

# WAVELETEK ÉS ALKALMAZÁSUK AZ IDŐSOROK ELEMZÉSÉBEN

ZSEMBROVSZKY PETRO<sup>1</sup>

MATÁV *Piacelemző Osztály*

A hagyományos Box-Jenkins idősoros statisztikai modellek (ARMA, ARIMA) csak olyan hosszú stacionárius idősorokon adnak megbízható eredményeket, amilyenek a legtöbb esetben nem állnak rendelkezésre. Ezért fordul a figyelem egyre inkább olyan módszerek felé, amelyek viszonylag rövid idősorokon is képesek hatékony elemzést adni és szilárd bázist biztosítani az előrejelzések számára. Az utóbbi években éppen ezért gyors fejlődésnek indult néhány, a tradicionálistól eltérő idősorelemzési irányzat: a wavelet elemzés, a fraktál analízis és a neurális hálózatok alkalmazása. Ez a cikk az említett módszerek közül a wavelet analízis elméleti alapjait és alkalmazási lehetőségeit mutatja be egy, a távközléshez kapcsolódó gazdasági idősor példáján.

A bevezetést követően először bemutatjuk a waveletek lényegét és kapcsolatukat más, többé-kevésbé ismert idősorelemzési módszerekkel, majd összefoglaljuk a waveletekkel szoros rokonságot mutató Fourier-analízis elemeit. Ezt követően soravesszük a wavelet elemzés alapjait, egy leggyakrabban alkalmazott módszerét, lehetséges továbbfejlesztéseit, alkalmazását, majd az említett számszerűen is kidolgozott feladatot. A cikket a fontosabb hivatkozások jegyzéke zárja.

## 1. A waveletekről általában

A waveletek<sup>2</sup> bizonyos matematikai követelményeket kielégítő függvények. E követelmények közül fontos megemlíteni azt, hogy az integráljuk 0 legyen, továbbá legyenek az időtartományban *lokalizáltak*, ami biztosítja, hogy az egyes időpontokhoz tartozó idősor-értékek jól leírhatók legyenek a waveletek segítségével. A gyakorlatban használt wavelet függvények emellett különböző számú eltűnő momentummal rendelkeznek.

A wavelet elemzés alapötlete, nevezetesen az, hogy az idősor alkalmas tulajdonságú függvények superponálásával közelíthető, nem új. Hasonló

<sup>1</sup>A szerző ehelyütt szeretne köszönetet mondani Lois Lászlónak (BME) a tanulmány alapos és értékes bírálatáért.

<sup>2</sup>A wavelet angol szó hullámocskát jelent. Mivel a magyar nyelvű szakirodalomban még nem honosodott meg a fogalom, nem látjuk célszerűnek lefordítani.

elemzési módszer a múlt század kezdetétől létezik, amikor *Fourier* megmutatta, hogyan lehet szinusz és koszinusz bázisfüggvényekkel más függvényeket reprezentálni. Mivel a wavelet elemzés a Fourier elemzéshez áll közel, és mivel ez utóbbi valamelyest ismert a közgazdászok idősorlemzési gyakorlatában, a továbbiakban a wavelet elemzés alapjait a Fourier-analízishez kapcsolva fejtjük ki.

A Fourier és wavelet módszerek közös jellemzője az, hogy lineáris transzformációk: a keresett függvényt a bázisfüggvények lineáris kombinációjával közelítik, illetve reprezentálják. A Fourier-transzformáció bázisfüggvényei szinusz és koszinusz függvények, ezért ezek csak a frekvenciatartományban lokalizáltak, az időtartományban nem. A waveletek, melyek hasonló bázisfüggvényekkel dolgoznak, időben (vagy térben) is lokalizáltak. A wavelet módszer fontos eleme a skála szerinti elemzés, ami annyit jelent, hogy a wavelet-algoritmusok az adattranszformálást különböző skálán azaz különböző felbontásban végzik. Ez nagymértékben alkalmas diszkrét jellegű, hirtelen megszakításokat tartalmazó, nem-stacionárius jelekből álló idősorok elemzésére. Mivel a wavelet elemzés nem trigonometrikus függvényekre épül, a nem periodikus idősorok reprezentálásához kevesebb wavelet bázisfüggvény kell, mint amennyi szinusz-koszinusz kellene Fourier-analízis esetén.

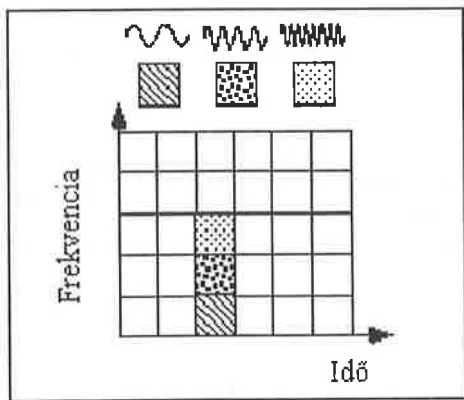
Ami a technikai megoldást illeti, jóllehet a Fourier-analízis elmélete már régen kialakult, az alkalmazások szempontjából az igazi áttörés azonban csak a számítógépek megjelenésével kezdődött. Számos algoritmust fejlesztettek ki a Fourier-analízis alkalmazásához, de a számítógépeknél a FFT (Fast Fourier Transform) diszkrét változata bizonyult a leghatékonyabbnak. A waveleteknél ennek a analógja a DWT (Discrete Wavelet Transform). Mivel mindkét módszer lényegi eleme az a transzformáció, amellyel az eredeti idősor adatait az új bázisba átvizsik, a továbbiakban az említett két transzformációt hasonlítjuk össze.

A FFT és a DWT sok *hasonlóságot* mutat. Mindkettő olyan lineáris művelet, ami  $\log_2 N$  (ahol  $N$  az idősor hossza, megfelelő számú 0-val kiegészítve) különféle hosszúságú együtthatókból álló adatstruktúrát generál, és melyet ez a művelet rendszerint kitölt és átalakít a különböző  $2^n$  ( $n = 0, 1, \dots, \log_2 N$ ) hosszúságú adatvektorokba. Mindkét transzformációban a bennük szereplő mátrixok hasonló matematikai tulajdonságokkal rendelkeznek; mind a FFT, mind pedig a DWT esetén a lineáris transzformáció *ortonormált*, így a transzformáció matrixa is ortonormált, azaz a matrix inverze megegyezik transzponáltjával. A FFT transzformáció olyan bázist használ, amely szinusz és koszinusz függvényeket tartalmaz. A wavelet transzformációnál a megfelelő bázisfüggvények, az ún. waveletek, anya waveletek vagy elemző waveletek, hasonlóak bár, de bonyolultabbak. A két transzformáció más szempontból is hasonló. Mivel a bázis függvények mindkét esetben frekvencia-lokalizáltak, az

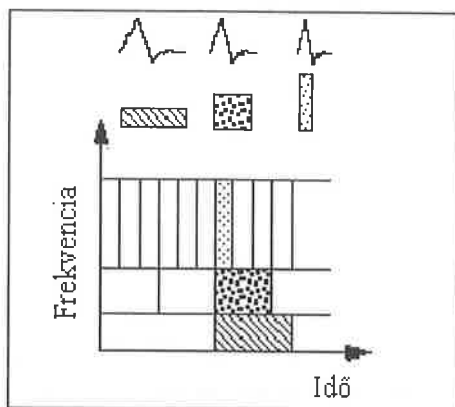
olyan, a frekvenciatartománybeli elemzésre épülő matematikai eszközök mint az energiaspektrum (periodogram) és skálagram (a későbbiekben definiáljuk), az elemzés lényeges részét képezik, és hasznosak lesznek a frekvencia kiválogatásánál, valamint az energiaeloszlás kiszámításánál.

A hasonlóságok mellett azonban a két módszer közt lényeges *eltérések* is vannak. Talán a leglényegesebb eltérés az, hogy szemben a Fourier-analízissel a wavelet függvények nem csak a frekvenciatartományon, de időben vagy térben is lokalizáltak. Ez a fajta lokalizáció, a wavelet frekvencia-lokalizálásával egyetemben, a wavelet tartományba való transzformáláskor sok, waveleteket használó függvényt és operátort ritkává tesz. Ez olyan hasznos alkalmazásokat eredményez mint az adatkompresszió, a kép sajtáságok felfedezése és a zaj eltávolítása az idősorokból. Jóllehet ezek a fogalmak elsődlegesen a jelfeldolgozásban játszanak komoly szerepet, az utolsó fejezetben éppen azt szeretnénk bemutatni, hogy maga a technika közgazdasági idősorok komponensekre való bontásában is jól alkalmazható.

Annak érdekében, hogy megmutassuk az idő-frekvencia felbontási különbségeket a wavelet és Fourier-transzformáció között, nézzük meg a bázis függvény lefedését az idő-frekvencia síkban. Az 1. ábrán a WFT-t (Window Fourier Transform) látjuk, ahol az ablak egy egyszerű négyzet hullám. A négyzet hullámú ablak úgy vágja le a szinusz vagy koszinusz függvényeket, hogy azok *beférjenek* a meghatározott szélességű ablakba. Mivel a WFT-ben minden frekvencia számára egy ablakot használunk, a felbontás azonos az idő-frekvencia sík minden helyén.



1. ábra: Fourier bázisfüggvények, idő-frekvencia mozaiklapok és az idő-frekvencia sík lefedése

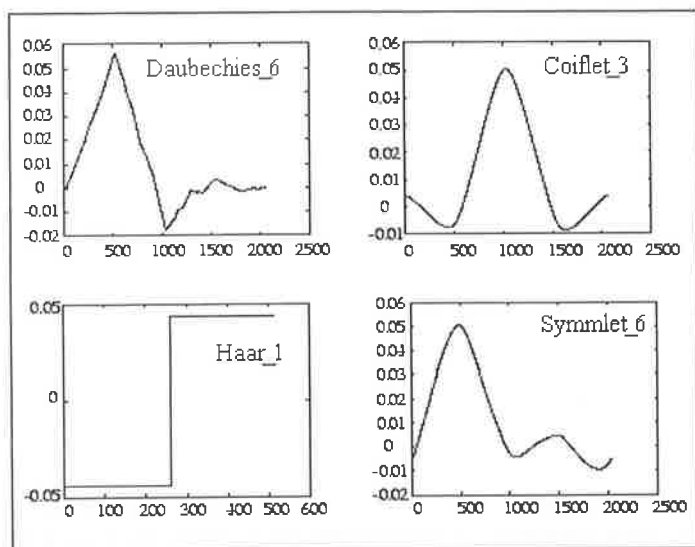


2. ábra: Daubechies wavelet bázis függvények, idő-frekvencia mozaiklapok és az idő-frekvencia sík lefedése.

A wavelet transzformáció előnye ezzel szemben az, hogy az ablakok változóak. Ahhoz hogy elkülönítsük a jel diszkontinuitását, nagyon rövid bázisfüggvényeket kellene alkalmazni. Ugyanakkor, a részletes frekvencia-analízishez nagyon hosszú bázisfüggvények kellene. A megoldás a rövid, magas frekvenciájú és hosszú, alacsony frekvenciájú bázisfüggvények egyidejű alkalmazása. Ezt a szerencsés eszközt a wavelet-transzformáció adja. A 2. ábra az idő-frekvencia sík lefedését mutatja be speciális, gyakran használt bázisfüggvények, a Daubechies wavelet bázisfüggvények segítségével.

A waveletek valójában tehát olyan alapfüggvények, amelyek lineáris kombinációival kikeverhetők az idősorok. A különböző alakú waveletek *családokat* alkotnak, ilyen pl. a már említett Daubechies család. Minden wavelet családon belül léteznek wavelet *osztályok*, melyek az együttthatók számával és az iterációs szinttel (lásd később) különbözhetnek meg. A wavelet-családon belül a waveleteket leggyakrabban az eltűnő momentumaik számával osztályozzák. Így például, a *Coiflet* wavelet családban léteznek a két és három eltűnő momentummal rendelkező Coiflet-ek. A 3. ábra néhány különböző wavelet családot mutat be.

A 3. ábrán látható waveletek közül a későbbiekben elsősorban a már említett Daubechies waveletekkel foglalkozunk. Az ugyanitt látható Haar-wavelet érdekessége az, hogy ez volt az első wavelet, amit a szakirodalom említett (1909), s ez első alkalmazója, Haar Alfréd magyar matematikus nevét viseli. A Haar-wavelet fontos tulajdonsága az, hogy kompakt tartójú, ami azt jelenti, hogy megszűnik egy véges időintervallumon kívül. Sajnos, a Haar-wavelet nem differenciálható folytonosan, ami korlátozza a felhasználását.



3. ábra: Néhány különböző wavelet család.

A wavelet név utáni szám az eltűnő momentumok számát fejezi ki.

## 2. A Fourier-analízis alapjai

A Fourier-analízis, mely már közgazdasági alkalmazásokban is létjogosultságot nyert, hatékony és jól bevált eszköz a stacionárius idősorok elemzésében. Ezért ahhoz, hogy a wavelet elemzés lényegét jobban megértsük, célszerűnek látjuk összefoglalni a Fourier-analízis alapfogalmait.

Ha az  $\{X_t, t = 0, \pm 1, \dots\}$  diszkrét stacionárius folyamat, akkor dekompozíciójához szinusz és koszinusz függvények lineáris kombinációja alkalmazható. Formálisan:

$$X_t = \int_{-\infty}^{\infty} (\cos \lambda t + i \sin \lambda t) dZ(\lambda) = \int_{-\pi}^{\pi} e^{i\lambda t} dZ(\lambda), \quad (2.1)$$

ahol  $\{Z(\lambda), -\pi \leq \lambda \leq \pi\}$  ortogonális folyamat. A folyamat változójának varianciája

$$\text{Var}\{X_t\} = \int_{-\pi}^{\pi} dF(\lambda), \quad (2.2)$$

ahol  $E\{|dZ(\lambda)|^2\} = dF(\lambda)$ . Az  $F(\lambda)$ -t a folyamat *spektrális eloszlásfüggvényének* nevezzük. Ha  $dF(\lambda) = f(\lambda) d\lambda$ , akkor az  $f(\lambda)$  a spektrális sűrűségfüggvény, vagy egyszerűen  $X_t$  *spektruma*. Az (2.2) összefüggés azt mutatja meg, hogy miképp kapcsolódik az  $X_t$  idősor varianciája a spektrum komponenseihez, maga a spektrum pedig azt, hogy az egyes frekvenciák milyen súllyal szerepelnek az  $X_t$  előállításában.

Ahhoz, hogy  $X_t$ -nek az  $f(\lambda)$  spektrumát megfigyelések alapján meghatározzuk (becsüljük), a

$$d(\lambda_j) = d_j = \frac{1}{(2\pi N)^{1/2}} \sum_{t=0}^{N-1} X_t e^{i\lambda_j t}, \quad j = 0, 1, \dots, [N/2], \quad (2.3)$$

alakú transzformációt kell elvégezni. Ebből kapható az elemzés egyik legfontosabb eredménye, a *periodogram*, mely az

$$I_j = |d_j|^2, \quad (2.4)$$

kifejezéssel definiálható, és a spektrum aszimptotikusan torzítatlan becslőfüggvényének tekinthető. A periodogram azt mutatja meg, hogy a vizsgált idősorban milyen frekvenciájú hullámösszetevők milyen súllyal fordulnak elő.

Az a tény, hogy a spektrum, feltételezve hogy az  $X_t$ -k  $\gamma_u$  autokovarianciái összeadhatók, az

$$f(\lambda) = \frac{1}{2\pi} \sum_{u=-\infty}^{+\infty} \gamma_u e^{-i\lambda u} \quad (2.5)$$

kifejezéssel írható le, azt mutatja, hogy  $f(\cdot)$  értékét az egyedi  $\lambda$  frekvencián olyan  $X_t$  értékek határozzák meg, amelyek a  $t$  széles tartományán szét vannak szórva. Ilymódon a spektrum komponensei az időben nem jól lokalizáltak, ami pedig azt jelenti, hogy néhány változás az  $X_t$  értékekben az összes  $d_j$  együtthatható megváltozásához vezethet.

### 3. Wavelet elemzés

A wavelet elmélet eredetét nehéz meghatározni. Kifejlődéséhez sok tudományterület járult hozzá. Az elmélet magja a jelfeldolgozás, a fizika és a matematika szintézise. A waveletek alkalmazási skálája széles és felhasználásuk rohamosan növekszik. De tulajdonképpen mi is a wavelet és miért olyan népszerű? Ezt a kérdést szeretnénk az alábbiakban kissé részletesebben megvizsgálni.

A wavelet elemzés alapja a  $\psi(x)$  bázis-, elemző- vagy anya-wavelet, ami egy speciális tulajdonságokkal rendelkező függvény. Maguk a waveletek pedig

az anya-waveletek dilatációinak (ez egyaránt jelent kitágítást vagy zsugorítást) és transzlációinak (áthelyezések, eltolások) véges halmazai.

Néhány anya-wavelet tulajdonsága az, hogy nullává válik bizonyos intervallumon kívül (kompakt tartójú), ezzel biztosítva az idősor komponensekre való gyors bomlását. Az  $\psi(x)$  elemző wavelet eltolásai (transzlációi) és skálázása (dilatációja) a wavelet ortogonális bázisát határozzák meg:

$$\psi_{(s,l)}(x) = 2^{-\frac{s}{2}} \psi(2^{-s}x - l),$$

ahol  $s, l$  egész számok. Ahhoz, hogy generáljuk pl. a Daubechies családhoz tartozó waveleteket, skáláznunk és dilatálnunk kell a  $\psi(x)$  elemző waveletet  $s, l$  számok segítségével. Az  $s$  skála index a wavelet szélességét jelzi, az  $l$  hely index pedig a wavelet pozícióját adja meg. Meg kell említeni, hogy az elemző wavelet függvényeket kettő hatványaival skálázzuk és egész számmal toljuk el. A dilatációval és transzlációval képzett waveletek *önhasonlósága* teszi a wavelet bázisokat különösen érdekessé. Ha ugyanis az elemző (bázis-) függvény egyértelműen meghatározott, akkor ezáltal már a bázisról is mindent tudunk.

Az elemző és skálázó függvények kapcsolatát az alábbi összefüggés adja meg:

$$\phi(x) = \sum_{k=-1}^{N-2} (-1)^k c_{k+1} \psi(2x + k),$$

ahol  $\phi(x)$  a  $\psi$  bázisfüggvény skálázó függvénye, és  $c_k$ -k a wavelet együtthatók. A wavelet együtthatóknak lineáris és kvadratikus követelményeket kell teljesíteniük a következő formában:

$$\sum_{k=0}^{N-1} c_k = 2, \quad \sum_{k=0}^{N-1} c_k c_l = 2\delta_{l,0},$$

ahol  $\delta$  a Kronecker-féle delta függvény és  $l$  a lokalizálási index. Megjegyezzük, hogy a második feltétel a transzformáció ortonormáltóságát hivatott biztosítani. Az elemző waveletek és a megfelelő skálázó függvények tehát egymással szoros kapcsolatban állnak: az elemző waveletek az adott skálázó függvényekből kaphatók meg.

Szerencsés, ha a kutató az adott wavelet rendszerhez úgy választhatja meg definiálódó együtthatóit, hogy azok jól illeszkedjenek az adott feladathoz. *Daubechies* olyan különleges családokat fejlesztett ki wavelet rendszerek számára, amelyek kiválóan alkalmasak polinomiális viselkedésű idősorok leírására.

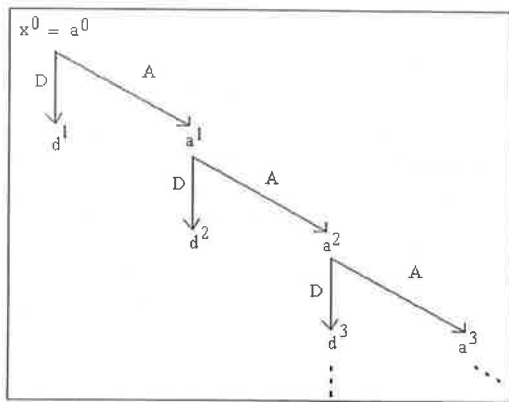
Daubechies a  $\{c_0, \dots, c_n\}$  együtthatókat szűrőnek tekintette. Ezeket az együtthatókat a nyers adatvektorra alkalmazandó transzformációs mátrixba helyezzük el. Az együtthatókat két domináns séma szerint rendezzük. Az

egyik a mozgó átlaghoz hasonlóan aluláteresztő (simító) szűrőként működik, és közelítő információkat ad. A másik (feluláteresztő szűrő) az adatok különbségi (egyedi) információit emeli ki. Az együtthatók ezen kettős rendezését a jelfeldolgozási irodalomban *QMF (Quadrature Mirror Filter)* párnak hívják.

A wavelet együtthatómátrixot hierarchikus algoritmusban alkalmazzuk, melyet néha *piramis algoritmusnak* neveznek. A wavelet együtthatókat úgy rendezzük el, hogy a páratlan sorok aluláteresztő szűrőt tartalmaznak, a páros sorokban pedig különböző előjelű wavelet együtthatók helyezkednek el, melyek az adatokból a különbségi (egyedi) információkat tartalmazzák. Az együtthatómátrixot először a teljes hosszúságú adatvektorra alkalmazzuk, majd kiválasztjuk a simított részt, és a mátrixot újra alkalmazzuk addig amíg triviális számú simított adat nem marad. Egy-egy mátrix ismételt alkalmazása az adatok magasabb szintű felbontását hozza ki, miközben egyidejűleg a megmaradó adatokat simítja. A DWT kimenete a megmaradó közelítő komponens(ek)ből és felhalmozott különbségi információkból áll.

#### 4. A DWT működésének főbb elemei

Ahhoz, hogy megértsük, mi is történik pontosan az adatokkal az egydimenziós DWT alatt, tekintsük át hogyan hat a DWT egy  $N = 2^n$  hosszúságú, egymástól  $\Delta t$  távolságban megfigyelt  $\mathbf{x}$  adatvektorra. A dekompozíció szematikusan a 4. ábrán látható.



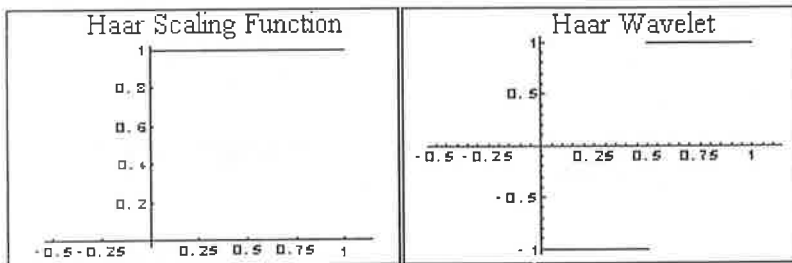
4. ábra: A DWT működési sémája



A DWT először az  $\mathbf{x}$  adatvektort két  $N/2$  hosszúságú,  $2\Delta t$  megfigyelési távolságú  $\mathbf{a}^1$ ,  $\mathbf{d}^1$  részvektorra vágja ketté. A  $\mathbf{d}^1$  vektor az  $\mathbf{x}$  vektor DWT-jának első eleme lesz, és a folyamatot az  $\mathbf{a}^1$ -ből folytatjuk tovább, hogy megkaphassuk  $\mathbf{a}^2$ -t és  $\mathbf{d}^2$ -t. Ezt ismételjük addig, ameddig az  $n$  lépésben az  $\mathbf{a}^n$  és  $\mathbf{d}^n$ -ben egy pont nem marad, és az  $\mathbf{x}$  vektort a  $\mathbf{d}^1$ ,  $\mathbf{d}^2, \dots, \mathbf{d}^n$ , és  $\mathbf{a}^n$  vektor helyettesíti. Minden szinten a két részvektor hosszának összege megegyezik a szülővektor hosszával, ily módon a  $\mathbf{d}$  vektorhossz összege plusz a végső  $\mathbf{a}^n$  pont ugyanakkora, mint az eredeti vektor  $N$  hossza. Mivel az első szinten  $\mathbf{x}$  vektor az  $\mathbf{a}$  szerepét játssza, célszerű néha az  $\mathbf{x}=\mathbf{a}^0$  értelmezés. Az  $\mathbf{a}^k \rightarrow \mathbf{a}^{k+1}$  transzformációt  $A$ -val, az  $\mathbf{a}^k \rightarrow \mathbf{d}^{k+1}$  transzformációt pedig  $D$ -vel jelöljük. Fontos megjegyezni, hogy az  $A$  és a  $D$  ortogonálisak. A dekompozíció  $A$  fázisa lényegében az adatok *átlagolásának* felel meg az előző szinten. A frekvencia tartományban ez ekvivalens az előző szinten történő alacsony frekvenciájú szűréssel. A dekompozíció  $D$  fázisa lényegében a folyó szint *átlagolt adatai és az előző szint adatai különbségének* felel meg. Ez ekvivalens az előző szinten való magas frekvenciájú szűréssel.

Mivel az  $A$  és a  $D$  transzformáció *ortogonális*, az inverz algoritmus jól kondicionált. Ez azt jelenti hogy a transzponált matrixokat a mindenkori  $\mathbf{a}$ -ra és  $\mathbf{d}$ -re alkalmazva, az előző szinten levő  $\mathbf{a}$  vektor visszaállítható. (Megjegyezzük, hogy az eddig mondottak érvényesek az FFT algoritmusra is.)

Az  $A$  transzformáció mögött álló  $\phi(t)$  függvény a *skaláló függvény*, a  $D$  mögötti  $\psi(t)$  függvény pedig az *anya- vagy elemző wavelet*. Skaláló függvények és elemző waveletek csak az összekötő párokban fordulnak elő. Néhány párosítás az 5. és a 6. ábrákon látható.



5. ábra: Skaláló és elemző Haar-wavelet

Az algoritmus a  $\psi(t)$  waveletet a  $\phi(t)$  skaláló függvényből állítja elő. A rájuk vonatkozó követelmények:

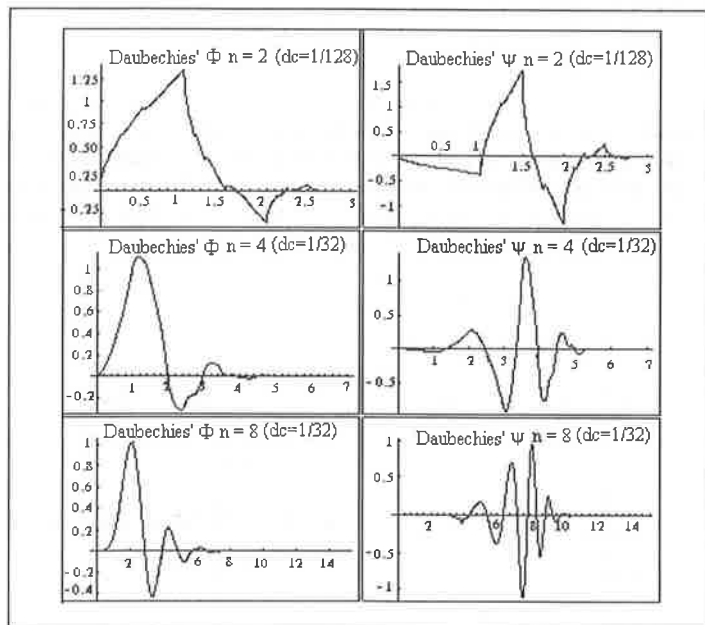
$$\int \phi(t) dt = 1 \quad (4.1)$$

és

$$\int \psi(t) dt = 0 \quad (4.2)$$

konzisztensek az átlagolás és a különbség értelmezésének megfelelően. A továbbiakban a skálázó függvények és az elemző waveletek néhány alapvető tulajdonságáról lesz szó.

Mint azt már említettük, a skálázó függvények és elemző waveletek az idő- és a frekvencia tartományban egyaránt lokalizáltak. A gyakorlatban használt skálázó függvények és elemző waveletek  $1/t$ -nél nagyobb sebességgel bomlanak az időben és egyidejűleg  $1/f$ -nél nagyobb sebességű bomlással rendelkeznek a frekvencia tartományban.



6. ábra: Skálázó és elemző Daubechies waveletek

A waveletekkel kapcsolatban fontos fogalom a többszörös felbontású elemzés. Feltételezzük hogy a

$$\phi(t - k) \quad k = \dots, -2, -1, 0, 1, 2, \dots$$

eltolások az adattartomány ortonormált bázisát alkotják (egy egység a  $k$ -ban  $\Delta t$ -nek felel meg). Ekkor a

$$\phi_{jk}(t) = 2^{-j/2} \phi(2^{-j}t - k) = 2^{-j/2} \phi(2^{-j}(t - k2^j)), \quad (4.3)$$

$j, k = \dots, -2, -1, 0, 1, 2, \dots$ , eltolások  $2^j$  nagyságú translációs lépessel bázist alkotnak a  $j$ -edik átlagolt tartomány számára. Bár az eltoló  $\phi$ -k ortonormáltak egy-egy szinten belül (rögzített  $j$ ), nem ortonormáltak a szintek között.

$$\int \phi_{jk}(t) \phi_{j'k'}(t) dt = \delta_{k,k'} \quad (4.4)$$

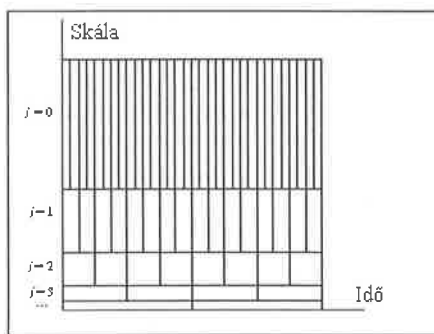
Ahhoz, hogy megkaphassuk a szintek közti ortonormált bázist, forduljunk a különbségi függvényhez - a wavelethez. Valóban, a  $\psi(t)$  elemző wavelet

$$\psi_{jk}(t) = 2^{-j/2} \psi(2^j t - k), \quad j, k = \dots, -2, -1, 0, 1, 2, \dots, \quad (4.5)$$

skalázott eltolásai az  $L^2$  ortonormált bázisát alkotják.

$$\int \psi_{jk}(t) \psi_{j'k'}(t) dt = \delta_{j,j'} \delta_{k,k'}. \quad (4.6)$$

Most elérkeztünk a wavelet elemzés egyik legfontosabb állításához, miszerint a DWT az adatelemzést különböző skálákon (oktávonként) végzi. Mivel a  $2^j$  translációs lépés a szinthez (oktávhoz) kapcsolódik, a magas frekvenciájú események ( $j$  közel áll 0-hoz) kevés translációs lépés alatt értelmezhetőek. Alacsony frekvenciájú események viszont ( $j \gg 0$ ) nagyobb időablakot követelnek. A wavelet bázis idő-frekvencia síkja a 7. ábrán látható.



7. ábra: Wavelet bázis idő-frekvencia síkja

Mivel a wavelet transzformáció komponenseit  $j$ -vel (ami skálát vagy frekvencia-szintet ad) és  $k$ -val (ami lokalizációt ad) indexeljük, ez azt jelenti

hogy a wavelet transzformáció az adatokat egyidejűleg az idő-frekvencia tartományban ábrázolja. Ez azt is jelenti, hogy az eredetileg egydimenziós tömböt két dimenzióssá transzformáljuk.

A (4.2) egyenlet szerint az elemző wavelet integrálja (nulladik momentuma) eltűnik. A gyakorlatban használt elemző waveletek legtöbb esetben több (mondjuk  $M$ ) eltűnő momentummal rendelkeznek:

$$\int t^m \psi(t) dt = 0, \quad m = 0, 1, 2, \dots, M - 1.$$

Az eltűnő momentumok nagyobb száma a wavelet és skálázó függvény nagyobb simítottságával áll szoros kapcsolatban.

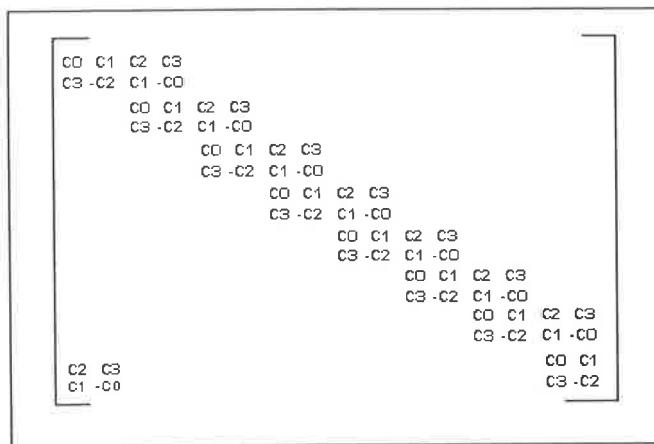
## 5. A Daubechies-féle Daub4 transzformáció

A Daub4 waveletet, amely a legkisebb a Daubechies családban, egyszerűsége miatt demonstrálásra és egyszerű alkalmazásokhoz gyakran használják. Ez a legkönnyebben megtanulható és a legjobban lokalizált wavelet transzformáció négy wavelet együtthatón alapszik. Daubechies a következő négytagú "bűvös" számhalmazt (együtthatókat) alkalmazta:

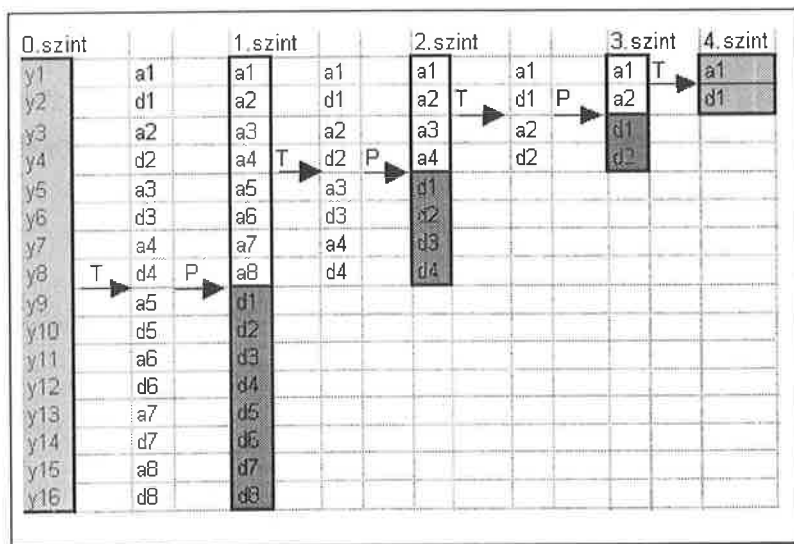
$$[c_0, c_1, c_2, c_3] = \frac{1}{4\sqrt{2}} [1 + \sqrt{3}, 3 + \sqrt{3}, 3 - \sqrt{3}, 1 - \sqrt{3}].$$

Ezekből a számokból építjük fel a  $(2n \times 2n)$  méretű transzformációs mátrixot a bemenő adatvektor hosszához igazítva. A 8. ábrán a mátrix felépítése, a 9. ábrán pedig az algoritmus végrehajtása során kialakult adatvektorok szerkezete, és egyben a transzformáció lényege látható.

A transzformációs mátrixot (ez esetben a Daub4-et) az input  $(y_1..y_{16})$  vektorra alkalmazzuk. Így megkapjuk a közelített (simított)  $(a_1..a_8)$  adatvektort, amely át van lapolva az egyedi hatásokat tartalmazó különbségi  $(d_1..d_8)$  adatokkal. Annak érdekében, hogy szétválasszuk a közelített és a különbségi adatokat, az eredményvektort átrendezzük (permutáljuk). A különbségi adatokat tároljuk, a közelítőket (simítottakat) pedig tovább transzformáljuk. Minden ismétlés a simított adatokat osztja két részre. Az eljárás bármelyik ponton félbeszakítható, de általában addig folytatjuk, amíg csak két pont nem marad. A végleges kimenet pontosan annyi adatpontot tartalmaz, amennyi a bemeneten volt.



8. ábra: A Daub4 transzformációs mátrix

9. ábra: A transzformáció végrehajtása ( $T$  – transzformálás,  $P$  – permutálás)

## 6. A waveletek alkalmazása

Közgazdasági idősorok elemzésénél klasszikus probléma a komponensekre való bontás: a trend, a szezonális komponens, az üzleti ciklus és a szabálytalan (véletlen) komponens szétválasztása. Mivel a dekompozíció nem egyértelmű és a komponensek sem függetlenek egymástól (mint ahogy azt a hagyományos dekompozíciós idősormodellek feltételezik), az egyes összetevők identifikálása nem egyszerű. Gazdasági idősorokban a szezonális komponenseket általában nem nehéz megtalálni. Először is többnyire jellemző az állandó periódus, ami havi adatok esetében 12 hónap. Másodszor, bizonyos feltételezések vagy az elmélet segít meghatározni az idősor szezonálisát. A hosszabb távú (egy évnél hosszabb ingadozású) ciklusok felkutatása már nehezebb feladat. Ezt a komponenset csak a szezonális komponens és a véletlen ingadozások kiküszöbölésével találhatjuk meg. Ugyancsak nehézséget jelent az egyes komponensek szétválasztásánál az, ha a szezonális komponens nem egészen szabályosan alakul. Mivel az idősorelemzés végső célja az előrejelzés, a dekompozíciót is ennek rendeljük alá. Általában ugyanis könnyebb előrejelezni az idősor egyes komponenseit mint az egész idősort. A wavelet elemzés ebben a dekompozícióban nyújt hasznos segítséget.

Az adatelemzés szempontjából a waveletek az adathalmaz által generált függvények ábrázolását segítik elő. Az adott  $2^n$  méretű  $\mathbf{y} = (y_0, \dots, y_{2^n-1})$  adatvektor a  $[0, 1]$ -en, ami az

$$f(x) = y_k, \quad x \in [k/2^n, (k+1)/2^n] - \text{el} \quad (6.1)$$

definiált  $f$  függvényvel értelmezhető. Ez az  $f$  függvény négyzetesen integrálható és wavelet dekompozíciója

$$f(x) = c_{00}\phi(x) + \sum_j d_{jk}\psi_{jk}(x),$$

alakú, ahol  $\phi(x)$  (lehet periodikus is) egy skálázó függvény.

Ha adott egy  $\mathbf{a} = (a_0, \dots, a_{N-1})$  sorozat, akkor ennek diszkrét Fourier-transzformáltja a

$$b_j = \sum_{l=0}^{N-1} a_l e^{-i(2\pi j/N)l}, \quad j = 0, 1, \dots, N-1 \quad (6.2)$$

által definiált másik  $\mathbf{b} = (b_0, b_1, \dots, b_{N-1})$  sorozat. A  $\mathbf{b}$  sorozat annak a sorozatnak a lineáris kombinációja, amelyik az időtartományból a frekvencia-tartományba vezet. Az  $\mathbf{a}$  és a  $\mathbf{b}$  a diszkrét Fourier-transzformáció *Fourier-eggyütthetői*. Hasonlóan, ha adott egy  $\psi_{jk}$  wavelet bázis, akkor az  $N = 2^n$  dimenziójú  $\mathbf{y}$  adatvektorhoz rendelünk hozzá egy másik ugyanilyen dimenziójú

$\mathbf{d}$  vektort (diszkrét wavelet transzformáció), amelyik  $\mathbf{y}$  vektorunkat az idő-és frekvenciatarományban ábrázolja. A  $\mathbf{d}$  vektor azokat az együtthatókat tartalmazza, amelyek az  $\mathbf{y}$ -t a kiválasztott wavelet bázisból előállítják. A  $\mathbf{d}$  elemei a

$$\mathbf{d} = (c_{00}, d_{00}, d_{10}, d_{11}, d_{20}, \dots, d_{n-1, 2^{n-1}}) \quad (6.3)$$

elrendezésben helyezkednek el, ahol  $c_{00}$  a skálázó függvény együtthatója. A  $c_{00}$ -án kívül van még egy másik 0-ás szintű  $d_{00}$  együttható, két 1-es szintű együttható ( $d_{10}, d_{11}$ ) és általában  $2^j$  számú  $j$  szintű  $d_{j0}, d_{j1}, \dots, d_{j, 2^j-1}$  együttható. (Az utolsó szint az  $(n-1)$ ). Legvégül kiszámítjuk a  $\mathbf{d}$  skálagramját, ami a wavelet analógja a Fourier-analízisben szereplő periodogramnak. Ha a  $\mathbf{d}$  az  $\mathbf{y}$  diszkrét wavelet transzformációjának együtthatóvektora, akkor a  $\mathbf{d}$  "energiát" a  $j$ -edik szinten az

$$E(j) = \sum_{k=0}^{2^j-1} d_{jk}^2, \quad j = 0, \dots, n-1 \quad (6.4)$$

képlet definiálja.

A  $\mathbf{d}$  skálagram a

$$(c_{00}^2, E(0), E(1), \dots, E(n-1)) \quad (6.5)$$

módon definiált energia-vektor. A DWT  $\mathbf{d}$  skálagramja az idősor különböző frekvenciájú ciklusokra való bontásához használható.

A gazdasági idősorokban legegyszerűbb esetben a 12 havi szezonális komponenst és a hosszútávú trendet keressük. Ezért célszerű egy  $\mathbf{x}$  gazdasági adatsor wavelet együtthatóinak skálagramjában két csúcspontot keresni. A  $2^n$  dimenziójú  $\mathbf{x} = (x_t)$  idősor dekompozíciója során leggyakrabban két ugyanolyan dimenziójú adathalmaz kapható:  $\mathbf{y} = (y_t)$ , és  $\mathbf{z} = (z_t)$ . Mindkettő külön-külön ( $x_t$ ) ingadozásait ábrázolja különböző frekvenciákon.

Ha az  $\mathbf{x}$  wavelet dekompozíciójának  $j$ -edik alacsony szintjén a legtöbb  $d_{jk}$  együttható nagy, akkor ez azt jelenti hogy hosszú periódusú, alacsony frekvenciájú komponens van jelen az  $\mathbf{x}$ -ben. Ha ellenben a  $j$ -edik magas szinten a legtöbb együttható nagy, akkor rövid periódusú, magas frekvenciájú komponenst találunk.

Vágjuk most ketté a  $\mathbf{d} - t$   $\mathbf{d}^{(1)}$ -re és  $\mathbf{d}^{(2)}$ -re, majd mindkét részre alkalmazzuk a  $\psi^{-1}$  inverz wavelet transzformációt. Ezáltal  $\mathbf{x}$  két komponensét azonosítottuk. Sőt, feltételezzük, hogy a  $\mathbf{d}$  skálagram két csúcspontja a  $j_1 < j_2$  szinten van. Osszuk most a szintek  $\{0, 1, 2, \dots, n-1\}$  halmazát a két  $\mathbf{A} = \{0, 1, \dots, j_1\}$  és  $\mathbf{B} = \{j_1 + 1, \dots, n-1\}$  részhalmazra úgy, hogy az  $\mathbf{A}$  halmazban levő szintek a  $j_1$  (első csúcspont szintje) körül, a  $\mathbf{B}$  halmaz szintjei pedig a  $j_2$  (második csúcspont szintje) körül helyezkedjenek el. Ekkor a  $\mathbf{d}^{(1)}$  és  $\mathbf{d}^{(2)}$

a következőképpen definiálható:

$$\mathbf{d}^{(1)} = (c_{00}, d_{00}, d_{10}, d_{11}, \dots, d_{j0}, \dots, d_{j,2^{n-1}-1}, 0, \dots, 0) \quad (6.6)$$

$$\mathbf{d}^{(2)} = (0, \dots, d_{j+1,1}, \dots, d_{j+1,2^{j+1}-1}, \dots, d_{n-1,2^{n-1}-1}) \quad (6.7)$$

és a komponensek, amelyekre  $\mathbf{x}$ -et bontottuk, a következők:

$$\mathbf{y} = \psi^{-1}\mathbf{d}^{(1)} \quad \text{és} \quad \mathbf{z} = \psi^{-1}\mathbf{d}^{(2)} \quad (6.8)$$

Általában az  $\mathbf{y}$  a gazdasági adatsor hosszú távú trendjét képviseli, ami gyakran összefügg más, a gazdaság üzleti ciklusait tükröző mutatókkal is. A  $\mathbf{z}$  gazdasági idősorok esetén leggyakrabban a 12 havi szezonális viselkedést követi, a szezonális ingadozás jellemző komponensének tekinthető. Természetesen, amennyiben a skálagramnak több csúcspontja is van, az idősorban több ciklus is azonosítható.

## 7. Illusztratív példa

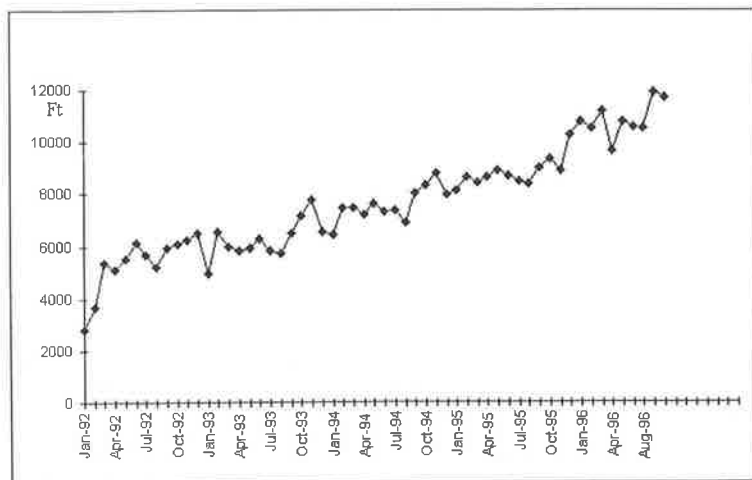
Az alábbiakban azt mutatjuk meg, hogy hogyan lehet DWT és a skálagram segítségével egy idősor komponenseit megtalálni és szétválasztani. A vizsgált idősor a távközlési piacon kialakult átlagos számlaértéket ábrázolja havi gyakorisággal 1992 és 1996 között (57 hónap). Az idősor, melyen a DWT működését demonstráljuk, a 10. ábrán látható.

Az egyszerűség kedvéért a megadott időorból kivonjuk annak átlagát, elosztjuk a szórásával, és a megmaradt zéro-átlagú és egységnyi varianciájú idősorhoz (11. ábra) alkalmazunk diszkrét wavelet transzformációt.

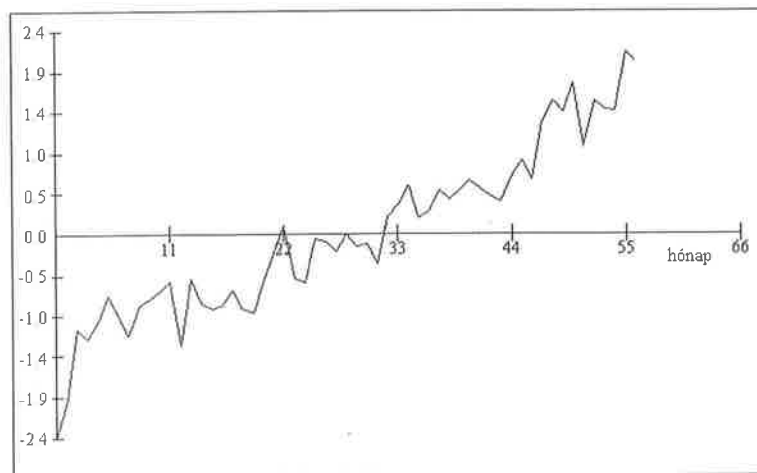
Az elemzéshez először wavelet bázist kell választani; legyen ez a DAUB/4. A bemenő adatok és az eredmények egy részét (első 24 megfigyelés) az 1. táblázatban mutatjuk be. Mivel a DWT-hez a kiinduló adatsor hossza  $2^n$  kell hogy legyen, a hiányzó számok 0-val pótolhatók. Így feladatunkban a DWT munka- és eredményvektor hossza 64 lett.

A "fogyar" oszlopban az eredeti adatsor, a "z-score"-ban a standardizált adatsor, a "Daub4DWT"-ben a DWT együtthatók, a "Band33-64"-tól a "Band1-4"-ig az együtthatókból visszanyert részletes idősorok található "frekvenciasáv" bontásban. Az utolsó 5 oszlop adatai az együtthatók oszlopaiból az inverz transzformációval kaphatók. A kiinduló adatsor (ebben az esetben a z-score) az utolsó 5 oszlop soronkénti összege. A skálagram, mely a "Daub4DWT" oszlopból (6.3 képlet) számítható ki a (6.5) képlet szerint, a 12. ábrán látható. A  $c_{00}$ -ás szint (állandó komponens) az ábrán nem szerepel.





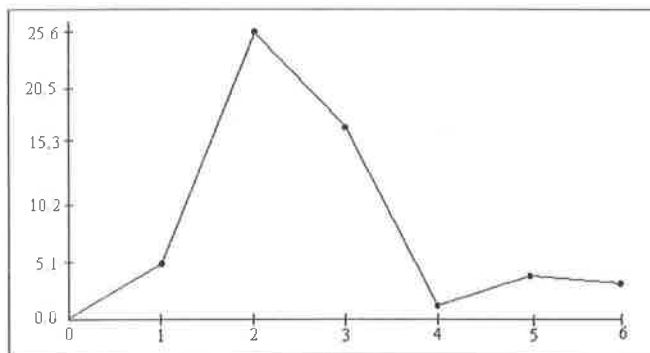
10. ábra: Az átlagos számlaérték alakulása



11. ábra: A standardizált számlaérték idősorának alakulása

ho	month	fogyar	"z-score"	"Daub4 DWT"	"Band 33-64"	"Band 17-32"	"Band 9-16"	"Band 5-8"	"Band 1-4"
1	Jan-92	2784	-2.39528	-3.0364E-07	-0.437469268	-0.616778002	0.01240274	-1.1201808	-0.2372561
2	Feb-92	3700	-1.94273	2.192150081	-0.45146374	-0.332716768	-0.12224553	-0.7301934	-0.3061107
3	Mar-92	5345	-1.12283	3.218991505	0.271007485	-0.291471107	-0.22629278	-0.5038855	-0.3721899
4	Apr-92	5120	-1.23498	-3.895122795	-0.060896159	0.002409894	-0.29973901	-0.441257	-0.4354937
5	May-92	5540	-1.02564	1.380779799	-0.038540934	0.22859746	-0.38138476	-0.3347708	-0.4995412
6	Jun-92	6130	-0.73157	0.093319788	0.242247489	0.081116546	-0.10216009	-0.3919641	-0.5608133
7	Jul-92	5550	-0.97081	0.230637471	-0.056815113	0.033759798	0.08036964	-0.4052995	-0.622829
8	Aug-92	5205	-1.19261	-3.88682671	-0.237383302	-0.061065875	0.16620444	-0.3747772	-0.6855884
9	Sep-92	5905	-0.84372	0.683840079	0.125661334	-0.143172288	0.27794857	-0.3560066	-0.7481486
10	Oct-92	6070	-0.76148	-0.088244205	-0.022800061	0.116153289	0.13194572	-0.1788441	-0.8079333
11	Nov-92	6220	-0.68672	0.197943759	-0.113129778	0.28399244	0.05500597	-0.0441226	-0.8684617
12	Dec-92	6480	-0.55713	-0.142365139	0.312670303	-0.035350664	0.04712931	0.0481579	-0.9297338
13	Jan-93	4955	-1.31722	-0.09611798	-0.274812519	-0.224153676	0.02074725	0.1518105	-0.9908066
14	Feb-93	6525	-0.5347	0.677306438	0.274027931	0.010065402	0.02080961	0.2130221	-1.0526231
15	Mar-93	5955	-0.8188	-0.082665889	-0.134884296	0.130936052	0.0137862	0.2856057	-1.1142403
16	Apr-93	5830	-0.8811	0.17304763	-0.087489797	0.012810726	-0.00032296	0.3695613	-1.1756582
17	May-93	5910	-0.84123	0.495450913	-0.000756411	-0.04127582	-0.0125335	0.4504698	-1.2371296
18	Jun-93	6260	-0.66678	-0.147593094	0.172269257	-0.102790817	-0.04528699	0.2422272	-0.9331976
19	Jul-93	5825	-0.89359	-0.395695829	-0.033026779	-0.162315367	-0.07253602	0.1114624	-0.7271752
20	Aug-93	5735	-0.92845	-0.125647549	0.188646371	-0.084634805	-0.09428057	0.0581754	-0.6190623
21	Sep-93	6476	-0.55912	0.135214665	0.102684061	-0.043718241	-0.11750004	-0.0158717	-0.4847146
22	Oct-93	7155	-0.22069	-0.661675785	0.037789952	0.241848623	-0.03961556	-0.012441	-0.4482765
23	Nov-93	7764	0.082842	-0.10987374	0.025176483	0.461861636	0.01117821	-0.0297705	-0.3856035
24	Dec-93	6534	-0.53021	0.195455946	-0.155182806	-0.045354986	0.03488124	-0.06786	-0.2966958

1. táblázat: Kiinduló adatok és DWT transzformáció



12. ábra: Skálagram

ho	fogyar	z-score	Daub4 DWT	scalogram	trendlow	Daub4 InvDWT	sezonhigh	Daub4 InvDWT
1	2784	-2.399281	-3.0364E-007	9.2197E-014	-3.0364E-007	-1.357436968	0	-1.041844521
2	3700	-1.94273	2.192150081	4.80552198	2.192150081	-1.036304136	0	-0.9064260214
3	5345	-1.122832	3.218991505	25.5650649	3.218991505	-0.8760753992	0	-0.2467564255
4	5120	-1.234976	-3.899122795	17.075877	-3.899122795	-0.8767507581	0	-0.3582252858
5	5540	-1.02564	1.380779799	1.03963567	1.380779799	-0.8343119945	0	-0.1913282434
6	6130	-0.731573	0.093319788	4.12766886	0.093319788	-0.9527773267	0	0.2212039393
7	5650	-0.970814	0.230637471	2.04349377	0.230637471	-1.028128536	0	0.05731433675
8	5205	-1.19261	-3.88682671		-3.88682671	-1.060365623	0	-0.1322447417
9	5905	-0.843718	0.683840079		0	-1.104155105	0.68384008	0.2604376227
10	6070	-0.761478	-0.088244205		0	-0.986774222	-0.0882442	0.2252989531
11	6220	-0.686716	0.197943759		0	-0.9125843511	0.19794376	0.2258686344
12	6480	-0.557127	-0.142365139		0	-0.8815758914	-0.14236514	0.3244489588
13	4955	-1.317215	-0.09611798		0	-0.8389961499	-0.09611798	-0.4782189552
14	6525	-0.534698	0.677306436		0	-0.8396010199	0.67730644	0.3049029427
15	5955	-0.818797	-0.082665889		0	-0.8286346081	-0.08266589	0.00983795207
16	5830	-0.881099	-0.17304763		0	-0.8060969146	-0.17304763	-0.07500203277
17	5910	-0.841225	0.495450913		0	-0.7866597367	0.49545091	-0.05456573095
18	6260	-0.666779	-0.147593094		0	-0.6909704432	-0.14759309	0.02419144097
19	5825	-0.883591	-0.395695829		0	-0.6157128425	-0.39569583	-0.2678781691
20	5735	-0.928449	-0.125647549		0	-0.5608869346	-0.12564755	-0.3675617523
21	6476	-0.559121	0.135214665		0	-0.5005863712	0.13521466	-0.05853422667
22	7155	-0.220694	-0.661675785		0	-0.4607175005	-0.66167578	0.240023027
23	7764	0.0828424	-0.10987374		0	-0.4153739742	-0.10987374	0.4982163442
24	6534	-0.530212	0.195455946		0	-0.3645557923	0.19545595	-0.1656565537
25	6450	-0.572079	-0.412630169		0	-0.31520454	-0.41263017	-0.256874948

2. táblázat: Az idősor komponenseinek szétválasztása

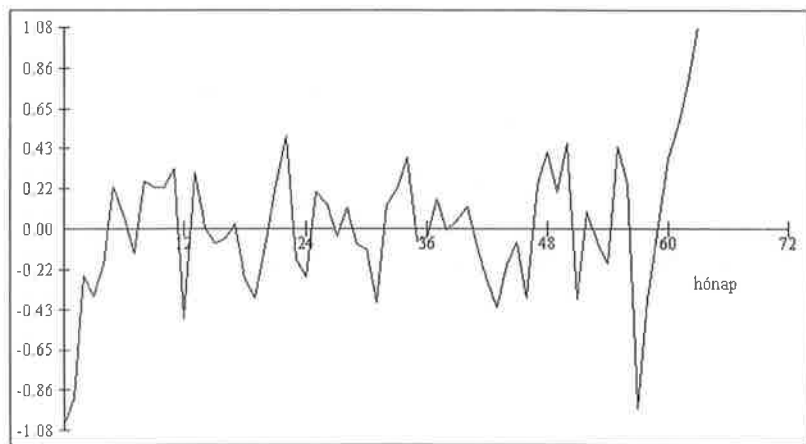
A grafikonon két csúcspont vehető észre: a 2 és 5 szinten. Ekkor a (6.6)-os és a (6.7)-es magyarázat szerint  $d$ -t felosztjuk  $d^{(1)}$ -re és  $d^{(2)}$ -re. A  $d^{(1)}$ -be 0,1,2,3 szintű együtthatók a  $d^{(2)}$ -be pedig 4,5,6 szintű együtthatók kerülnek. Az így felosztott két együtthatóhalmazra külön-külön alkalmazunk inverz DWT transzformációt. Az eljárás lebonyolításának módja a 2. táblázatban követhető. (A táblázat ezúttal is csak az első 24 értékre vonatkozó eredményeket mutatja.)

A TRENDLOW oszlopban a  $d^{(1)}$  a SEZONHIGH oszlopban a  $d^{(2)}$  együtthatók találhatók. Ezekből kaphatók inverz transzformáció után azok az adatsorok, amelyek a táblázatban közvetlenül a megfelelő együtthatóoszlop után következnek. A Daub4InvDWT nevű 7. oszlop a számított hosszútávú komponens, a 9. oszlop pedig a szezonális ingadozások standardizált, becslött idősorát mutatja. (A kettő összege minden időpontban kiadja a z-score 3. oszlopban található megfigyeléseit.) A szezonális normalizált idősor a 13. ábrán látható. Érdeemes figyelni arra, hogy a szezonális komponens ez esetben —eltérően a szokásostól— nem mutat teljesen szabályos, ismétlődő formát.

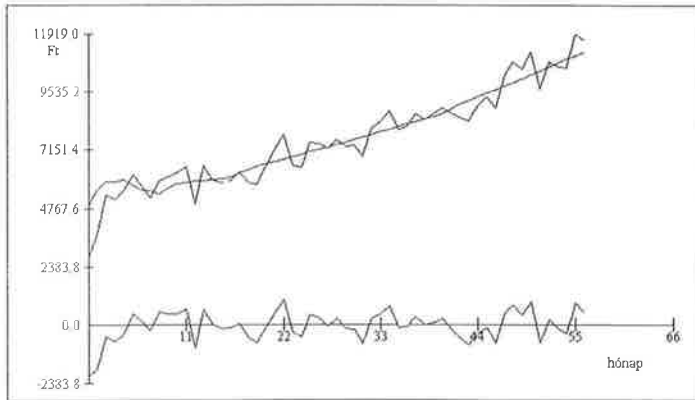
Az eredeti idősrora visszatérve (visszafelé alkalmazva a standardizálást) most már megkaphatók az  $x_t$  idősor azonosított komponensei. A  $x$ ,  $y$  és  $z$  összevont diagramja a 14. ábrán látható, ahol a felső tört görbe az eredeti adatsort (FOGYAR), simított görbe a trendet és az alsó tört görbe a szezonális komponenst ábrázolja. Ez valójában az elemzés végeredménye.

A diagramból jól látható, hogy a fogyasztói ár növekvő tendenciát mutat és minden évben ismétlődő, bár nem teljesen szabályos ingadozásoknak van kitéve.

Ezzel a kis bevezető jellegű példával azt mutattuk meg, hogy a wavelet analízis alkalmas az idősorok belső struktúrájába való betekintésre, és ezáltal az elemzés, valamint az előrejelzés hasznos eszköze lehet.



13. ábra: A standardizált változó szezonális idősrora



14. ábra: Az átlagos számlaérték idősorának komponensei

## Irodalom

1. Arino, M. A.: Time series forecasts via wavelets: an application to car sales in the Spanish market. Discussion Paper 95-30, ISDS, Duke University, 1996.
2. Arino, M. A. and Vidacovic B.: On wavelet scalograms and their applications in economic time series. Discussion Paper, 95-21, ISDS, Duke University, 1995.
3. Daubechies, I.: Ten Lectures on Wavelets. SIAM, CBMS-NST Conference Series, 61, 1992.
4. Graps, A.: An Introduction to Wavelets. IEEE Computational Science and Engineering. Vol. 2, num. 2, 1995.
5. Morettin, P. A.: From Fourier to Wavelet Analysis of Time Series. Department of Statistics, University of Sao Paulo, 1996.
6. Pintér I.: Beszédjelek perceptuális wavelet-reprezentációja és zajos beszéd tisztítása, Híradástechnika, XLV. 1994, szept.

## WAVELETS AND THEIR APPLICATION IN ANALYZING TIME SERIES

Wavelets are mathematical functions that cut up data into different frequency components, and then study each component with a resolution matched to its scale. They have advantages over traditional Fourier methods in analyzing physical situations where the signal contains discontinuities and sharp spikes. This paper introduces wavelets to the interested economic persons outside of the digital signal processing field. We shortly mention about prehistory of wavelets, compare wavelet

transform with the well known Fourier transform. For the purpose of best understanding we introduce effective Daubechies algorithm for Fast Wavelet Transform (FWT) with application to telecommunication time series. We use scalogramm to separate the final result into low and high frequency components.