb simítási érték mindig az előző kétszeresénél eggyel nagyobb;

ndd – a számítás közben kieső adatok pótlására szolgáló becslés tartományának hossza: lehetséges értékei: 1, 2, ... $ndt/(2 \cdot ndp)$;

var – ezt a paramétert nem kötelező megadni, de ha megadjuk, akkor az iterációs eljárás végeztével a végeredményben az outlierek irreguláris értékeit korrigálni lehet. Minden olyan irreguláris adat vágásra kerül, ahol: $|irreguláris_adat| \geq var * irreguláris_sor_szórása$ azaz, ha az irreguláris adat nagyobb, mint az irreguláris sor szórásának var -szorosa, akkor azzal az értékkel helyettesítjük. A var szokásos értékei: 2 és 3.

Az iterációs eljárás lépései az alábbiak.

a) A kezdeti $T[0]$ féltrend meghatározása:

- (1) Vegyük az eredeti $D[0]$, már transzformált idősort.
- (2) Képezzük a (1) sor ndp hosszú mozgómediánját.
- (3) Képezzük a (2) sor ndp hosszú mozgóátlagát.
- (4) Képezzük a (3) sor 3 hosszú mozgóátlagát. Az 1., 2., 3. lépések során két ciklus adatai elvesznek. A 4. lépés az eltolódott adatok helyreigazításához szükséges.
- (5) Képezzük a (4) sor nts hosszú súlyozott mozgóátlagát. Ezzel elkészítettünk egy durva trendet.
- (6) Vonjuk ki a (4) sorból az (5) sor adatait.
- (7) Képezzük a (6) sor nts hosszú súlyozott mozgóátlagát.
- (8) Adjuk össze a (7) és a (5) sor adatait.

A 8. lépés után rendelkezésünkre áll a $T[0]$ kezdeti féltrend.

b) Kezdeti $S[0]$ félszezon meghatározása:

- (9) A $D[0]$ adatsorból vonjuk ki a $T[0]$ sor adatait.

Az i fusson végig 1-től egészen a periódus hosszáig (ndp), és minden lehetséges értékére hajtsuk végre az alábbi lépéseket:

- (10) Képezzük a (9) idősor i -edik szeletét, azaz vegyük sorra a következő sorszámú elemeket: i , $i+ndp$, $i+2ndp$, $i+3ndp$, ...

- (11) Vegyük a (10) lépésben keletkezett részsorozat nss hosszú súlyozott mozgómediánját.
 (12) Vegyük a (11) lépésben kapott sor nss hosszú súlyozott mozgóátlagát.
 (13) A (12) sor adataira alkalmazzuk a szélső ndd adatokon alapuló súlyozott legkisebb négyzetes becslés módszerét annak érdekében, hogy a részsorozat elejéről és végéről hiányzó adatokat pótoljuk.
 (14) Rendezzük a (13) lépésben kapott ndp darab (13) részsorozatot újra egy sorozatba, hogy minden elem az eredeti helyére kerüljön vissza.

A 14. lépés után rendelkezésünkre áll az $S[0]$ kezdeti félszezon.

c) Kezdeti $I[0]$ félirreguláris meghatározása:

- (15) A (9)-es adatsorból vonjuk le a félszezont (14).

A 15. lépés után rendelkezésünkre áll az $I[0]$ kezdeti félirreguláris komponens.

d) Az új trend $T[n]$ meghatározása.

- (16) Vonjuk ki az (1) $D[0]$ transzformált adatsorból a (14) $S[n-1]$ adatsort. Ezzel a trend plusz irreguláris részére egy újabb közelítést kapunk.
 (17) Vegyük a (16) sor nts hosszú súlyozott mozgómediánját.
 (18) Az iterált durva trendet a (17) adatsor nts hosszúságú súlyozott mozgóátlagaként kapjuk.
 (19) A (16) adatsorból levonva a (18) adatsort, megkapjuk az iterált irreguláris részt.
 (20) Vegyük a (19) adatok nts hosszú súlyozott mozgómediánját.
 (21) Képezzük a (20) adatsor nts hosszú súlyozott mozgóátlagát.
 (22) A $T[n]$ sort a (18) durva trend és a (21) sor összegeként állíthatjuk elő.

e) Az új szezon $S[n]$ meghatározása:

- (23) Vonjuk ki az (1) $D[0]$ adatsorból a (22) $T[n]$ adatsort.
 (24) A (23) lépésben kapott idősből vonjuk ki a (14) $S[n-1]$ értékeit.

Az i fusson végig 1-től egészen a periódus hosszáig (ndp), és minden lehetséges értékre hajtjuk végre az alábbi lépéseket:

- (25) A kezdeti félszezon meghatározásánál leírtak szerint képezzük a (24) szezonosorozat i -edik részsorozatát. A keletkezett sorozat t -edik elemét jelöljük I_t -vel.
 (26) Vegyük a (25) sor abszolút értékeit.
 (27) Képezzük a (26) sor $2nss+1$ hosszú súlyozott mozgómediánját. A keletkezett sorozat t -edik elemét jelöljük M_t -vel.
 (28) A rezisztens súlyokat, az R_t sorozatot a következő szabály alapján képezzük: $R_t = B(I_t/6M_t)$, ahol B a duplanégyzetfüggvény.
 (29) Képezzük a (28) sorozat nss hosszú mozgóátlagát.
 (30) Képezzünk i -edik részsorozatot a (25) lépésnek megfelelően, de most tegyük ezt a (23) lépésben kapott sor adatai alapján. Ezt a részsorozatot $(S+I)_t$ -vel jelölhetjük.
 (31) A (28) és a (30) sorozat elemeit szorozzuk össze.
 (32) Képezzük a (31) sorozat nss hosszú mozgóátlagát.
 (33) Vegyük a (29) és a (32) sorozat hányadosát. Ez tulajdonképpen az $(S+I)_t$ sorozat rezisztensen súlyozott mozgóátlaga. Ennek tömör képlete a $W_t, R_t, R_t(S+I)_t$ adatok segítségével a következő, ahol $(S+I)_t$ az $S+I$ sor t -edik elemét jelenti:

$$\frac{\sum W_t \cdot R_t(S+I)_t}{\sum W_t \cdot R_t} = \frac{\sum W_t \cdot [R_t(S+I)_t] / \sum W_t}{\sum W_t \cdot [R_t] / \sum W_t} = \frac{\text{Csökkenő súlyozottságú átlag } [R_t(S+I)_t]}{\text{Csökkenő súlyozottságú átlag } [R_t]} \quad /9/$$

- (34) Az $S[n]$ sort úgy kapjuk meg, hogy a (33) lépés során keletkezett ndp darab részsorozatot újra egy listaiba rendezzük úgy, hogy mindegyik az eredeti helyére kerüljön.

f) Az új-irreguláris $I[n]$ meghatározása:

(35) Az (1) $D[0]$ transzformált adatsorból vonjuk le a (22) $T[n]$ -et, valamint a (34) $S[n]$ -et, hiszen $I[n]=D[0]-S[n]-T[n]$.

g) A szezonális komponens $S[n]$ finomítása.

Ha már elértük a kellő iterációs számot, akkor térjünk át a befejező lépések végrehajtására. Egyébként pedig az i fusson végig 1-től egészen a periódus hosszáig (ndp), és minden lehetséges értékére hajtsuk végre a következő lépéseket:

(36) Képezzük a (34) $S[n]$ sorozat i -edik részsorozatát.

(37) A (36) sor adataira alkalmazzuk a szélső ndd adatokon alapuló súlyozott legkisebb négyzetes becslés módszerét annak érdekében, hogy a részsorozat elejére és végére olyan új adatokat generáljunk, amelyek a következő lépések során elhasználnak.

(38) Rendezzük teljes sorozatba a (37) lépésben kapott ndp darab részsorozatot a szokott módon.

(39) Vegyük a (38) sor ndp hosszú mozgóátlagát.

(40) Képezzük a (39) sor 2 hosszú mozgóátlagát.

(41) Számítsuk ki a (40) sorozat nts hosszú súlyozott mozgóátlagát.

A továbbiakban az így keletkezett sor elejéről és a végéről összesen ndp darab szimmetrikus adatot figyelmen kívül hagyunk:

(42) Az új iterációhoz felhasználandó $S[n]$ sorozatot úgy kapjuk, ha az eredeti (34) $S[n]$ sorozatból kivonjuk a (41) lépés során kiszámolt sort.

(43) A számítást immár a finomított szezonális komponens segítségével folytassuk a (16) ponttól újra-kezdve.

e) Befejező lépések:

(44) Erre a zajcsendesítési lépésre csak akkor kerül sor, ha megadtuk a var értéket. Ez az irreguláris faktor adatsorát módosítja úgy, hogy az outlierok értékeit „levágja”. Egy irreguláris adatot akkor minősítünk „kilógónak”, szélsőségesnek (outliernek), ha az értéke kívül esik a $[-szórás(I[N])-var; +szórás(I[N])+var]$ intervallumon. Az ilyen kívül eső értékeket a hozzájuk közelebb eső intervallumvégpont értékével helyettesítjük, azaz a kilógó részt levágjuk.

(45) Legvégül, a kapott idősorokat visszatranszformáljuk a P függvény inverze segítségével, majd az $expr$ mezőben leírt képleteket kiszámítva elkészítjük a végeredményeket tartalmazó file-t. A $D[0]$ inverze természetesen mindig az eredeti adatsor lesz, ezért ezt a számpéldában nem is tüntettük fel.

Az eljárás lépéseit a következőkben foglaljuk össze.

<i>A SABL dekompozíciós folyamat</i>		
Input	Művelet	Output
Y =eredeti adatok	Adatok transzformálása	$D[0] = \begin{cases} Y^p & \text{ha } p > 0 \\ \lg Y & \text{ha } p = 0 \\ -Y^p & \text{ha } p < 0 \end{cases}$
$D[0]$ =transzformált adatok	Első trendsimítás	$T[1]=D[0]$ adatok első trendsimítása
$D[0]-T[1]$ =nyers szezonális	Szezonális simítás	$S[1]=D[0]-T[1]$ simított szezonális
$S[1]$	A szezonális trend simítása és eltávolítása	$S[2]=S[1]$ mínusz $S[1]$ trend simítása
$S[2]$	Elvesztett szezonális komponensek becslése	$S[3]=a$ kiegészített szezonálisértékek hozzáadása a sor elejéhez és végéhez

$D[0]-S[3]$ =nyers trend	Trendsimitás	$T[2]=D[0]-S[3]$ trend simítása
$D[0]-T[2]$ =nyers szezonális	Szezonális simítása	$S[4]=D[0]-T[2]$ simított szezonálitása
$S[4]$	A szezonális trend simítása és eltávolítása	$S[5]=S[4]$ mínusz $S[4]$ trend simítása
$D[0]-S[5]$ =nyers trend	Trendsimitás	$T[n]=D[0]-S[5]$ trend simítása
$D[0]-T[n]$ =nyers szezonális	Szezonális simítása	$S[6]=D[0]-T[n]$ simított szezonálitása
$S[6]$	A szezonális trend simítása és eltávolítása	$S[n]=S[6]$ mínusz $S[6]$ trend simítása
$D[0], T[n], S[n]$	Irreguláris komponens számítása	$I[n]=D[0]-T[n]-S[n]$
$D[0], T[n], S[n], I[n]$	Grafikus ábrák készítése	Grafikus ábrák értelmezése és diagnosztizálása

Forrás: [2] 206. old.

A SABL tehát elsődlegesen az adatállomány transzformációját végzi el, eredeti formájában közvetlen előrejelzésre nem alkalmas. A helyesen transzformált adatok additív kapcsolatú trend- és periódus- (szezons- vagy konjunktúra-) komponenseket tartalmaznak, irreguláris részt viszont nem. A SABL által transzformált sor felhasználható a különböző időszorosítási módszerekkel (például trend-extrapoláció, exponenciális simítás, ARIMA stb.) történő becslésekhez.

A SABL-módszer alkalmazására számítógépes program készült,¹ amely a következőképpen működik:

A paraméter neve	A paraméter jelentése	Lehetséges érték
<i>in</i>	az idősort tartalmazó adat-file neve, egy adatsor	*.dat
<i>out</i>	a file neve, amelybe az eredményeket tesszük	*.out
<i>ndp</i>	az idősor egy ciklusának hossza (havi adatoknál 12, negyedéves adatoknál 4 stb.)	1..[ndt/3]
<i>nts</i>	a trendsimitás nagysága (a mozgóközeppek: -átlag, illetve -medián elemszáma)	1..65535
<i>nss</i>	a szezonális simítás nagysága (a mozgóközeppek: -átlag, illetve -medián elemszáma)	1..65535
<i>ndd</i>	az iterációban kieső adatok helyreállításához használandó regressziós adatok száma (maximum a periódusok számának fele)	1..[ndt/ (2·ndp)]
<i>ndt</i>	az idősor hossza (az adatok száma), a program kiszámolja, nem kell megadni	1..16384
<i>niter</i>	az algoritmus során végrehajtandó iterációs lépések száma	0..65535
<i>p</i>	a be- és kimeneti adatok transzformációs paramétere: bármely valós szám	bármely valós szám
<i>var</i>	irreguláris rész levágása, az alábbi összefüggés szerint: irreguláris adat ≥ var · szórás(irreguláris sor)	pozitív valós szám
<i>expr</i>	az eredmény-file-ba kinyomtatandó értékek képletei pontosvesszővel elválasztva, ahol használhatók a következők: zárójelzett műveletek (*/-+^) valós konstansok a következő betűk: <i>D</i> (eredeti adatok), <i>T</i> (trend komponens), <i>S</i> (szezónális komponens), <i>I</i> (irreguláris rész).	<i>D, T, S, I</i>

A bemutatott számítógépes programban egyedül az „*in*” paraméter megadása kötelező, az összes többi nem.

¹ Kruzslicz Ferenc – Kiss Tibor – Sipos Béla: SABL decomposition of time series. Janus Pannonius Tudományegyetem. Pécs. 1997.

GYAKORLATI ALKALMAZÁSOK

A módszer alkalmazására két példát is bemutatunk. Először a BUX-indexet² prognosztizáljuk, majd a Kondratyev-hullámok illesztésére használjuk a SABL-eljárást.

A BUX-index prognosztizálása

A BUX-index napi értékei rendelkezésünkre állnak. Az éveket azonos hosszúságra: 252 tőzsdenapra egészítettük ki. Ha egy évben kevesebb adat állt rendelkezésre, akkor egy hétköznapra eső ünnep esetén az előző tőzsdenap indexét használtuk. A szombat-vasárnap mindig szünnap. Az 1991. január 2. és 1997. április 30. közötti időszak, azaz 1594 nap adatait vettük figyelembe. Az induló paraméterek a következők:

$ndp=252$, azaz éves szezonaritást kerestünk első lépésben;
 $nts=30$ és $nss=30$, a trend és szezonális simítás fokát azért választottuk erőse, hogy a rövidebb ingadozásokat (2, 3, 5, 6, 10, 15 napos) kiküszöböljük;
 $ndt=1594$ (az adatsor hossza napokban, a program kiszámítja, nem kell megadni);
 $ndd=2$, a mozgóátlag-számítás során kieső adatok helyreállításához használandó adatok száma, szokásos értéke 1 vagy 2;
 $p=1$, azaz nem transzformáljuk az eredeti adatokat, feltételezzük az additív kapcsolatot; eredmények a d , t , s , i adatsorokat, vagyis az idősor trend, szezonális és irreguláris komponensét kérjük az eredeti adatsor mellett;
 $niter=5$, azaz a finomító lépések száma öt, ami már megfelelő finomítást biztosít;
a var paraméternek nem adunk értéket, így a kilógó adatok nem kerülnek vágásra.

A futtatás eredményeként kapott szezonális komponens ábrája azt mutatta, hogy az évesnél kisebb: féléves hullámzás van a idősorban, ezért a periódus értékét módosítottuk, és az $ndp=126$ periódushosszt határoztuk meg a részletes elemzés céljára. Tekintettel az adatbázis rendkívül nagy méretére – bár a teljes adatsor rendelkezésre áll – csak az első hónap (1991. január) és az 1997. év áprilisi adatait, valamint az egyhavi „ex ante” prognózis értékeit szemléltetjük.

A SABL-módszer önmagában nem alkalmas előrejelzésre, azonban van olyan technikája, amely a számítások során az előzetesen levágott adatokat kipótolja. Ehhez az előzőekben ismertetett ndd értéket használja fel, amely megmondja, hogy az idősor végétől számítva hány értéket vegyen figyelembe egy egyszerű lineáris trend számításához, amelynek segítségével azután a visszapótlást elvégzi. Ez lokális előrejelzés, amennyiben azonban ugyanezt az idősor teljes hosszára vonatkozó becslésekre is elvégezzük, a módszerhez szervesen igazodó előrejelzést kapunk. A SABL által ily módon szolgáltatott előrejelzés hossza nem haladhatja meg a periódushosszat, ez esetben az ötöt.

A módszert kétféleképpen is teszteljük: először a SABL-becslésekre globális trendet illesztünk (külső becslés) és a kiegészítést arra végezzük el. Ez lesz az első összehasonlítási alap. Ezután mindkét előrejelzésre kiszámítjuk az U statisztikákat [9], és ezzel az eredeti adatoktól (második összehasonlítási alap) való eltérést vizsgáljuk. Ezek segítségével abszolút eltérés és relatív eltérés is vizsgálható. Az adatsort 1997. április 11-ig használjuk a hagyományos SABL-becslésre, a többi adatot az előrejelzés pontosságának tesztelésére használjuk fel.

² A BUX-index a Budapesti Értéktőzsde hivatalos részvényindexe. Leírása megtalálható: Magyar Pénzügyi és Tőzsdei Almanach. VI. évf. II. köt. Szerk.: *Kerekes György*. TAS-11 Kft. Budapest. 1996.1087. old.

A SABL dekompozíciós módszerrel trend-, szezonális és irreguláris komponensre bontjuk az idősort, és kimutatjuk a féléves (azaz 126 napos) szezonális hullámváz jelenlétét. Az eredményeket az 1. táblában terjedelmi korlátok miatt rövidítve közöljük.

1. tábla

A BUX-index és prognosztizálása

Tőzsdenap	BUX-index	SABL-trend-komponens (<i>n</i> _{ts} =30)	Harmadfokú parabola	Szezonálisitás (<i>n</i> _{ss} =30)	Prognózis (harmadfokú parabola + szezonálisitás)
A tőzsdei adatok					
1991.01.02.	1000,00	1006,56	723,97	-16,81	707,15
1991.01.03.	1001,02	1007,12	726,62	-11,10	715,52
1991.01.04.	996,33	1007,94	729,26	-13,46	715,80
1991.01.07.	997,06	1009,07	731,89	-16,38	715,52
1991.01.08.	1001,91	1010,54	734,51	-14,75	719,76
1991.01.09.	997,92	1012,35	737,12	-12,54	724,58
1991.01.10.	1000,65	1014,50	739,71	-5,82	733,90
1991.01.11.	1004,69	1016,99	742,30	-9,96	732,34
1991.01.14.	990,62	1019,82	744,87	-15,23	729,64
1991.01.15.	991,61	1022,98	747,43	-16,21	731,22
1991.01.16.	990,74	1026,46	749,98	-8,89	741,09
1991.01.17.	990,36	1030,25	752,52	-10,41	742,11
1991.01.18.	996,49	1034,33	755,04	-10,90	744,15
1991.01.21.	1014,75	1038,67	757,56	-20,56	737,00
1991.01.22.	1025,55	1043,20	760,06	-19,23	740,84
1991.01.23.	1032,34	1047,80	762,55	-16,40	746,15
1991.01.24.	1056,37	1052,34	765,03	-11,67	753,36
1991.01.25.	1056,30	1056,65	767,50	-17,44	750,06
1991.01.28.	1056,37	1060,63	769,96	-16,16	753,80
1991.01.29.	1058,75	1064,18	772,41	-9,04	763,37
1991.01.30.	1063,61	1067,20	774,84	-8,23	766,61
1991.01.31.	1068,12	1069,62	777,26	-5,73	771,53
M					
1997.04.01.	5319,20	5241,20	5893,82	33,06	5926,38
1997.04.02.	5353,49	5267,02	5907,64	26,95	5935,10
1997.04.03.	5301,15	5294,98	5921,48	23,60	5945,82
1997.04.04.	5315,46	5324,67	5935,34	29,71	5965,09
1997.04.07.	5398,69	5355,79	5949,24	32,26	5981,05
1997.04.08.	5424,05	5388,15	5963,16	30,66	5993,33
1997.04.09.	5481,65	5421,53	5977,10	30,83	6007,08
1997.04.10.	5464,77	5455,69	5991,07	25,23	6016,14
1997.04.11.	5500,25	5489,95	6005,07	28,33	6032,97
Ellenőrző adatok					
1997.04.14.	5442,59		6019,10	24,47	6044,22
1997.04.15.	5506,52		6033,15	18,04	6051,69
1997.04.16.	5567,90		6047,22	16,28	6063,57
1997.04.17.	5662,39		6061,32	14,66	6075,14
1997.04.18.	5703,78		6075,45	12,69	6087,89
1997.04.21.	5902,04		6089,61	4,78	6094,74

(Tábla folytatása a következő oldalon.)

(Folytatás.)

Tőzsdenap	BUX-index	SABL-trend-komponens ($n_s=30$)	Harmadfokú parabola	Szezonális ($n_s=30$)	Prognózis (harmadfokú parabola + szezonális)
1997.04.22.	5795,81		6103,79	7,46	6111,14
1997.04.23.	5840,65		6118,00	12,06	6129,65
1997.04.24.	5789,04		6132,23	17,03	6148,38
1997.04.25.	5804,05		6146,49	9,07	6155,82
1997.04.28.	5847,71		6160,78	9,42	6171,07
1997.04.29.	5952,92		6175,09	7,78	6184,00
1997.04.30.	5985,17		6189,43	10,70	6200,68
			Előrejelzett értékek		
1			6203,79	12,29	6216,09
2			6218,19	9,80	6227,98
3			6232,60	8,22	6240,83
4			6247,05	9,11	6256,16
5			6261,52	9,93	6271,45
6			6276,02	9,33	6285,35
7			6290,55	2,47	6293,01
8			6305,10	7,28	6312,38
9			6319,67	3,01	6322,69
10			6334,28	0,50	6334,78
11			6348,91	-8,74	6340,18
12			6363,57	-9,14	6354,43
13			6378,25	-7,36	6370,90
14			6392,97	-10,43	6382,53
15			6407,71	-18,30	6389,41
16			6422,47	-20,90	6401,57
17			6437,26	-13,37	6423,89
18			6452,08	-22,96	6429,12
19			6466,93	-30,22	6436,71
20			6481,80	-25,08	6456,72
21			6496,70	-32,57	6464,13
22			6511,63	-33,33	6478,29
23			6526,58	-32,29	6494,29
24			6541,56	-34,54	6507,02
25			6556,57	-35,49	6521,08
26			6571,61	-32,82	6538,78
27			6586,67	-39,04	6547,63
28			6601,76	-32,93	6568,83
29			6616,87	-34,20	6582,68
30			6632,02	-32,36	6599,66

Az idősor komponensekre bontását kis (7), közepes (15) és nagymértékű (30) trend-, illetve szezonális simítással is elvégeztük, a trend nem volt érzékeny a simítás értékére. Tekintettel a napi adatok intenzív hullámzására, a szezonális meghatározását nagyfokú simítással végeztük el. A kapott SABL-trendkomponens értékeire az Excel-program segítségével trendet illesztettünk. A harmadfokú parabola adta a legjobb illeszkedést. A kapott trend egyenlete: $y=4 \cdot 10^{-6}t^3+0,0059t^2+2,6117t+721,3$ és $R_3^2=0,905$. A másodfokú parabola

R_2^2 értéke 0,8427 volt. F -próbával teszteltük, hogy a magasabb fokszámú parabola jobb közelítést ad-e. Legyen m -edrendű a magasabb, r -edrendű az alacsonyabb fokszámú parabola ($m > r$). A próbafüggvény képlete [12]:

$$F = \frac{(R_m^2 - R_r^2)/(m-r)}{(1 - R_m^2)/(n-m-1)} \quad /10/$$

ahol n a mintaelemszám. A számláló szabadságfoka $m-r$, a nevezőé $n-m-1$. A nullhipotézis az, hogy az m -edfokú polinom nem ad szignifikánsan jobb közelítést az r -edfokúnál. Ha $F \geq F_{0,05} [(m-r), (n-m-1)]$, akkor 5 százalékos szignifikanciaszinten az m -edfokú parabola jobb közelítést nyújt.

Példánkban a paraméterek értékei a következők: $m=3$, $r=2$, $R_3^2=0,905$, $R_2^2=0,8427$ és $n=1594$, a számláló szabadságfoka 1, a nevezőé 1590. Az adatokat /10/-be helyettesítve:

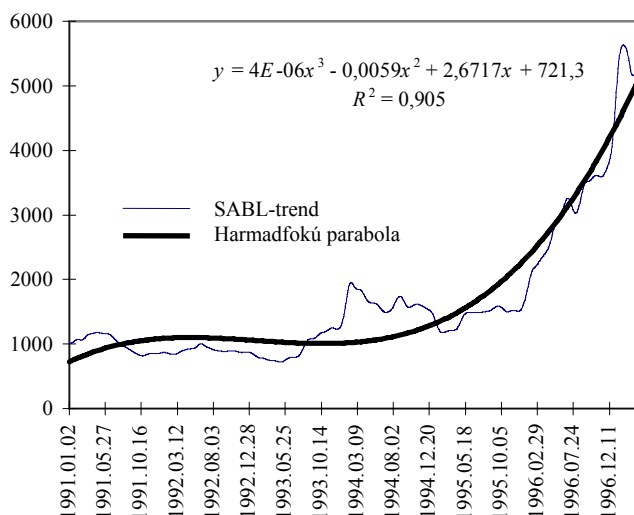
$$F = \frac{(0,905 - 0,8427) : 1}{(1 - 0,905) : (1594 - 4)} = 1042,71$$

$F_{0,05}$ az eloszlástáblázat szerint 3,84, azaz $F > F_{0,05}$, tehát azt állítjuk, hogy 5 százalékos szignifikanciaszinten a harmadfokú parabola valóban jobban illeszkedik.

Célszerű az adatsor különbségeit is elemezni. [13] Az eredeti idősor különbségei ($y_t - y_{t-1}$) másodfokú parabola mentén szóródnak, ami szintén alátámasztja azt, hogy a trend harmadfokú polinomot követ, hiszen a harmadfokú parabola deriváltja másodfokú.

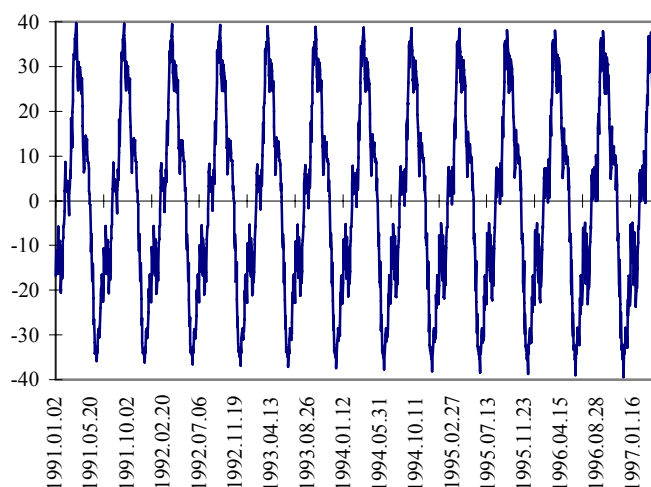
A módszer tisztán matematikai statisztikai szempontból értékeli a közelítést ([11] 183–185. old.), a parabolák fokszáma növelésének ugyanis korlátja van. A harmadfokú polinomnál magasabb fokszámú polinom már nem fogható fel trendként, mivel az már az idősor kilengéseit is követi. A trend viszont nem követheti az idősor minden ingadozását, hiszen ezek a véletlen vagy a periodikus komponens következményei.

2. ábra. A BUX-index és a parabolikus trend nagyfokú simítás mellett



Az R_3^2 többszörös determinációs együttható értéke nagyon magas, ezért az illeszkedés igen jónak mondható. Grafikusan szemléltetjük az eredményeket a 2. ábrán, a szezonális komponens alakulását pedig a 3. ábrán. Megjegyzendő, hogy a Budapesti Értéktőzsde a dinamikus növekedés időszakában van. Később, az érettség szakaszában a magas fokszámú polinomok valószínűleg nem nyújtanak majd megfelelő prognózist.

3. ábra. A BUX-index szezonális hullámzását erős simitással



A 3. ábra határozottan mutatja a BUX-index szezonális hullámzását, melynek legutóbbi mélypontja 1996. december 3-án volt, a csúcspont tavasszal következett, 1997. március 13-án.

A 2. táblában az 1997. április 11 utáni eredeti adatok, a külső trendillesztésen alapuló adatok és a SABL saját, lokális trenden alapuló visszapótlásos előrejelzése található. Mivel az előrejelzés a visszapótlás lényegéből következően csak a ciklus periódushosszára vonatkozik, ezért a program csak öt napra (a következő hét) szolgáltat előrejelzést. Az alkalmazott paraméterek: $ndt=1581$, $ndp=5$, $nts=3$, $nss=17$, $ndd=2$, $niter=5$, $p=0,5$. A trendsimítás értéke kicsi ($nts=3$), hogy a változásokat jobban tudja követni a program. Másodfokú parabolát alkalmaztunk ($p=0,5$), mivel a SABL az utolsó adatokat használja fel előrejelzésre.

2. tábla

A SABL- és külső trenden alapuló előrejelzés

Tőzsdenap	BUX-index	Trenden alapuló előrejelzés	SABL-előrejelzés
1997.04.14	5442,59	6044,22	5539,23
1997.04.15	5506,52	6051,69	5581,19
1997.04.16	5567,90	6063,57	5588,92
1997.04.17	5662,39	6075,14	5633,49
1997.04.18	5703,78	6087,89	5630,01

A külső, trendillesztésen alapuló előrejelzés U statisztika értéke 6,84, míg a lokális illesztés U értéke 0,81. Ez abszolút összehasonlításban annyit jelent, hogy a saját módszer által szolgáltatott U érték a 0-tól kevésbé tér el, és elfogadható. A relatív összehasonlítás azt mutatja, hogy a lokális, saját előrejelzés sokkal jobb eredményt ad, mint a trendillesztésen alapuló előrejelzés. Természetesen a tőzsdeindexek esetén várható hirtelen változásokat egyik megoldás sem tudja követni. A Theil-féle U statisztika képlete:

$$U = \sqrt{\frac{\sum_{i=1}^{n-1} \left(\frac{F_{i+1} - X_{i+1}}{X_i} \right)^2}{\sum_{i=1}^{n-1} \left(\frac{X_{i+1} - X_i}{X_i} \right)^2}}$$

Ha $U=0$, akkor $F_i=X_i$, vagyis az előrejelzés megegyezik a valósággal. Különben U értéke 0-tól különbözik.

Kondratyev-hullámok illesztése SABL-eljárással

Mielőtt a SABL-módszert a hosszú hullámok illesztésére alkalmaznánk, ismerkedjünk meg röviden a hosszú ciklusok fogalmával. A hosszú hullámok első átfogó tudományos magyarázatát *N. D. Kondratyev* adta 1922-ben megjelent könyvének [6] V. fejezetében. Kondratyev műve eddig – tudomásunk szerint – csak orosz nyelven jelent meg, célszerű ezért néhány gondolatát ismertetnünk.

A gazdasági konjunktúrák mozgása ritmikus. A konjunktúrák emelkedő periódusát, többé vagy kevésbé élesen, a leszálló konjunktúrák periódusa váltja fel. Az ilyen ingadozások ciklusának két fő típusát különböztetjük meg: a nagy ciklust, amely mintegy ötven évet ölel át, és a kisebb ipari-tőkés ciklust, amely általában 8–11 évet foglal magába. A nagy ciklus időszaka általában néhány kis ciklust is magába foglal, melyek során a konjunktúra az emelkedés állapotából a depresszió állapotába kerül, ez az átmenet történhet élesen, válság formájában és olykor nyugodtan, krízis nélkül is. Kondratyev az elemzés során arra a következtetésre jutott, és a legfontosabb mutatók mozgását számba véve ma már tényekre alapozva állíthatjuk: az első világháború ténylegesen akkor kezdődött, amikor a világgazdaság és a legfontosabb országok gazdasága a nagy ciklus felemelkedő ágában volt. ([6], [7]).

A konjunktúraciklusok következő típusai különböztethetők meg [3]:

- a) a Kitchin- vagy készletciklus, amely 3–5 éves időtartamú, rövid távú ciklus;
- b) a Juglar- vagy gép- és eszközbefektetési ciklus, melynek periódusa 7–11 év;
- c) a Kuznets- vagy építési befektetési ciklus középtávú, 15–25 éves;
- d) a Kondratyev-ciklus, azaz a hosszú hullámok, az alapvető tőkejavak létrehozásával összefüggő hullámok amelyeknek időtartama 30–50 év;
- e) az évszázados (szekuláris) trendek változása.

Az utóbbi, a szekuláris trend az alábbiak szerint alakult Európában [13] (zárójelben a csúcspont éve):

1. szekuláris trend: 1250–(1350)–1510
2. szekuláris trend: 1510–(1650)–1740
3. szekuláris trend: 1740–(1817)–1896
4. szekuláris trend: 1896–(1973)–

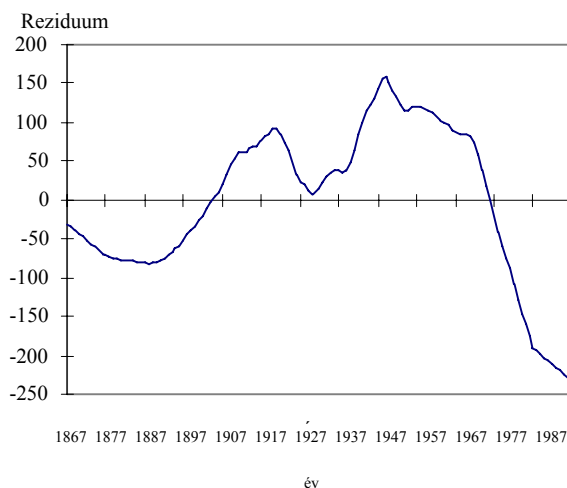
A ciklusokat, felfedezőjükről, *J. A. Schumpeterről* [10] nevezték el. Mind a négy ciklus egyidejűleg létezik és Schumpeter nyomán kialakult az a nézet, hogy egymáshoz illeszkednek. Más vélemény szerint a négy ciklus egymástól függetlenül hat. Megfigyelhető az is, hogy néhány összetett ciklus különböző hosszúságú ciklikus hullámzásokból épül fel. Ez a szemlélet vezethet ahhoz, amely szerint például az 1929–1933-as gazdasági válság súlyosságát a Kitchin-, Juglar-, Kuznets- és Kondratyev-ciklusok, a gazdasági ciklus motorját jelentő befektetési hullám leszálló ágának egybeesésével magyarázhatjuk.

Vizsgáljuk meg a Kondratyev-hullámok jelenlétét az Egyesült Államok acéltermelésében az 1867–1996 közötti időszakban a SABL-eljárás segítségével. A SABL-dekompozíció révén a trendet – amely ez esetben az évszázados trend – eltávolítjuk az idősből, és a ciklikus hullámzásokat vesszük vizsgálat alá.³ Az induló paraméterek:

$ndp=28$ – azaz a Kondratyev ciklus hosszát 28 évesnek feltételezzük a reziduomok elemzése alapján;
 $nts=15$ és $nss=15$ – a trend és a periodikus simítás mértékét közepesre állítva kiküszöböltük a rövidebb (3, 5, 15 éves) ingadozásokat;
 $ndt=130$ – az adatsor hossza (a program számítja ki);
 $ndd=2$ – a mozgóátlag-számítás során kieső adatok helyreállításához használandó adatok száma;
 $p=1$ – azaz nem transzformáljuk az eredeti adatokat, feltételezzük az additív kapcsolatot; eredménynek a d , t , s , i adatsorokat, vagyis az idősor trend-, szezonális és irreguláris komponensét kérjük az eredeti adatsor mellett;
 $niter=5$ – a finomító lépések száma öt;
a var paraméternek nem adunk értéket, a szélsőséges adatokat nem vágjuk le.

A 18 tagú mozgóátlag számítása után kapott reziduomokat közöljük a 4. ábrán. Látható, hogy a legutóbbi maximumok időpontja, 1921 és 1949 között 28 év telt el.

4. ábra. Az Egyesült Államok acéltermelése
18 tagú mozgóátlagának reziduumai



³ Az adatok forrása: Világgazdasági idősorok (1860–1960). (Közgazdasági és Jogi Könyvkiadó. Budapest. 1965.); Nemzetközi statisztikai évkönyvek. (Központi Statisztikai Hivatal. Budapest.); illetve Tények könyve. Ráció Kiadó–Greger Média Kft. Debrecen 1992. és 1993. évi kötetei, valamint Nemzetközi statisztikai zsebkönyv. (Központi Statisztikai Hivatal. Budapest. 1994.) és *Monthly Statistics* (United Nations) és az INTERNET számítógépes információs hálózat.

A 3. táblában rövidítve közöljük az Egyesült Államok acéltermelésével kapcsolatos számítások eredményeit.

3. tábla

Az Egyesült Államok acéltermelésének prognosztizálása

Év	Acéltermelés (kilogramm/fő)	Az acéltermelés 18 tagú mozgó- átlaga	SABL-trend komponens (<i>n</i> _S =15)	Harmadfokú parabola	Szezonális (<i>n</i> _S =15)	Prognózis (parabola + szezonális)
Statisztikai adatok						
1867	0,54	13,956	25,36	17,91	0,33	18,24
1868	0,71	13,956	26,28	15,48	-5,43	10,05
1869	0,82	13,956	27,08	13,44	-9,43	4,02
1870	1,75	13,956	27,76	11,79	-13,81	-2,03
1871	1,81	13,956	28,31	10,50	-16,93	-6,44
1872	3,45	13,956	28,70	9,57	-17,50	-7,92
1873	4,7	13,956	28,89	9,00	-17,53	-8,52
1874	4,97	13,956	28,86	8,78	-17,75	-8,96
1875	8,79	13,956	28,58	8,90	-16,38	-7,48
1876	11,76	13,956	28,10	9,36	-14,05	-4,70
1877	12,28	16,022	27,59	10,13	-10,25	-0,12
1878	15,44	18,819	27,18	11,22	-7,62	3,60
1879	19,31	21,691	26,97	12,62	-4,88	7,74
1880	25,21	24,492	26,98	14,33	-1,75	12,58
1881	31,29	27,807	27,21	16,32	0,70	17,03
1882	33,41	31,219	27,62	18,60	1,24	19,84
1883	31,44	34,793	28,23	21,16	2,72	23,89
1884	28,46	38,223	29,11	23,99	6,63	30,62
M						
1976	540,63	509,69	501,05	392,86	16,69	409,54
1977	525,47	494,74	489,49	386,68	15,01	401,69
1978	568,19	478,47	478,00	380,04	8,38	388,43
1979	558,07	464,23	466,80	372,93	0,59	373,52
1980	442,69	453,41	455,97	365,35	-4,71	360,64
1981	473,87	443,46	445,54	357,27	-8,83	348,44
1982	284,84	428,99	435,47	348,70	-13,93	334,78
1983	321,12	411,54	425,69	339,63	-17,46	322,18
1984	322,78	398,95	416,22	330,05	-17,54	312,51
1985	330,99	388,94	407,17	319,95	-17,80	302,15
1986	301,69	378,66	398,73	309,33	-18,07	291,25
Ellenőrző adatok						
1987	336,56	368,05		298,17	-16,94	281,22
1988	365,39	368,05		286,47	-14,29	272,17
1989	355,36	368,05		274,22	-9,87	264,34
1990	392,05	368,05		261,41	-7,29	254,12
1991	315,41	368,05		248,04	-4,55	243,49
1992	330,56	368,05		234,09	-1,58	232,51
1993	335,42	368,05		219,57	0,64	220,21
1994	340,3	368,05		204,46	0,44	204,89
1995	355,81	368,05		188,75	1,60	190,34
1996	355,82	368,05		172,44	5,51	177,94

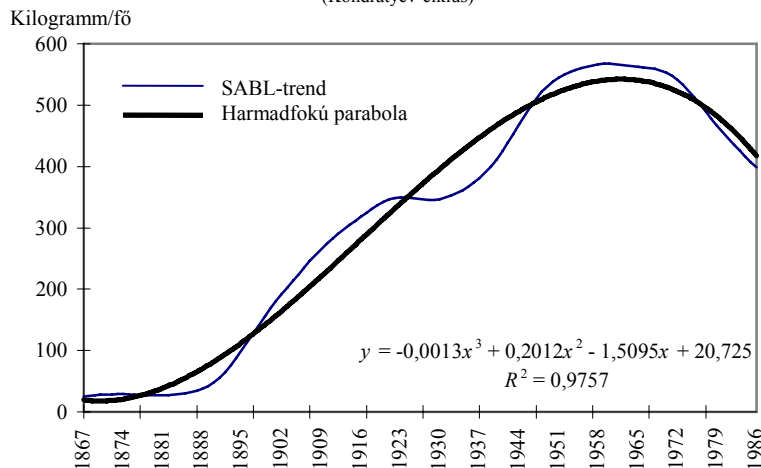
Kondratyev vizsgálataiban 9 tagú mozgóátlagolást végzett a rövidebb ciklusok: a Kitchin- (3 éves) és a Juglar- (9 éves) hullámzás kiküszöbölése érdekében. (Ezzel [12]-ben részletesen foglalkoztam.) Mi az eredeti adatok 18 tagú mozgóátlagát tekintettük kiinduló értéknek, annak érdekében, hogy a Kuznets-ciklust is kiküszöböljük a modelltől. Ez összhangban van azzal, hogy [3] megmutatta, Európában a Kuznets-ciklus hossza átlagosan 18 év. A SABL-dekompozícióhoz közepes (15) trend- és szezonális simítást használtunk. A mozgóátlagolás technikájának következtében az első és az utolsó tíz év adatai azonosak, ezért a dekompozíció során az utolsó tíz évet „levágtuk”, így a trend illesztésénél $n=120$. Ismét az Excel-program segítségével illesztettük az adatokhoz a trendet. A legjobb közelítést a harmadfokú parabola szolgáltatja. A trend egyenlete: $y=0,013x^3+0,2012x^2-1,5095x+20,725$ és $R_3^2 = 0,9757$, ami szoros kapcsolatra utal. A harmadfokú parabola tesztelése a másodfokúval szemben azt mutatta, hogy jelentősen javult az illeszkedés. A harmadfokú parabolát ismét teszteltük a másodfokúval szemben. A paraméterek értékei a következők: $m=3$, $r=2$, $R_3^2=0,9757$, $R_2^2=0,9242$ és $n=120$, a számláló szabadságfoka 1, a nevezőé 116. Az adatokat a /10/ képletbe helyettesítve:

$$F = \frac{0,9757 - 0,9242h}{0,9757h^{20-4h}} = 254,84$$

Az $F_{0,05}$ értéke az eloszlástáblázatban 3,92, tehát $F > F_{0,05}$, ezért azt mondhatjuk, hogy 5 százalékos szignifikanciaszinten a harmadfokú parabola valóban jobban illeszkedik.

Az idősor különbségei másodfokú parabolát követnek, ami igazolja, hogy a harmadfokú parabola illesztése helytálló. Az 5. ábrán szemléltetjük a 18 tagú mozgóátlagra illesztett harmadfokú trendet.

5. ábra. Az Egyesült Államok egy főre jutó acéltermelése 18 tagú mozgóátlag alapján (Kondratyev-ciklus)

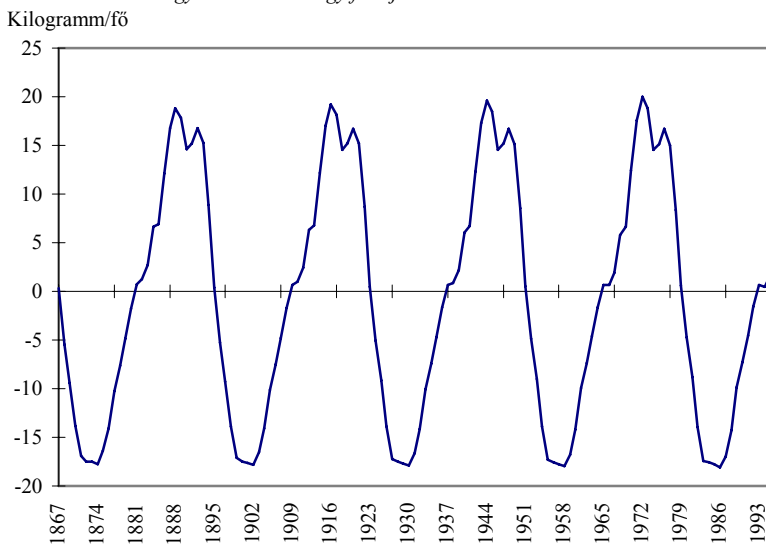


A szekuláris trend, mint már említettük, 1973-ban érte el fordulópontját, azóta leszálló ágban van, ami egybeesik a Kondratyev-ciklusnak a hetvenes évek közepétől tartó leszálló tendenciájával. Mindezt már a reziduumok tanulmányozása is előrevetítette. (Lásd a 4.

ábrát.) Az Egyesült Államok acéltermelésének reziduuma szintén csökkennek a hetvenes évek közepétől, majd 1976-tól tartósan negatív értéket vesznek fel, ami az alaptendencia megfordulására (az évszázados trend és a Kondratyev-hullám fordulópontjára) utal. Az eredmények igazolták várakozásainkat. Az acéltermelés a hetvenes évek közepétől-végétől csökken, ekkor volt az úgynevezett második olajárrobbanás, amelynek jól ismert következménye az egyes országokban és a világ gazdaságban végbement mélyreható szerkezeti átalakulás.

A SABL-dekompozíció „periodikus” komponensét tekinthetjük most a modell hosszú ciklusú elemének, amit a 6. ábrán mutatunk be. A ciklus maximumát 1972-ben, minimumát 1986-ban érte el, 1992-ig negatív szakaszban volt. Jelenleg pozitív értéket mutat, de még mindig nagyon alacsony.

6. ábra. Az Egyesült Államok egy főre jutó acéltermelésének hosszú hullámai



4. tábla

A SABL- és külső trenden alapuló előrejelzés

Év	Acéltermelés	A trenden alapuló előrejelzés	A SABL-előrejelzés
1987	336,56	281,22	310
1988	365,39	272,17	330
1989	355,36	264,34	350
1990	392,05	254,12	370
1991	315,41	243,49	350
1992	330,56	232,51	350
1993	335,42	220,21	310
1994	340,3	204,89	300
1995	355,81	190,34	330
1996	355,82	177,94	330

Ezt az adatsort is teszteltük a SABL belső előrejelzési technikájával. Az alkalmazott paraméterek: $ndt=130$, $ndp=5$, $nts=3$, $nss=17$, $ndd=2$, $niter=5$, $p=0,5$. Az eredményeket az 4. táblában foglaltuk össze.

Ez esetben is a trenden alapuló előrejelzés U statisztika értéke (4,39) lényegesen magasabb, mint a SABL belső előrejelzése esetén ($U=1,199$), tehát az előrejelzésre célszerűbb SABL belső előrejelzést használni.

IRODALOM

- [1] Cleveland, W. S. – Devlin, S. J.: Calendar effects in monthly time series; Detection by spectrum. *Journal of the American Statistical Association*. 1980. évi 75. sz. 489–496. old.
- [2] Cleveland, W. S. – Dunn, D. M. – Terpenning, I. J.: SABL – A resistant seasonal adjustment procedure with graphical methods for interpretation and diagnosis. Bell Laboratories. Megjelent: Seasonal analysis of economic time series. Szerk.: A. Zellner. Government Printing Office. Washington. 1978. 201–231. old.
- [3] Duijn, J. J. Van: The long wave in economic life. George Allen and Unwin. London. 1983. 239 old.
- [4] Hajdu Ottó – Hermann Sándor – Pintér József – Rappai Gábor – Rédey Katalin: Statisztika I–II. Janus Pannonius Tudományegyetem. Pécs. 1994. 319 és 245 old.
- [5] Kerékgyártó Györgyné–Mundruczó György: Statisztikai módszerek a gazdasági elemzésben. Budapesti Közgazdaságtudományi Egyetem. AULA Kiadó. Budapest. 1995. 571 old.
- [6] Kondratyev, N. D.: Mirovoje hozajsztvo i ego konjunkturü vo vremja i poszle vojnü. Vologda. 1922.
- [7] Kondratyev, N. D.: A gazdasági fejlődés hosszú hullámai. *Történelmi Szemle*. 1980. évi 2. sz. 241–269. old.
- [8] Levenbach, H. – Cleary, J. P.: The beginning forecaster: The forecasting process through data analysis. Lifetime Learning Publications. Belmont. California. 1981. 372 old.
- [9] Makridakis, S. – Wheelwright, S. C. – McGee, V. E.: Forecasting. Methods and applications. 2. Ed. John Wiley & Sons, New York–Chichester–Brisbane–Toronto–Singapore. 1983. 926 old.
- [10] Schumpeter, J. A.: Business cycles. A theoretical, historical and statistical analysis of the capitalist process. 1–2. köt. McGraw-Hill. New York. 1939. 1132 old.
- [11] Sipos Béla: Vállalati árelőrejelzések. Közgazdasági és Jogi Könyvkiadó. Budapest. 1985. 292 old.
- [12] Sipos Béla: A Kondratyev-ciklus empirikus vizsgálata és prognosztizálása. *Statisztikai Szemle*. 1986. évi 12. sz. 1209–1237. old.
- [13] Sipos Béla: Vállalati prognosztika (Piaci és termelési prognózisok készítése). Janus Pannonius Tudományegyetem Kiadó. Pécs. 1995. 222 old.
- [14] Sipos Béla – Szentmiklósi Miklós: A hosszú hullámok alakulása a mezőgazdaságban. *Statisztikai Szemle*. 1991. évi 10. sz. 798–812. old.

TÁRGYSZÓ: Konjunktúraciklus-modell.

SUMMARY

Data decomposition is an important phase of forecasting process. SABL procedure for adjusting seasonal time series decomposes raw data or transformed data into additive, trend seasonal and irregular component by using resistant linear and nonlinear smoothing techniques. Resistance in this context means the property of being insensitive to gross perturbations of a small part of the data. The procedure uses a smoothing process which consists of moving averages, medians, tapered means and medians, and bisquare estimates.

The advanced SABL procedure is able to decompose and forecast data of time series.