STATISZTIKAI MÓDSZERTANI FÜZETEK, 43

SZEZONÁLIS KIIGAZÍTÁS

BUDAPEST, 2005

ISSN 0324-5985
ISBN 963 215 827 X

Készült:
a KSH Statisztikai kutatási és oktatási főosztályának
Mintavételi és módszertani osztályán

Főosztályvezető:
dr. Szép Katalin

Összeállította:
Bauer Péter
Földesi Erika

Közreműködött:
Berki Natália
Fábián László

Lektorálta:
dr. Sugár András

E–mail: marketing.ksh@office.ksh.hu

Információszolgálat
Telefon : (36–1) 345–6789; Fax: (36–1) 345–6788
Internet: http://www.ksh.hu

TARTALOMJEGYZÉK

1. BEVEZETÉS

A gazdasági vagy társadalmi folyamatok vizsgálatához gyakran idősorokat használnak, amelyek a folyamatok időbeli alakulását írják le. Az idősorok viselkedését nagymértékben befolyásolhatják olyan tényezők, amelyek különböző évek azonos időszakaiban (például hónap vagy negyedév) azonos irányban és közel azonos mértékben hatnak az idősor alakulására. Ilyen tényezők lehetnek az időjárás, a különféle adminisztratív hatások vagy a társadalmi hagyományok. Ezeket a tényezőket együttesen szezonális hatásnak nevezzük. Az elemzők gyakran a folyamatok olyan jellemzőire kíváncsiak, amelyeket a nagymértékű szezonális hatás elfed, ezért szükség van ennek eltávolítására. A szezonalitás kiszűrését szezonális kiigazításnak nevezzük.

Ebben a kiadványban arra vállalkozunk, hogy a szezonális kiigazítás módszertanával kapcsolatos fő kérdésekről, a használt módszerekről és a KSH szezonális kiigazítási gyakorlatáról átfogó, ugyanakkor részletes ismertetést adjunk, amely mindazok számára jól használható, akik a KSH által publikált szezonálisan vagy munkanappal kiigazított adatokat igényesen, az adatok előállításánál alkalmazott módszerek ismeretében kívánják felhasználni.

Jelenleg a két legelterjedtebb szezonális kiigazító módszer az X12-ARIMA és a TRAMO/SEATS. A kiadványban részletesen ismertetjük mindkét módszert, valamint a korábbi, helyenként még használatos X11-ARIMA módszert is.

Beszámolunk a folyamatról, melynek során a TRAMO/SEATS módszert választottuk a KSH hivatalos szezonális kiigazító módszerének, és a választást befolyásoló tényezőkről, szempontokról, így a szezonális kiigazítással kapcsolatos nemzetközi elvárásokról és gyakorlatról is.

Végül bemutatjuk a KSH szezonális kiigazítási gyakorlatát, és annak kialakításánál felmerülő választási lehetőségeket, a választások indokait.

A második fejezet egyes részei (2.5 és 2.8-2.11-ig) a sztochasztikus idősorelemzésben járatosak számára ajánlottak, a többi fejezet alapfokú statisztikai ismeretekkel is megérthető.

2. SZEZONÁLIS KIIGAZÍTÁSI MÓDSZEREK

Ebben a fejezetben a történeti áttekintés után röviden ismertetjük az idősorelemzés alapvető fogalmait. Először a dekompozíciós idősormodellek segítségével bemutatjuk az idősorok komponenseit, azok meghatározásának módszereit, majd ezen túlmutatva, röviden ismertetjük a későbbiekben tárgyalásra kerülő X11-ARIMA, X12-ARIMA és TRAMO/SEATS eljárásokat, amelyek azonban már a sztochasztikus elemzéshez tartoznak. A következő fejezetekből pedig részletesen megismerhetjük a sztochasztikus folyamatok és a módszerek jellemzőit.

2.1. A szezonális kiigazítás módszereinek fejlődése – rövid történeti áttekintés[1]

(A fejezetben használt szakkifejezések magyarázatára a következő részekben kerül sor.)

A 19. század végén Angliában, a csillagászat és a meteorológia területén tevékenykedő kutatók fogalmazták meg először, hogy egy megfigyelt idősor több, meg nem figyelhető tényező összhatásának az eredménye. A korabeli kutatók úgy gondolták, hogy a két változó közötti „hamis" korrelációt a trend okozza, amelyet ezért előbb ki kell szűrni az idősorból. Poynting (1884) és Hooker (1901) a trendet és a szezonális hatásokat az árak több évre történő átlagolásával próbálta kiszűrni. Spencer (1904) és Andersen (1914) mutatta be magasabb fokú polinomok alkalmazását a trend leválasztására. A közgazdaságtan területén tevékenykedő kutatók a gazdasági ciklus kimutatása érdekében választották le a szezonális hatásokat és a trendet.

Az 1920-as és 1930-as években kezdődött meg az igazán aktív kutatás a szezonális kiigazítás területén Person (1919) munkája következtében, amelyben felírta a híres képletet, ami szerint (multiplikatív modellt

[1] Fischer (1995) alapján.

feltételezve) egy idősor kifejezhető a következő formában:

$$X_t = S_t \times T_t \times C_t \times R_t$$

ahol

S_t a szezonális ingadozást leíró komponens;

T_t a hosszútávú trend;

C_t a középtávú ciklus;

R_t a véletlen tényező.

Person módszere állandó szezonális tényezőket használt, annak ellenére, hogy a kor statisztikai irodalmában több publikáció szólt arról, hogy a fix szezonalitás feltételezése sok esetben nem helytálló. Sydensticker és Britten (1922) vezették be formális szezonális kiigazító modellben a változó szezonális tényezőket. Crum 1925-ben módosította Person módszerét, hogy alkalmassá váljon a változó szezonalitás kezelésére.

A első átfogó szezonális kiigazító rendszert Macauley (1931) dolgozta ki. A módszer három alaplépésből áll:

– havi adatokat feltételezve 12 tagú centrírozott mozgóátlag leválasztása és ezek átlagolása után fix szezonális tényezők előállítása;
– a trend becslése lineárisan vagy magasabb fokszámú polinom illesztésével;
– a mozgóátlagok osztása a trend becsült értékével a ciklikus komponens értékének becslését adja.

Ezt a módszert nevezik ma klasszikus dekompozíciónak, amely sok modern szezonkiigazító eljárásnak képezi alapját, így a legelterjedtebb X11-nek is.

Az 1950-es évek során két fontos újítás történt. Az első az exponenciális simító eljárások elterjedése volt, amelyekkel az addigi számítások mennyiségét jelentősen csökkenteni lehetett, legalább ugyanolyan jó eredmények mellett, mint a korábban használt módszerekkel. A másik fontos fejlődést a számítógépek megjelenése jelentette, ami szintén a számítások gyorsaságának növekedését segítette elő. Arra, amire korábban több nap kellett, most pár másodperc is elég volt. A kutatók így sokkal bonyolultabb modelleket dolgozhattak ki, hiszen az új verziókat könnyen tesztelhették nagy számú idősoron is.

A Census I. módszer 1954-ben jelent meg; ez annyival lépett túl a Macauley-féle eljáráson, hogy az idősort egyszerű extrapolálással előre-hátra meghosszabbította, hogy pótolja a mozgóátlagolás során elvesztett elemeket. Ennek továbbfejlesztett változata a Census II. (1955), amely az akkor használatos technikák elektronikus verziója volt – kidolgozásában Julius Shishkin játszott úttörő szerepet.

A Census II-t több kritika is érte, mégpedig azzal kapcsolatban, hogy a módszer sok tekintetben ad hoc jellegű eljárásokra épül, amelyeket nem támaszt alá semmilyen statisztikai elmélet – a szezonális ingadozásoknak csak a kiugró részeit szűri ki, a többit benne hagyja stb. A modell folyamatos felülvizsgálatát követően alakult ki 1965-re az X11-es változat, amely a közelmúltig a legelterjedtebb szezonális kiigazító eljárás volt. Ennek legfőbb újdonsága a munkanaptényező regressziós módszerrel történő kezelése. Ezen kívül az új változatban a felhasználó választhat az additív és a multiplikatív modell között és meghatározhatja, hogy milyen mozgóátlagot kíván használni.

Box és Jenkins az autoregresszív mozgóátlagolásra irányuló, 1970-es években folytatott kutatásainak hatására fejlődött ki 1980-ra a kanadai statisztikai hivatalnál az X11-ARIMA változat, ami abban különbözött elődeitől, hogy az idősor előre és hátra történő meghosszabbítását az ARIMA-modellezés segítségével végezte el. Ez a becslés minőségét javította. Az újabb X11-ARIMA/88 és X11-ARIMA/2000 verziók főleg diagnosztikákkal bővítették az addigi módszertant. Az 1990-es évek közepén jelent meg az X12-ARIMA program, amely új alapokra helyezte a munkanapok, ünnepnapok és outlierek hatásának elemzését és a hiányzó adatok pótlását; valamint tovább bővítette a diagnosztikák körét.

A mozgóátlagolású technikák – amelyekről tudjuk, hogy erősen ad hoc jellegűek – mellett kifejlődtek modellszemléletű megközelítések is, amelyek között érdemes megkülönböztetni a determinisztikus és sztochasztikus jellegű modelleket. A determinisztikus modellek a trendet és a szezonalitást egy előre elrendelt pályának tekintik; amelyre a véletlen csak olyan módon gyakorol hatást, hogy eltéríti az idősor tényleges értékét ettől a pályától. A determinisztikus jellegű modellek a regressziószámításra épülnek, amely a trendet és a szezonalitást valamilyen determinisztikus módon megadott függvényekkel kezeli. Ebbe a modellcsaládba tartoznak a DAINTIES és a BV4 programok.

A sztochasztikus módszerek a véletlennek jelentős hatást tulajdonítanak, ez a modellezésben fontos szerepet játszik. Ezek története Yule autoregresszív (1927), illetve Slutsky mozgóátlagolású modelljéig (1937) nyúlik vissza. Wold alkalmazta először a mozgóátlagolású modellt valós adatokra, illetve ő dolgozta ki a vegyes ARMA-modellek használatát (1954). Azonban a számítások nehézkessége miatt az ARMA-modelleket csak nagyon kevesen használták, egészen a számítógépek széles körű elterjedéséig, illetve amíg Box és Jenkins meg nem fogalmazta azokat a kritériumokat, amelyekkel minden idősorra meghatározható egy konkrét típusú és fokú ARIMA-modell. Ennek általános képlete szezonális idősorra:

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D z_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t,$$

ahol B a késleltetési operátor ($Bz_t = z_{t-1}$), így $\nabla z_t = (1-B)z_t$, továbbá $\phi_p(B), \Phi_P(B^s), \theta_q(B), \Theta_Q(B^s)$ a nemszezonális és szezonális autoregresszív és mozgóátlag polinomok, ε_t pedig fehér zaj folyamat.

A log (0,1,1)(0,1,1) modell (az ún. Airline-modell, ami arról kapta a nevét, hogy egy légitársaság adatait használták a módszer bemutatására), viszonylag kevés paraméterrel jól illeszthető sok idősorra.

A modell első gyakorlati megvalósítása az Angol Nemzeti Bankban történt a '80-as években. Ezt fejlesztették tovább a Spanyol Nemzeti Bankban Augustin Maravall irányítása alatt, ennek eredménye a TRAMO-SEATS program. Annak ellenére, hogy a program egy teljesen új megközelítést alkalmaz, az utóbbi években nagyon széles körben elterjedt.

2.2. A szezonális kiigazítás alapjai

Az idősorelemzés alapfogalmairól a Hunyadi-Mundruczó-Vita (1996) könyvben olvashat részletesebben.

Az idősorelemzés típusai

Az idősorelemzés során olyan időpontra vagy időszakra vonatkozó adatok sorozatát vizsgáljuk, amelyek szabályos időközönként (havonta, negyedévente) állnak rendelkezésünkre. Ezek az adatok lehetnek alapadatok, vagy az alapadatokból képzett indexek is.

Az idősorok vizsgálatánál különbséget kell tennünk a determinisztikus és a sztochasztikus elemzés között. A determinisztikus modell abból indul ki, hogy az idősorok egy előre meghatározható (determinált) pályát követnek, amely hosszú távon nem változik Ezek a modellek a véletlen tényező létét elfogadják, de igyekeznek hatását kiszűrni.

Ezzel szemben a sztochasztikus elemzés a véletlen tényezőt a folyamat alapvető részének tekinti. A rövid távú hatásokat vizsgálja és valószínűségi változókat alkalmaz a folyamat modellezéséhez.

Az idősorok komponensei és kapcsolódási lehetőségei

A megfigyelt idősoroknak **négy fő komponens**ét különböztethetjük meg:

Trend ($\hat{y}$): a hosszú távú alapirányzatot mutatja. A trendet meghatározhatjuk az analitikus trendszámítás segítségével, amikor a tartós irányzatot függvényszerűen írjuk le, vagy mozgóátlagok segítségével. Mindkét esetet később részletesen megvizsgáljuk.

Szezonális komponens (s, s^*): a trendtől való eltérés, a rövid távú (éven belüli), szabályos ingadozás mértéke.

Ciklikus komponens (c,c^*): a szabálytalan hosszabb távú ingadozást mutatja.

Véletlen tényező (ε,ν): előre nem jelezhető hatások, véletlen változók, amelyről feltételezzük, hogy várható értékük 0 vagy 1.

Az egyes komponensek többféleképpen kapcsolódhatnak egymáshoz. Összegszerű kapcsolódás esetén **additív modell**ről beszélünk:

$y = \hat{y} + s + c + \varepsilon$

Ez a modell abból indul ki, hogy a trend és az eredeti adatok különbsége az azonos időszakokban (hónapokban, negyedévekben) közel állandó, függetlenül attól, hogy a trend milyen tendenciát mutat, azaz a szezonális eltérés mértéke nagyjából megegyezik.

Multiplikatív modellnek nevezzük a dekompozíciót, amikor szorzatszerűen kapcsolódnak egymáshoz a tényezők:

$y = \hat{y} \cdot s^* \cdot c^* \cdot \nu$

A szorzatszerű összekapcsolódásnál abból indulunk ki, hogy a trend és az eredeti adatok eltérésének nem a mértéke, hanem a trendhez viszonyított aránya tekinthető állandónak, azaz a szezonindexek nagysága lesz közel állandó. Ilyenkor növekvő trend esetén a „kilengések" is egyre nagyobbak, míg csökkenő trend esetén ezek mértéke is csökkenő.

Logadditív modellről beszélhetünk, ha az idősor logaritmusára írunk fel additív modellt. Multiplikatív módon dekomponált idősor logaritmusát felírhatjuk a komponensek logaritmusainak összegeként. Ily módon a multiplikatív modell visszavezethető a logadditív modell esetére, ezért ezt a két utóbbi modelltípust gyakran nem különböztetik meg.

Az angol statisztikai hivatalban fejlesztették ki a **pszeudoadditív** modelltípust, amely akkor használatos, amikor a felbontás alapvetően multiplikatív, de az idősor zéró vagy zéróhoz közeli értékeket is felvehet:

$y= \hat{y} \cdot (s + \nu - 1)$

A modelltípust csak az X családba tartozó szezonális kiigazítási módszerek támogatják, a TRAMO-SEATS nem.

A ciklikus komponens számításával a továbbiakban részletesen nem foglalkozunk, hiszen ezt csak nagyon hosszú idősorok esetén érdemes külön vizsgálni. A továbbiakban – az általános gyakorlatnak megfelelően – a trendbe a ciklikus komponenst is beleértjük, mind a determinisztikus, mind a sztochasztikus elemzésnél.

Trendtípusok az analitikus trendszámításban

Az analitikus trendszámítás során leggyakrabban **lineáris trendfüggvény**t alkalmaznak az alapirányzat meghatározására, mert rövid időszakokra jól illeszkedik, és ennek kiszámítása, értelmezése a legegyszerűbb. Ebben az esetben az adatok állandó mértékű változását feltételezzük.

A lineáris trend mellett nemlineáris összefüggést is feltételezhetünk az idősorról. Ebben az esetben számolhatunk **exponenciális trend**et, ha az alapadatok változásának állandó ütemét feltételezzük, vagy polinomiális trendet, illetve ennek speciális esetét a parabolikus trendet is.

Mozgóátlagolású trend

A mozgóátlagolású trendszámítás lényegében átmenetet képez a determinisztikus, függvények segítségével meghatározott trend és a sztochasztikus, valószínűségeken alapuló trendszámítás között. Mozgóátlagolás esetén a trendet nem egy képlettel meghatározott függvény jelenti, hanem az eredeti adatok átlagolásával kapjuk meg a trend értékeit. Ugyanakkor ebben az esetben – a sztochasztikus elemzéssel szemben – a trend és a többi komponens meghatározása nem kapcsolódik szorosan össze, nem egyszerre, hanem egymás után történik kiszámításuk.

A t. időszak trendértékét a t. érték és a környezetében lévő elemek átlagaként számítjuk ki. Az átlagolással egyrészt „simítjuk az idősort”, azaz csökkentjük a véletlen tag szerepét, másrészt a t értékének változtatásával az alaptendenciát próbáljuk követni.

A mozgóátlag tagszámának meghatározásánál figyelembe kell vennünk az idősor éven belüli szezonalitásának alakulását. Ha az éven belül periodicitás figyelhető meg, akkor ennek megfelelően kell megválasztani a tagszámot. Azaz például szezonalitást tartalmazó negyedéves adatok esetében 4 tagú mozgóátlagot, havi adatok esetében 12 tagú mozgóátlagot célszerű választani. Amennyiben az idősor semmilyen periodicitást nem tartalmaz, akkor szabadon határozhatjuk meg a tagszámot.

A szezonális komponens

A determinisztikus idősorelemzésnél a szezonalitás hatását állandónak tekintjük, azaz az idősor teljes hosszán az azonos időszakokban ugyanakkora a szezonalitás nagysága. Az állandó szezonalitás feltétele a sztochasztikus elemzésnél már nem teljesül, ott az ún. mozgó szezonalitást is figyelembe veszik a módszerek.

A szezonális komponens számítási módjánál alapvető kérdés annak eldöntése, hogy az idősorelemzést additív vagy multiplikatív modellel végezzük. Szerencsés esetben az idősor adataiból is látható, hogy melyik modell írja le jobban az idősor viselkedését. Amennyiben az idősor elemei között additív kapcsolódás van, akkor a szezonális hatást számszerűsítő komponenst **szezonális eltérés**nek nevezzük. Multiplikatív modell esetén pedig **szezonindex**et számolhatunk.

A szezonálisan kiigazított idősor

A meghatározott komponensek segítségével számítható ki az úgy nevezett **szezonálisan kiigazított idősor**. Ebben az esetben az eredeti adatokat „tisztítjuk meg” a szezonális hatástól, azaz minden egyes értékből kivonjuk az adott időszakhoz (hónaphoz, negyedévhez) tartozó szezonális eltérés értékét, vagy osztunk a szezonindex értékével. Az így kapott idősor tehát a trendet és a véletlen tényezőt tartalmazza.

Az outlierek

Outliernek (kiugró értéknek) az olyan adatot tekintjük, amely nem illeszkedik a megfigyelt idősor tendenciájába, kívül esik a trend és a szezonális tényező általános mintája alapján várható értékhatárokon. Az outliernek tekintett érték hátterében általában valamilyen egyszeri gazdasági, vagy társadalmi esemény áll, amelynek hatására az idősor adatai hosszabb vagy rövidebb időre eltérnek a korábbi tendenciától. Ilyen esemény lehet például az ipari termelés esetén, egy iparágon belül egy meghatározó üzem bezárása, amelynek hatására a termelés színvonala visszaesik, hiszen a továbbiakban az adat csak a megmaradt üzemek termelését mutatja. Hasonló hatású lehet például egy új jogszabály megjelenése, egy új támogatási forma, vagy adó bevezetése.

Az előzőekben bemutatott determinisztikus módszerek nem veszik figyelembe az outliereket, a sztochasztikus módszereken alapuló szezonális kiigazító módszerek azonban igen.[2] Ugyanakkor ha a kiigazítás során nem vesszük figyelembe a kilógó értékeket, akkor mind a trendre, mind pedig a szezonális tényezőre torz becslést kapunk.

Az outlierek több fajtáját különböztetjük meg, de ezek közül a három leggyakrabban alkalmazott: (1. ábra)

---

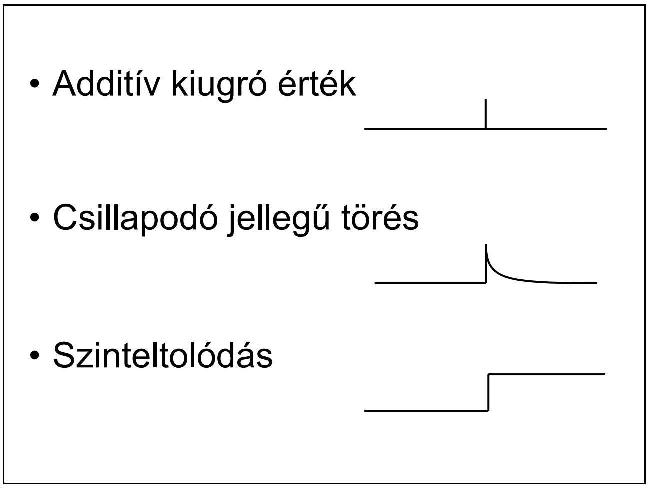
[2] A meghatározás az X11-ARIMA program esetén igaz. Az X12-ARIMA és a TRAMO/SEATS program esetében az outlierek meghatározása a komponensekre bontás előtt történik. Ott az eredeti idősorra illesztett modelltől jelentősen eltérő értékek tekinthetőek outliernek.

Additív outlier (AO): a hatás csak egy megfigyelés értékét befolyásolja.

Csillapodó jellegű törés (Temporary change, TC): egy megfigyelés értéke kiugró, majd a következő megfigyeléseknél fokozatosan (exponenciálisan) csökken a kiugrás mértéke, míg végül az idősor visszaáll az eredeti szintre.

Szinteltolódás (Level Shift, LS): egy időszaktól kezdve az összes megfigyelés értéke ugyanannyival ugrik ki a korábbi értékekhez képest, azaz az idősor szintje tartósan megváltozik.

1. ábra: A kiugró értékek típusai



A szezonálisan kiigazított adatsor mindhárom outliertípust tartalmazhatja, a trendbe azonban csak a szinteltolódás kerül bele, mert ennek hatása hosszú távon érvényesül.

Munkanaphatás

Az idősor alakulására az is hatással lehet, hogy egy-egy időszakban (például hónapban) mennyi a munkanapok száma. A problémát az okozza, hogy a munkanapok száma nemcsak időszakról időszakra változhat, hanem még a különböző évek azonos időszakaiban is különbözhet, így közönséges szezonális hatásként nem kezelhető.

A munkanapok számát az adott időszakban lévő, nem hétvégére eső ünnepek száma is befolyásolja. Mivel az egyes országokban a nemzeti ünnepek eltérően alakulnak, ezért mindenhol az adott ország ünnepnapjait célszerű figyelembe venni.

A munkanapok száma nem befolyásolja minden idősor adatát. Például egy olyan iparágban, ahol az üzemek folyamatosan termelnek függetlenül attól, hogy munkanap vagy szabadnap van, ott a munkanapok száma nincs hatással a megfigyelt értékekre.

A húsvét hatásának kiemelt kezelésére is szükség lehet bizonyos idősoroknál (például kiskereskedelmi forgalom), mert a húsvét mozgó ünnep: lehet márciusban vagy áprilisban is, és hatása befolyásolhatja a húsvétot megelőző egy hetes időszakot is.

Ha a megfigyelt idősor értékeiből csak a munkanapok és a húsvét hatását szűrjük ki, de a szezonalitást nem, akkor a **munkanappal kiigazított adatsor**t kapjuk.

2.3. Az X szezonális kiigazítási módszer-család röviden

Az X család legrégebbi még ma is használt tagja az X11 módszer. A komponensekre bontást ismételt mozgóátlagolásokkal valósítja meg. Ez a dekompozíciós rész gyakorlatilag változatlan a család legújabb tagjánál, az X12-ARIMA-nál is. Az eljárás képes kezelni a komponensek additív, multiplikatív, logadditív és pszeudo-additív összekapcsolódását is.

A dekompozíció lényege az, hogy először egy mozgóátlagolás segítségével egy kezdeti trendbecslést állítunk elő, a trend leválasztása után a szezonalitást is mozgóátlagolással becsüljük, és így kapunk egy kezdeti szezonálisan kiigazított idősort. A második lépésben ebből az idősorból kiindulva egy ún. Henderson-trendszűrő segítségével (ami egy speciális mozgóátlagolású szűrő) nyerjük a trend második becslését, majd újra megbecsüljük a szezonalitást, és megkapjuk a szezonálisan kiigazított idősor második, és egyben végső becslését. A harmadik lépésben a trendet még egyszer megbecsüljük a végső szezonálisan kiigazított idősorból Henderson-szűrő segítségével.

A leírt eljárás problémája, hogy az idősorok elején és végén a mozgóátlagoláshoz olyan értékekre is szükség van, amik nem állnak (még) rendelkezésre. Az X11-ARIMA és X12-ARIMA eljárásoknál a hiányzó adatokat úgy pótolják, hogy ARIMA-modell segítségével előrejelzik, illetve „visszajelzik" az idősort.

Az outlierek kezelése az X11-módszer esetében meglehetősen ad hoc jellegű, a becsült irreguláris komponens alapján történik. Az X12-ARIMA az ún. regARIMA-módszert (regresszió ARIMA-hibataggal) használja az outlierek becslésére. A módszer az idősorra egy ARIMA-hibataggal rendelkező lineáris regressziót becsül, ahol a magyarázó változók a megfelelő outliereket leíró változók. Az outlierek automatikus keresése bonyolult iterációs eljárással történik.

Az X12-ARIMA esetében a munkanaphatás, húsvéthatás, szökőnaphatás és egyéb felhasználó által definiált hatások kezelése is a regARIMA módszerrel történik, a hatást leíró változó magyarázóváltozóként a regresszióban szerepel.

Az X család, különösen az X12-ARIMA számos diagnosztikát bocsát a felhasználó rendelkezésére. A legfontosabbak: a szezonalitás szignifikanciájának vizsgálata varianciaanalízis segítségével; az M-statisztikák, amelyek a szezonális és irreguláris komponensre vonatkozó stabilitási mutatók; a Q-statisztika, amely az M-statisztikák lineáris kombinációja; és a Sliding spans és Revision Analysis, amelyek a szezonális kiigazítás eredményeinek stabilitását vizsgálják.

2.4. A TRAMO/SEATS módszerről röviden

A TRAMO és a SEATS programot a Spanyol Nemzeti Bankban fejlesztették ki. A TRAMO (**T**ime Series **R**egression with **A**RIMA Noise, **M**issing Observations and **O**utliers – Idősorregresszió ARIMA-zajjal, hiányzó megfigyelésekkel és outlierekkel /kiugró értékekkel/) és a SEATS (**S**ignal **E**xtraction in **A**RIMA **T**ime **S**eries – Jelkinyerés ARIMA idősorokban) programot együtt használhatjuk idősorok szezonális kiigazítására. A TRAMO által végzett előkészítés és a SEATS által végzett felbontás is teljes mértékben modell alapú, szemben például az X12-ARIMA eljárással.

A TRAMO egy olyan regressziós modellt illeszt az idősorra, ahol a hiba tag egy ARIMA folyamat, és automatikusan azonosítható a modell és becsülhetők a paraméterei. A regressziós változókat megadhatja a felhasználó, vagy a program generálja. A program által generált változók lehetnek például a munkanaphatás, húsvéthatás változók, illetve az outliereket leíró változók.

A TRAMO teszteli a munkanaphatás és a húsvéthatás meglétét, és ennek megfelelően kiigazítja az idősort. Meghatározza az outlierek helyét, típusukat (additív, szinteltolódás, vagy csillapodó jellegű törés) és megtisztítja hatásuktól az idősort. Amennyiben az ARIMA-modellt automatikusan határozzuk meg, akkor az outlierek detektálása és az ARIMA-modell és paramétereinek becslése párhuzamosan, iteratív módon történik. A TRAMO a hiányzó megfigyelésekhez kísérleti (nagyon nagy vagy nagyon kicsi) értéket rendel és additív outlierként kezeli.

Az így megtisztított (linearizált) idősort kapja meg a SEATS, amely ARIMA-modell alapú módszert alkalmaz az idősor nem megfigyelt komponensekre bontásához. Ezek a komponensek a trend, a ciklikus, a szezonális és az irreguláris komponens.

Feltételezi, hogy a komponensek is ARIMA-modellt követnek. A komponensekre való felbontás frekvencia tartományban történik: a modell spektrumát felbontja a különböző komponensek spektrumainak összegére. Ezzel meghatározza a komponensekre az ARIMA-modelleket, majd megbecsli a komponensek értékeit.

Végül bevonja a TRAMO által kiszűrt outliereket és naptári hatásokat (munkanap- és húsvéthatás) a különféle komponensekbe, és így kapja az úgynevezett végső komponenseket a következőképpen. Az additív és csillapodó jellegű törés outliereket az irreguláris komponensbe, a szinteltolódás outliereket a trend komponensbe illeszti, a naptári hatások a szezonális komponensbe kerülnek. Így a szezonálisan kiigazított idősorban (ami a trend komponens és az irreguláris komponens összege, ha a modell additív, illetve szorzata, ha multiplikatív) benne lesznek az outlierek is.

Lényeges megérteni, hogy milyen sok függ a TRAMO által kiválasztott (vagy általunk megadott) ARIMA-modelltől. Ettől függnek az outlierek, a munkanaphatás, illetve az idősor komponensekre bontása is. Az ARIMA-modell, illetve paramétereinek megváltoztatásán kívül nincs mód a komponensekre bontást befolyásolni, ellentétben például az X12-ARIMA-val, ahol a komponensekre bontáshoz kiválaszthatjuk a megfelelő szűrőt.

2.5. Bevezetés a sztochasztikus módszerekhez

A determinisztikus idősorelemzést azért nevezik determinisztikusnak, mert az idősornak a vizsgálat szempontjából lényeges komponensei determinisztikusak, azaz a múltbeli értékek alapján a jövőbeli értékek pontosan megjósolhatók. Ilyenkor a véletlen hatását egy külön összetevőbe tömörítjük és lehetőleg minél jobban elimináljuk a becslések, illetve az előrejelzések során. Ez a komponens fogja képezni a becslések maradékát – a *reziduumokat* – és egyben a becslések hibáját is a determinisztikus modellben.

A sztochasztikus idősorelemzésnél viszont a vizsgálat szempontjából lényeges komponensekben is figyelembe vesszük a véletlen hatását, a komponenseket egy-egy sztochasztikus folyamat realizációinak tekintjük.

Az alábbiakban ismertetendő X11-ARIMA és X12-ARIMA, de főképpen a TRAMO/SEATS elszakad a determinisztikus idősorelemzéstől és már sztochasztikus módszereket használ. Míg az X család ad hoc jellegű dekompozíciós módszere viszonylag könnyen megérthető mélyebb sztochasztikus idősorelemzési ismeretek nélkül is, addig a TRAMO/SEATS erősen támaszkodik az ARIMA-modellekkel kapcsolatos ismeretekre.

A következőkben felsoroljuk a következő részek megértéséhez szükséges alapfogalmakat, alapismereteket, feltételezve, hogy ezeket az olvasó ismeri, vagy utánanéz a – gyakorlatilag kizárólag angol nyelven hozzáférhető – szakirodalomban. Irodalomként javasoljuk Brockwell és Davis (1996), illetve Hamilton (1994) könyvét. Magyarul Tusnády és Ziermann (1986) könyve (ARIMA-modellek nem, csak ARMA-modellek szerepelnek), valamint néhány egyetemi jegyzet elérhető. A KSH módszertani füzetek (nemzetközi módszertani füzetek, ill. ökonometriai füzetek) sorozatban is megjelent néhány idősorokkal, illetve szezonális kiigazítással foglalkozó kötet. Halabuk és mások (1964) kifejezetten a szezonális kiigazítás korszerű módszereivel (Census II) foglalkoztak. Hrubos és mások (1968) különféle szezonális kiigazítási eljárásokat hasonlítottak össze. Hulyák (1977) a szezonális kiigazítás mellett ARIMA-modellekről és előrejelzésről is ír. Freschl és mások (1982) idősorelemzéssel és szezonális kiigazítással is foglalkoztak.

Szükséges fogalmak:

Sztochasztikus folyamat, stacionárius folyamat.

Autokovariancia függvény, autokorreláció függvény, ezek becslései.

Fehérzaj, autoregresszív (AR) folyamat, mozgóátlag (MA) folyamat.

Autoregresszív-mozgóátlag (ARMA) folyamat.

Autoregresszív integrált mozgóátlag (ARIMA) folyamat.

Néhány jelölésbeli konvenció, amit követünk: a visszatolás (avagy visszaléptetés, backshift, késleltetés, lag) operátort *B* betűvel jelöljük (a másik elterjedt jelölés az *L* betű). Az idősort jelölésben nem különböztetjük meg a sztochasztikus folyamattól, amelyből realizálódott: általában x_t-vel vagy y_t-vel jelöljük. Az innovációkat (vagy sokkokat) a_t-vel jelöljük.

Szükséges tudnivaló még a szezonális ARIMA-modellek fogalmának ismerete, mivel ez a fogalom a későbbiek folyamán gyakran előfordul, ezért most itt definiáljuk.

Amennyiben a vizsgált idősor szezonális, akkor a szezonális késleltetésekhez (12 megfigyelés havi adatoknál) tartozó autokorreláció szignifikánsan eltér nullától. Közönséges ARIMA-modellekkel az ilyen idősorok nem jellemezhetők jól, mert ezeknek az autokorrelációknak a figyelembevételéhez túl magas (12 vagy több) autoregresszív, illetve mozgóátlag rendű modellt kellene becsülni. A megoldást az ún. szezonális autoregresszív integrált mozgóátlag (szezonális ARIMA vagy SARIMA) modell jelenti. Valójában egy speciális ARIMA-modellről van szó, ahol bizonyos együtthatók zérusok, más együtthatókra pedig a felírásból eredő feltételek teljesülnek. Általánosítást az jelent, hogy lehetséges a szezonális differenciálás is (azaz $1\text{-}B^s$ operátor alkalmazása). Két típusuk az additív illetve a multiplikatív SARIMA-modell. Az additív típussal itt nem foglalkozunk. A multiplikatív *SARIMA*(*p*, *d*, *q*)(*P*, *D*, *Q*) modelltípus általános képlete a következő:

$$\phi_p(B)\Phi_P(B^S)\nabla^d\nabla_s^D y_t = \theta_q(B)\Theta_Q(B^S)a_t, \qquad (11)$$

ahol

$$\phi_p(B)\Phi_P(B^S) = (1+\phi_1 B + \phi_2 B^2 + ... + \phi_p B^p)(1+\Phi_1 B^S + \Phi_2 B^{2S} + ... + \Phi_P B^{PS})$$

az ún. autoregresszív operátorpolinom, ill.

$$\theta_q(B)\Theta_Q(B^S) = (1+\theta_1 B + \theta_2 B^2 + ... + \theta_q B^q)(1+\Theta_1 B^S + \Theta_2 B^{2S} + ... + \Theta_Q B^{QS})$$

az ún. mozgóátlag operátorpolinom,

és

$$\nabla^d \nabla_s^D = (1-B)^d (1-B^s)^D$$

a differenciaképzés operátora.

A TRAMO/SEATS-módszer dekompozíciós eljárása a spektrum felbontására épül, ezért a spektrál analízis (vagy frekvenciatartománybeli elemzés) lényegét röviden összefoglaljuk:

Régóta alkalmazott függvényvizsgálati technika a Fourier-analízis, amelynél egy periodikus függvényt különböző frekvenciájú és amplitúdójú szinuszok és koszinuszok végtelen összegére bontunk fel, ahol a frekvenciák előre adottak, és a függvényt a megfelelő frekvenciához tartozó amplitúdókkal jellemezzük. Ezt a felírást hívják Fourier-sornak. A kérdéses amplitúdók az ún. Fourier-együtthatók. Az eljárás általánosítható nem periodikus függvények és sorozatok vizsgálatára is.

A Fourier-analízis mintájára vezették be az idősorok vizsgálatára a spektrál analízist. A spektrál analízis azt a megközelítést alkalmazza, hogy egy stacionárius idősor autokovariancia függvényének értékeit írja fel egy ún. spektrális sűrűségfüggvény Fourier-együtthatóiként. A spektrális sűrűségfüggvényt röviden spektrumnak szokás nevezni, és a [-π, π] intervallumon értelmezett szimmetrikus, nemnegatív függvényről van szó. Spektrális sűrűségfüggvény nem minden stacionárius idősorhoz létezik, de a létezéshez elégséges, hogy az autokovarianciák abszolútértékeinek összege véges legyen. Ez a legtöbb gyakorlatban előforduló idősornál feltehető.[3]

[3] Megj.: például az ún. hosszú memóriájú (vagy hosszú távon összefüggő) folyamatokra a feltétel nem igaz. Az ilyen folyamatoknál az autokovariancia függvény lassan cseng le, ezért az autokovarianciák abszolút összege végtelen. Ennek ellenére ezeknek a folyamatoknak is van spektrális sűrűségfüggvényük.

Ilyenkor a spektrum egyszerűen felírható a következőképpen: $f(\lambda)=\frac{1}{2\pi}\sum_{h=-\infty}^{\infty}e^{-ih\lambda}\gamma(h)$,

ahol γ az autokovariancia függvény.

Ennek az összefüggésnek a segítségével bizonyos folyamatok spektruma egyszerűen számolható, például egy σ szórású fehérzaj spektruma $\frac{\sigma^2}{2\pi}$, azaz konstans.

A spektrumban megjelenő csúcsok az adott frekvenciához tartozó periodikus viselkedést jelzik az idősorban. A spektrum vizsgálata segítséget jelent akkor is, amikor úgynevezett időinvariáns lineáris szűrők hatását vizsgáljuk. Például az ARMA-folyamat is értelmezhető egy speciális idő-invariáns lineáris szűrő outputjaként, amit a fehérzajra, mint inputra alkalmaznak. Ilyen esetben az input folyamat spektrumának és a szűrő ún. négyzetes erősítésének (squared gain)[4] szorzataként áll elő az output spektruma. Az X család és a TRAMO/SEATS szezonális kiigazításának eredménye is értelmezhető az eredeti idősorra vett valamilyen időinvariáns lineáris szűrő outputjaként, ezért a spektrális megközelítés hasznosnak bizonyult a szezonális kiigazítás különféle módszereinek összehasonlításánál is.

A spektrum előállításánál az elméleti autokovariancia függvényből indulunk ki, tehát elméleti spektrumról van szó, amelyet az idősor tényleges realizációja alapján becsülni kell. A becslésre alapvetően két mód kínálkozik. Az egyik az ún. periodogrammal (illetve simított periodogrammal) való becslés, amelynél lényegében az autokovariancia függvény becslésének segítségével történik a spektrum becslése. A másik lehetőség, hogy az idősorra valamilyen modellt (például ARMA) becslünk, és a modell elméleti spektrumának ismeretében kiszámítjuk az idősor spektrumát. Ez utóbbi megközelítést alkalmazza a TRAMO/SEATS módszer a spektrum becslésére.

A spektrumot a fentiekben stacionárius idősorokra értelmeztük, de lehetőség van bizonyos nemstacionárius idősorokra a spektrumhoz hasonló ún. pszeudo-spektrumot értelmezni. A (szezonális) ARIMA-modell pszeudo-

[4] Más szóhasználattal: teljesítmény átviteli függvény (power transfer function).

spektrumát úgy kapjuk meg, hogy egy olyan ARMA-modell spektrumát írjuk fel, ahol a differenciaképzést beleértjük az autoregresszív részbe, azaz egységgyököket is megengedünk az autoregresszív polinomban. Ilyen módon a spektrumnak bizonyos helyeken pólusai lesznek, azaz a reguláris (azaz sima) differenciálásnak megfelelő zérus, illetve a szezonális differenciálásnak megfelelő szezonális frekvenciáknál a spektrum értéke végtelen. Képlettel: ha x_t egy szezonális ARIMA-folyamat, és $\phi(B)x_t = \theta(B)a_t$, ahol ϕ az autoregresszív polinom, amely a differenciálást is tartalmazza, θ a mozgóátlag polinom, a_t pedig □ szórású fehérzaj, akkor az x_t folyamat pszeudospektruma:

$$f_x(\lambda) = \frac{\sigma^2}{2\pi} \frac{\left|\theta(e^{-i\lambda})\right|^2}{\left|\phi(e^{-i\lambda})\right|^2}.$$

2.6. Az X11-ARIMA

Mint azt a korábbi részben említettük, az idősorok összetevői több módon is kapcsolódhatnak egymáshoz: leggyakrabban összegszerűen, illetve szorzatszerűen. Az előbbi additív, az utóbbi multiplikatív modellhez vezet. Mivel a gazdasági idősorok többsége multiplikatív összekapcsolódást feltételez, ezért a program ismertetésénél használt példák is erre a modelltípusra fognak vonatkozni. Multiplikatív modell esetén az idősor felbontása:

$$Y_t = T_t \times S_t \times I_t \times D_t \times E_t$$

ahol T_t a trend, S_t a szezonális komponens, I_t a véletlen hatás, D_t a munkanaphatás, E_t a húsvéthatás.

Az X11-ARIMA program fő lépései a következők:

1. ELŐZETES KORREKCIÓK

a) Az idősor előzetes korrekciója

b) A munkanaphatás (D_t) becslése és eltávolítása

c) A húsvéthatás (E_t) becslése és eltávolítása

$$Y_t^1 = Y_t / D_t \times E_t$$

d) ARIMA-modell segítségével az idősor előrejelzése és kiegészítése

2. A TREND, A SZEZONÁLIS TÉNYEZŐK ÉS A VÉLETLEN HATÁS ELŐZETES BECSLÉSE

$$\begin{aligned}
&\mathrm{M}[Y_t^1] = T_t^{\text{előz}}\\
&Y_t^2 = Y_t^1 / T_t^{\text{előz}} = S_t \times I_t\\
&\mathrm{M}[Y_t^2] = S_t^{\text{előz}}\\
&Y_t^3 = Y_t^1 / S_t^{\text{előz}} = T_t \times I_t\\
&\mathrm{M}[Y_t^3] = T_t^1
\end{aligned}$$

3. A KILÓGÓ ÉRTÉKEK BECSLÉSE ÉS HELYETTESÍTÉSE UTÁN A 2. SZAKASZ LÉPÉSEINEK MEGISMÉTLÉSE (T_t^1-et használjuk a trend előzetes becsléseként az első lépésben)

4. A KOMPONENSEK VÉGSŐ BECSLÉSE

A szezonálisan kiigazított idősor becslése:

$$Y_t^{\text{szez.kiigazított}} = Y_t^1 / S_t^{\text{vég}} \times D_t^{\text{vég}} \times H_t^{\text{vég}} = T_t^{\text{vég}} \times I_t^{\text{vég}}.$$

Most nézzük az egyes lépéseket részletesebben.

<u>1. a) Az idősor előzetes korrekciója</u>

Mielőtt a program megkezdi a kiigazítást, a felhasználónak lehetősége van arra, hogy az idősorra vonatkozó előzetes információit rögzítse, és a szezonális kiigazítás megkezdése előtt megváltoztassa mindazokat az adatokat, amelyek az idősor általános mintájába nem illeszkednek. Megadhatunk korrekciós tényezőket az egyes hónapokra, ha például figyelembe akarjuk venni bizonyos „mozgó" ünnepek – ilyen például a húsvét vagy a pünkösd – hatását az általunk megfigyelt folyamatra. Az előzetes igazításhoz tartozik az is, amikor rögzítjük az éves frissítés során nyert napi súlyokat, illetve a mozgóátlag, vagy az idősor meghosszabbítására használt ARIMA-modell típusát. Év közben ezeket a tényezőket nem változtatjuk, a szükségtelen revíziók elkerülése érdekében.

A program az előzetes kiigazítás tényezőit rögzíti és alkalmazza az idősorra, az eredményt a felhasználó ellenőrizheti. A további kiigazításokat az így nyert értékekkel hajtja végre.

1. b) A munkanaphatás (D_t) becslése és eltávolítása

A munkanaptényezőbe csak a nem éves ciklusban jelentkező hatások tartoznak bele, mivel az évenként rendszeresen ismétlődő hatásokat (pl. a december végi több napos leállások) a szezonális tényező már kimutatja.

A napi súlyokat regresszió segítségével a következőképpen becsüljük (multiplikatív modell):

$$I_t D_t - 1 = \frac{X_{1t}B_1 + X_{2t}B_2 + \cdots + X_{7t}B_7 + \varepsilon_t}{N_t},$$

ahol X_{jt} az adott nap előfordulásának gyakorisága az adott hónapban,

B_j a 7 napi súly, $\sum_1^7 B_j = 7$,

N_t pedig 31, 30 vagy 28,25 attól függően, hogy a hónap 31, 30 napos vagy február.

A súlyokat a legkisebb négyzetek módszerével becsüljük, majd előállítjuk a kombinált napi súlyokat úgy, hogy minden napra összeadjuk a regressziós becslés által nyert értéket és az előzetes igazítás során megadott napi súly értékét (csak multiplikatív modell esetén van lehetőségünk előzetes súlyok megadására). Ha az előzetes igazítás során nem adtunk meg napi súlyokat, akkor 1-hez adjuk hozzá a regressziós együtthatókat. A következő lépésben t-próbával teszteljük, hogy a kombinált súlyok szignifikánsan különböznek-e multiplikatív modell esetén az előzetes súlyoktól vagy 1-től, additív modellnél pedig 0-tól. A munkanaphatást pedig F-próbával teszteljük.

A következő lépésben kiszámítjuk a munkanaptényezőt az egyes hónapokra:

$$D_t = \frac{X_{1t}(b_1+1) + X_{2t}(b_2+1) + \cdots + X_{7t}(b_7+1)}{N_t}$$

(a jelölések megegyeznek az előző képletével, b_j pedig a napi súlyoknak a legkisebb négyzetek módszerével kapott becslése).

A modell frissítésénél a munkanapregresszió segítségével meghatározzuk a napi súlyokat, és teszteljük, hogy szignifikáns-e a hatásuk. Ha igen, a PDW parancs segítségével rögzítjük őket a parancsfájlban, és év közben változatlan napi súlyokkal futtatjuk a programot.

Az eredeti idősort az így kapott munkanaptényezőkkel osztva kapjuk a munkanaphatástól tisztított idősort.

1. c) A húsvéthatás (E_t) becslése és eltávolítása

A felhasználó az idősorra vonatkozó előzetes információi alapján eldöntheti, hogy tulajdonít-e jelentőséget az elhúzódó hatásnak. Alapbeállításnál a program F-teszttel ellenőrzi, hogy a húsvét hatása szignifikáns-e – ha legalább 10%-os szinten szignifikáns, akkor alkalmazza a húsvétkorrekciót. Az így kapott húsvétfaktorokat rögzítjük, és az év során húsvétregresszió nélkül futtatjuk a programot.

Amikor a húsvét áprilisban van, a húsvét hatását a szezonális tényező tartalmazza. Amikor a húsvét márciusra, vagy április elejére esik, további korrekcióra van szükségünk (az április elejére eső húsvét esetén az elhúzódó hatás miatt, ami befolyásolja a márciusi adatokat is), hiszen ebben az esetben a szezonális tényező nem fogja tartalmazni a húsvét hatását.

Ezt regresszió segítségével teszi meg a program. Ennek során becsli a húsvéttényezőket azokra a hónapokra, amikor a húsvét márciusra esik, illetve külön-külön – attól függően, hogy melyik napra esik – azokra a hónapokra, amikor a húsvétot április első napjaiban ünnepeljük.

1. d) Az idősor előrejelzése és kiegészítése ARIMA-modell segítségével

Az ARIMA-modellt az aszimmetrikus mozgóátlagolás elkerülése érdekében, az idősor előre illetve hátra történő meghosszabbítására használjuk. Az ARIMA-modellezés (integrált autoregresszív mozgóátlagolás) lényege, hogy az idősorok leírására kidolgozott autoregressziós (amelyek azt becslik, hogy a megfigyelés mostani, z_t értéke hogyan függ az előző időszakok $z_{t-1}, z_{t-2}, z_{t-3}, \ldots, z_1$ értékeitől) és mozgóátlagoláson alapuló eljárásokat (amelyek pedig azt mutatják, hogyan függ a megfigyelés mostani értéke az előző időszakok véletlen tényezőitől) egy közös modellbe építjük be.

Egy (p,d,q) paraméterekkel jellemezhető ARIMA-modell tartalmaz egy q-ad rendű mozgóátlagolású és egy p-ed rendű autoregresszív modellt, amelyeket az idősor eredeti elemeiből képzett d-ed fokú differenciákra írunk fel. Szezonális modellre a paraméterek kibővülnek (p,d,q) $(P,Q,D)_{12}$-re, ahol a nagybetűs paraméterek a szezonális tényezőt jellemzik, a 12 az indexben pedig azt mutatja, hogy havi idősorral van dolgunk. Az általános képlet multiplikatív modellre:

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D z_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_{t,}$$

ahol B a késleltetési operátor[5] ($Bz_t = z_{t-1}$); $\nabla z_t = (1-B)z_t$;

$\phi_p(B), \Phi_P(B^s), \theta_q(B), \Theta_Q(B^s)$ a nemszezonális és szezonális autoregresszív és mozgóátlag polinomok és ε_t fehér zaj folyamat.

A log $(0,1,1)(0,1,1)_{12}$ modellre (Airline-modell) – a legegyszerűbb, de nagyon sok idősorra jól illeszthető modell – ugyanezt felírva:

$$\nabla\nabla_{12} z_t = (1-\theta B)(1-\Theta B^{12})\varepsilon_{t,}$$

vagy:

$$z_t - z_{t-1} - z_{t-12} + z_{t-13} = \varepsilon_t - \theta\varepsilon_{t-1} - \Theta\varepsilon_{t-12} + \theta\Theta\varepsilon_{t-13}.$$

Automatikus futtatás esetén a program az öt előre beépített ARIMA-modellből választja ki az idősorhoz illeszkedőt, a modellek tesztelését a legegyszerűbbel $(0,1,1)(0,1,1)_{12}$ kezdi, és fokozatosan halad a bonyolultabbak felé. A vizsgált modellt a program nem fogadja el, ha a modell alapján az utolsó 3 évre számított és a tényleges adatok eltérése

---

[5] Angol elnevezése backshift operátor. A szakirodalomban előfordul még a L (lag operátor) jelölés is.

meghaladja a 15%-ot, vagy a reziduumok véletlenszerűségét ellenőrző χ^2-próba eredménye nem éri el az 5%-ot. Amennyiben a program talált egy megfelelő modellt, a többit már nem ellenőrzi. Évente egyszer, a modell felülvizsgálatánál automatikus modellillesztést végzünk, ezt a modellt a parancsfájlban rögzítjük, és a többi hónapban ezzel futtatjuk a programot.

2. a) A trend leválasztása

A trend előzetes becslésére a program egy alapbeállításban 12 (vagy a felhasználó által kérhető 24) tagú középre igazított mozgóátlagot illeszt az eredeti (vagy előzetesen igazított) idősorra:

$$T_t^{\text{elöz}} = \frac{1}{24} Y_{t-6} + \frac{1}{12}(Y_{t-5} + Y_{t-4} + \cdots + Y_{t+4} + Y_{t+5}) + \frac{1}{24} Y_{t+6}$$

Ha az idősorból leválasztjuk a trendet, a szezonalitás és a véletlen tényező együttes hatására kapunk előzetes becslést (SI-idősor). A leválasztás additív modell esetén $SI_t = Y_t - T_t$, multiplikatív modellre: $SI_t = Y_t / T_t$.

A trend végleges becslésére a szezonálisan kiigazított idősorra a program újból mozgóátlagolású trendet illeszt, de a súlyrendszer itt más: 9, 13 vagy 23 tagú Henderson-féle átlagolást használ. A tagszám attól függ, hogy hogyan viszonyul egymáshoz a trendkomponens és a véletlen tényező változásának átlagos értéke – ezt a program I/C mutatója jelzi. Erősebb irreguláris hatás esetén a trend nagyobb intervallumra kiterjedő mozgóátlagolással becsülhető. Minél nagyobb az I/C hányados értéke, annál nagyobb tagszámú mozgóátlagot illeszt a program.

$$\left.\begin{array}{rr} I/C & 9 \\ 1 \le I/C < 3{,}5 & 13 \\ 3{,}5 \le I/C & 23 \end{array}\right\} \text{tagú mozgóátlag}$$

A Henderson-féle mozgóátlagolás szimmetrikus, ahol a súlyok összege egyenlő 1-gyel. 2k+1 tagszámú Henderson-súlyokat használva a trend képlete a következő lesz:

$$T_t = \sum_{j=-k}^{k} h_j SI_{t+j}$$

A szimmetrikus mozgóátlagoláshoz a program az idősor mindkét végén az ARIMA-extrapolációval kiszámított értékekkel számol.

2. b) Szezonális kiigazítás

A szezonális tényezők meghatározásához is mozgóátlagokat számít a program, de itt az azonos hónapra vonatkozó SI-értékeket átlagoljuk, hogy a szezonális és véletlen hatást szétválasszuk. Erre a célra a program többszörös mozgóátlagolást (3x3, 3x5 vagy 3x9-es) alkalmaz, ami azt jelenti, hogy az egyszerű mozgóátlagokból képez újabb mozgóátlagokat. Például a 3x3 mozgóátlag esetében az azonos hónapokra egy háromtagú, majd még egyszer egy háromtagú mozgóátlagot illesztünk:

$$S_t = \frac{1}{3}\left[\frac{1}{3}(SI_{t-24} + SI_{t-12} + SI_t) + \frac{1}{3}(SI_{t-12} + SI_t + SI_{t+12}) + \frac{1}{3}(SI_t + SI_{t+12} + SI_{t+24})\right] =$$

$$\frac{1}{9}SI_{t-24} + \frac{2}{9}SI_{t-12} + \frac{3}{9}SI_t + \frac{2}{9}SI_{t+12} + \frac{1}{9}SI_{t+24}$$

Az előzetes igazításhoz a program 3x3-as mozgóátlagot illeszt. A végleges szezonális tényezők megállapításához a program vizsgálja a szezonalitás stabilitását. Ehhez minden hónapra kiszámítja a szezonális és irreguláris tényező évenkénti átlagos változását, valamint a két tényező arányát – ez az I/S hányados. Automatikus futtatás esetén a mozgóátlagok elemszáma az I/S hányados értékétől függ.

$$\left.\begin{array}{cc} I/S < 2{,}5 & 3\times 3 \\ 3{,}5 \le I/S < 5{,}5 & 3\times 5 \\ 6{,}5 \le I/S & 3\times 9 \end{array}\right\} \text{tagú mozgóátlag}$$

Ha az I/S értéke 2,5 és 3,5, valamint 5,5 és 6,5 közé esik, az X 11 ARIMA újraszámolja azt a legutolsó teljes év elhagyásával, és az így kapott érték alapján dönt. Ha a kapott érték továbbra is bizonytalan, akkor a 3x5-ös mozgóátlagolást választja.

Évente, a modell frissítésénél automatikus futtatással kiválasztjuk a megfelelő mozgóátlagot; ezt a parancsfájlban rögzítjük, és a következő frissítésig ezt használjuk. Mivel az I/S arány hónaponként eltérő lehet, lehetőség van havonként különböző elemszámokat előírni. Ez akkor hasznos, ha csak néhány hónap szezonális tényezője tűnik stabilnak, és el akarjuk kerülni, hogy a többi hónaphoz tartozó nagyobb vagy kisebb elemszámú mozgóátlagok rontsák az adott hónapra vonatkozó becsléseket.

Az SI idősorból leválasztva a szezonális tényezőket kapjuk a véletlen tényező értékét. Additív modellre $I_t = SI_t - S_t$, multiplikatív modell esetén $I_t = SI_t / S_t$.

3) Outlierek kezelése

Az X11-ARIMA program képes felismerni és az idősor mintájához jobban illeszkedő értékekkel helyettesíteni az additív outliereket. A szinteltolódás és az átmeneti szintváltás kezelésére nincs beépített opció. Egy lehetséges megoldást nyújt az idősorban bekövetkezett törés előtti rész elhagyása, ha a maradék rész elég hosszú, legalább 5 év. Ha ez nem teljesül, az előzetes igazítások során a felhasználó transzformálhatja az idősort úgy, hogy az megtartsa a szezonalitás jellegét, de a törés okozta szintváltást kiszűrje belőle.

Az additív outlierek felismeréséhez és értékeik korrigálásához a program a véletlen tényező idősorának minden 5 év hosszúságú szakaszára kiszámítja a szórást. A szakaszok középső évének minden értékéhez hozzárendel egy súlyt a következő módszerrel:

- Ha egy bizonyos hónap értékének az átlagtól vett eltérése nagyobb, mint az 5 éves szakasz szórásának 2,5-szerese, akkor a súly 0.
- Ha az érték átlagtól vett eltérése kisebb, mint az 5 éves szakasz szórásának 1,5-szerese, akkor a súly 100 (ezek az adatok jól illeszkednek).
- Az 1,5 és 2,5 közé eső értékekre a súly lineárisan csökken 100-tól 0-ig.

A 100-nál kisebb súlyú adatoknál az SI értéket az adatpont súlyával beszorzott SI értékének és az azt megelőző, illetve követő két legközelebbi, 100 súlyú adatpont értékének átlagával helyettesíti. Az első két évre a harmadik év szórásával számol, az utolsó kettőnél az azt megelőző utolsóval. Az outlierek korrekciójához az adat súlyozott SI értékének és a 4 legközelebbi 100 súlyú értéknek az átlagát használjuk

Ha valamely hónaphoz vagy évhez túl sok outlier tartozik, akkor az arra az időszakra vonatkozó szezonálisan kiigazított adatok nem lesznek megbízhatóak. Gyakori eset, hogy a márciusi és áprilisi adatok között szerepel sok outlier, ami a húsvéttényező nem megfelelő kezelésre utal. A gyakorlatban, amennyiben nyilvánvalóan kiugró adattal találkozunk, célszerű azt már az előzetes igazítás során eltávolítani. Ha az outlier oka ismert és azt egy valós esemény idézte elő, akkor az ideiglenes igazítás alkalmazandó, mivel így az outlier meg fog jelenni a szezonálisan kiigazított idősorban, de a szomszédos értékekre nem lesz hatással. Amennyiben az idősor elemzésekor nem találunk nyilvánvaló outliert, célszerű a programra bízni az outlierek automatikus kezelését.

A kiigazítás minőségének ellenőrzése

A szezonális kiigazítás eredményének értékelését a program számos diagnosztikai táblával segíti. Futása során az X11 ARIMA automatikusan elkészíti a diagnosztikai táblákat; amelyek az elemzés során felmerült problémák megoldását is segítik.

Az E táblákban az eredeti és a szezonálisan kiigazított idősor egy hónap alatti változásai követhetők nyomon. Az E5 táblában (eredeti adatok) az adott hónaphoz tartozó értékek előjele az évek során nem változik. Ha ebben a mintában változást észlelünk, akkor a szezonális tényező megtörése valószínűsíthető. Az outlierek a táblában kiugróan magas vagy alacsony értékként jelennek meg, amelyet a következő hónapban ellentétes irányú kiugró adat követ.

Ha az E6 tábla vizsgálata során (kiigazított adatok) bizonyos hónapokban ismétlődő változásokra bukkanunk, akkor feltehetően az idősorban még jelen van maradék szezonalitás, vagy pedig a program „túligazította” az idősort.

Az F táblák közül az F3 táblára érdemes részletesebben kitérni, amely összegző adatokat tartalmaz a kiigazítás eredményéről. A tábla segítségével gyors képet kaphatunk a kiigazítás sikerességéről. Az M1-M11-gyel jelölt összegző statisztikák eredménye minden esetben 0 és 3 közötti szám. Az M1-M6 statisztikák a véletlen tényezőhöz, az M7-M11 statisztikák a szezonális tényezőhöz kapcsolódnak. A kiigazítást sikeresnek tekinthetjük, ha valamennyi M-statisztika értéke 0 és 1 közé esik. Általában, minél kisebb értéket kapunk valamely M-statisztikára, annál jobbnak tekinthető a szezonális kiigazítás. Az M-statisztikák a következők:

Ml: az eredeti idősor 3 havi változásából az irreguláris tényező részaránya;

M2: a trend kiszűrésével nyert idősor szórásából az irreguláris komponens részaránya;

M3: a véletlen tényező átlagos egyhavi változása összehasonlítva a trend átlagos egyhavi változásával;

M4: az irreguláris tényező autokorrelációjának mértékét fejezi ki, vagyis azt, hogy felismerhető-e valamilyen minta az irreguláris tényezőben;

M5: hány hónap alatt haladja meg a trend változása az irreguláris komponens változását;

M6: a véletlen tényező egyévi változása a szezonális tényező egyévi változásához viszonyítva;

M7: a változó és állandó szezonalitás aránya;

M8: a szezonális tényező éves átlagos ingadozásának nagysága;

M9: a szezonális tényező átlagos lineáris jellegű változása;

M10: az M8 mutató, az idősor utolsó néhány évének adatára;

Ml l: az M9 mutató, az idősor utolsó néhány évének adatára;

Az F3 tábla után a program egy globális értékelési kritériummal jelzi a kiigazítás minőségét. Ez az M1-M11 mutatók súlyozott átlaga. A dekompozíció eredménye akkor tekinthető elfogadhatónak, ha értéke 1-nél kisebb.

2.7. X12-ARIMA[6]

Az X12-ARIMA módszer dekompozíciós része lényegében megegyezik az X11-ARIMA azonos lépéseivel. Amiben a két program eltér, az a munkanaphatás és az outlierek kezelése. Az X11-ARIMA esetében először végrehajtjuk a szezonális dekompozíciót, azaz leválasztjuk a trendet és a szezonális komponenst, majd megvizsgáljuk, hogy a maradéktényezőben maradtak-e outlierek, és, hogy a munkanaphatás és a húsvéthatás kimutatható-e szignifikánsan a véletlen hatásban. Ha igen, akkor az előzetes korrekciókat ez alapján módosítjuk.

---

[6] Sugár (1999a) alapján.

Az X12-nél ugyanakkor az eredeti idősorból kiindulva állapítjuk meg, hogy van-e szignifikáns munkanap-, húsvét- vagy outlierhatás. Ha ezen hatások bármelyike szignifikáns, akkor először igazítjuk az idősort, majd a korrigált idősorra hajtjuk végre a dekompozíciót.

Az X11-ARIMA-hoz hasonlóan – a mozgóátlag-technika alkalmazása miatt – az X12-ARIMA esetén is szükség van arra, hogy az idősort meghosszabbítsuk előre és hátra egyaránt. Az X12-nél a meghosszabbításhoz használt ARIMA-modell paramétereinek becslésével egyidőben történik a naptári hatások, outlierek regressziós változóinak meghatározása is.

Így a dekompozíció előtt az X12 egy regARIMA-modellt illeszt az idősorra: $Y_t = \sum \delta_i X_{it} + z_t$, ahol z_t egy szezonális ARIMA-folyamat: (p,d,q)(P,D,Q)$_S$. Az egyenletet kifejezve z_t-re és ezt behelyettesítve az ARIMA modell egyenletébe, a következő formát kapjuk:

$$\phi_p(B)\Phi_p(B^s)(1-B)^d(1-B^s)^D(Y_t - \sum \delta_i X_{it}) = \theta_q(B)\Theta_Q(B^s)\varepsilon_t,$$

ahol az előző fejezetben használt jelöléseknek megfelelően B a késleltetési operátor, $\phi_p(B)$, $\Phi_p(B^s)$, $\theta_q(B)$, $\Theta_Q(B^s)$ a nemszezonális és szezonális autoregresszív és mozgó átlag polinomok, ε_t fehérzaj.

Kiválasztjuk a megfelelő regressziós változókat, megbecsüljük az együtthatókat, képezzük az $Y_t - \sum \delta_i X_{it}$ változót. Ennek képezzük a megfelelő differenciáit, és becsüljük a szezonális ARIMA-modellt. Ez utóbbi, az X11-ARIMA módszerhez hasonlóan arra szolgál, hogy előre-hátra meghosszabbítsuk az idősort.

Az X12 regARIMA részében különböző regressziós változók alkalmazhatók, például trendkonstans, fix szezonális változó, munkanaphatás, a hónap hosszának kezelése, a szökőév kezelése, a húsvéthatás, outlierek.

Munkanaphatás

Az X11-ARIMA-tól eltérően a munkanapokat nem hét, hanem hat változóval kezeli. A változók értéke: az adott nap száma – vasárnapok száma. Lehetőség van olyan opció választására is, hogy a munkanapokat egy regressziós változóval kezelje a program. Ebben az esetben minden hónaphoz hozzárendeli, azaz számszerűsíti, hogy a munkanapok száma mennyivel több, mint a hétvégi napok száma: a [hétköznapok száma - $\frac{5}{2}$ (szombatok és vasárnapok száma)] értéket rendeli hozzá.

A húsvét hatása

A húsvéthatást is egyszerűbben kezeli, mint az X11-ARIMA. A regressziós változó értéke attól függ, hogy az ünnep előtt hány napig lehet érezni a hatását. Márciusban és áprilisban 1/v, ahol v a húsvét előtt adott hónapra eső napok számát jelenti, a többi hónapban 0 az értéke.

Szökőévhatás

Az X12-ARIMA már a szökőév hatását is figyelembe veszi. A szökőév kezelésére bevezetett regressziós változó értéke 0,75, ha szökőév februárja van, és -0,25 a többi februárban, illetve 0 a többi hónapban.

Outlierek típusa

Az X12-ARIMA is az outlierek (kiugró értékek) 4 típusát különbözteti meg: additív outlier, szinteltolódás, csillapodó jellegű törés, átmeneti szintváltás.

Az outlierek becslése történhet automatikus kereséssel,[7] de a felhasználónak is lehetősége van az időpontok megadására. Az automatikus keresés két részből áll: a forward és a backward változószelekciós részből.

A program először a véletlen tényezőre számol standard hibát. Megbecsli a regressziós együtthatókat és az ARIMA-modell paramétereit, majd a véletlen tényezők mediántól vett abszolút átlagos eltérését számítja ki. Ennek 1,49-szorosát tekinti a véletlen tényező szórásának. Ennek segítségével *t* értéket számol, és ezt hasonlítja a *t*-statisztika értékéhez. A forward eljárás keretében egy adott *t* kritikus értékhez (alapbeállításban 3,8) viszonyítja a modellbe be nem vont változók bevonása melletti *t* értéket. Kiválasztja a legnagyobb abszolút értékűt, és ha ez meghaladja a kritikus értéket, akkor bevonja a változót a modellbe, majd újra becsli a modellt. Az algoritmust addig folytatja, amíg talál a kritikus *t* értéknél nagyobb abszolút értéket.

Az előző forward eljárás során kapott változókból indul ki a backward algoritmus. A program itt lépésenként kihagyja azt a változót (az abszolút

[7] Kivéve az átmeneti szintváltás esetét, arra nincs automatikus keresési eljárás.

értékben legalacsonyabb *t* értékűt), amely nem felel meg az adott kritériumnak.

Végül az utolsó lépés az ARIMA-modell típusának meghatározása. A programban 5 alapmodell található, de kiindulópontként bármilyen tetszőleges modell megadható. Az alapmodellek szezonális modell része mindig (0,1,1) alakú, a nem szezonális rész modellje pedig (0,1,1), (0,1,2), (2,1,0), (0,2,2), (2,1,2) alakú lehet. A program automatikusan választ a modellek közül az Akaike-féle információs kritérium (AIC) alapján, de a diagnosztikák (autokorrelációs és parciális autokorrelációs függvények, erre vonatkozó Box-Pierce- és Ljung-Box-próbák) segítségével a felhasználó is dönthet az alkalmazandó modellről.

A modell illesztése után a program két kritérium teljesülését is figyeli: egyrészt az utolsó három évre az átlagos abszolút hibának az átlag százalékában kifejezett értéke ne haladja meg a 15%-ot, másrészt a Ljung-Box-statisztika értéke 5%-os szinten ne legyen szignifikáns. Ezek a kritériumok azonban nem befolyásolják a modellek közti választást, csak utólagos ellenőrzésre szolgálnak.

Amennyiben ezek a feltételek egyik modellre sem teljesülnek, a program nem illeszt ARIMA-becslést, hanem az X11-ARIMA hagyományos aszimmetrikus filterét használja.

A kiigazítás minőségének ellenőrzése

M-statisztikák, Q mutató

Az X11-ARIMA diagnosztikáit ez a program is tartalmazza. Mint arról korábban szó volt, ezek az eredmények stabilitását és a véletlen tényező szerepét értékelik.

Csúszó tartományok és változások követése elvén alapuló tesztek

Az X11-ARIMA tesztjeit kiegészültek ezekkel az új diagnosztikákkal is. A csúszó tartományok (sliding spans) elve a mintavételi hiba becslésekor alkalmazható bootstrap elvnek, a változások követése (revision history) elv pedig a jackknife elvnek feleltethető meg.

A csúszó tartományok tesztnél is az eredmények stabilitását ellenőrizzük, de úgy, hogy a szezonális kiigazítást évente csúsztatva végezzük el és az azonos időszakokra vonatkozó értékeket (például a szezonális tényező értékét) hasonlítjuk egymáshoz. Multiplikatív modell esetén a következőképpen járunk el: a megfigyelési tartományt mindig egy évvel csúsztatjuk. A program számolja minden időszakra a szezonális tényező minimális és maximális értékét, ha ezek között 0,03-nál, azaz 3 százalékpontnál nagyobb eltérés van, akkor az adott időszakra a szezonális tényezőt instabilnak tekinti. Ha ez az értékek több mint 15%-ra fennáll, az eredmények nem tekinthetők stabilnak.

A változások követésének elve az új adatok megjelenésének hatását vizsgálja és számszerűsíti. A változást általában láncindexszerűen, az előző időszakhoz viszonyítva számszerűsíti. Az új adat megjelenése az idősorban általában azzal jár, hogy a trend és a kiigazított értékek visszamenőleg is megváltoznak. Annak érdekében, hogy a stabilitást valamennyire biztosítani lehessen, az X12-ARIMA esetében is a modell rögzítését ajánlják. Ez konkrétan azt jelenti, hogy az előző év végéig vizsgálva az adatokat, számszerűsítik az outliereket, a munkanapok hatását, az ünnepeket, majd az idősor szezonális kiigazítása után megbecslik a következő év szezonális hatását. Az új adat bekerülésekor a regARIMA-modellt újrabecslik, de az előző év beállításait (ARIMA-modell, munkanaphatás) megtartják. Az így kapott tényezőket kiszűrik az idősorból, de a szezonális kiigazítást már a becsült szezontényezőkkel végzik.

Szignifikáns szezonalitás az eredeti idősorban

Az eredeti idősor helyett a program a Henderson-féle mozgóátlagolású trendek leválasztása után hajtja végre a tesztet. A nullhipotézis szerint a trend leválasztása után kapott tényezők havi átlaga megegyezik egymással, azaz nincs az idősorban szezonalitás. Ha a véletlen tényező normális eloszlású fehér zaj, akkor a hipotézist (11,N-12) szabadságfokú F-próbával tesztelhetjük. Amennyiben a nullhipotézis elfogadható, nincs értelme a szezonális kisimításnak, sőt káros is lehet.

Ezzel a teszttel lehet azt is ellenőrizni, hogy maradt-e szezonalitás az idősorban a kisimítás után. Ilyenkor a tesztet a véletlen tényező végső becslésére kell elvégezni.

Spektrális elemzés

Szintén új diagnosztikai elem az X12-ARIMA programban a spektrális elemzés megjelenése, amely lehetővé teszi, hogy a ciklikus jellegű munkanaphatást és szezonalitást frekvenciatartományon történő elemzéssel is megvizsgáljuk. A program becsüli és kiírja a véletlen tényező és a szezonálisan kiigazított idősor spektrumát, valamint jelöli a naptári és munkanap hosszakhoz tartozó frekvenciák helyeit.

2.8. A modellalapú megközelítés

Az itt következő ismertetés részletesebben megtalálható Maravall (1999) cikkében.

A módszer lényege

A szezonális kiigazítás ún. modell alapú megközelítése két alapvető elemet használ: egyszerű, paraméteres idősor modellt (legtöbbször ARIMA-modellt), és *signal extraction* (kb.: jelkinyerés) technikát, ami matematikailag jól definiált komponensek optimális becslését jelenti. Az egyik ilyen fajta megközelítést – amelyre a TRAMO-SEATS eljárás is épül – ARIMA-modell-alapú (AMB) módszernek hívják. Az alapgondolata a következő:

Az $[x_t] = [x_1, \ldots, x_t]$ megfigyelt idősort egymásra merőleges – azaz korrelálatlan – nem megfigyelt komponensek összegére bontjuk fel:

$$x_t = \sum_i x_{it} \,,$$

ahol az x_{it} komponensek felírhatók ARIMA-folyamatként. A komponensek ortogonalitásának követelménye arra a meggondolásra épül, hogy az idősor trendjét és a sor szezonalitását egymástól alapvetően független tényezők befolyásolják. Mivel ARIMA-folyamatok összege is ARIMA-folyamat, ezért a megfigyelt idősor is leírható ARIMA-modell-lel. Az egyes komponensek, így a trend, a szezonális komponens és az irreguláris komponens modelljét a megfigyelt idősorra becsült modellből kiindulva határozzuk meg, ehhez a komponensekre különféle feltevéseket kell tenni. Az egyik ilyen, hogy az irreguláris komponens egyszerű fehérzaj. Lényeges az is, hogy az AMB-megközelítésnél a komponensekről feltesszük, hogy a trend és a szezonális komponens stabilitása maximális, így az irreguláris komponens varianciája maximális. Ebben például eltér egy másik modell alapú megközelítéstől, az ún. strukturális idősor (STS) módszertől, amelynek lényege, hogy közvetlenül a komponensekre határozza meg az ARIMA-modellt.

Miután a nem megfigyelhető komponensek modelljeit meghatároztuk, a komponenseket minimum átlagos négyzetes hiba (MMSE) becsléssel becsüljük. Ez a nem megfigyelt komponensek megfigyelt idősorra vonatkozó feltételes várható értéke, amit signal extraction technikával számolunk, konkrétan a Wiener-Kolmogorov-szűrővel (az STS-nél Kalman-szűrővel).

Említésre méltó, hogy számos ad hoc szezonális kiigazító eljárás értelmezhető speciális ARIMA-modellekre vonatkozó modellalapú MMSE-becslésként.

Előnyök

Csökken annak veszélye, hogy olyan idősort igazítunk ki szezonálisan, amely nem is szezonális, mivel ilyenkor az AMB megközelítésnél – nagy valószínűséggel – az idősorra nemszezonális modellt illesztünk. Klasszikus példa, hogy az X11-gyel, illetve az X12-vel – melyek az ad hoc módszerek közé tartoznak – igazítva szezonális komponenst nyerhetünk ki a nyilvánvalóan nem szezonális fehérzajból.[8]

Az ad hoc módszereknél a kiigazítás rögzített szűrővel történik, vagy rögzített szűrők szűk választéka áll rendelkezésre. Ez azt jelenti, hogy a kiigazított sor az eredeti sornak valamilyen előzetesen lerögzített együtthatós mozgóátlagolásával áll elő. A modell alapú megközelítésnél viszont az együtthatókat a sor tulajdonságaihoz igazítjuk, vagyis az ARIMA-modellhez. A rögzített szűrővel az a gond, hogy nem veszi figyelembe az idősor szezonalitásának stabilitását. A szezonális frekvenciáknál fix szélességben szűr, így a szűrőnek megfelelőnél instabilabb szezonalitású (a szezonális frekvenciáknál szélesebb spektrális csúcsot mutató) idősorokat aluligazítja, azaz a kiigazított sorban szezonalitás marad. Ugyanakkor a stabilabb szezonalitású (a szezonális frekvenciáknál keskenyebb spektrális csúcsot mutató) idősorokat túligazítja, azaz „több” szezonalitást nyer ki az idősorból, mint ami abban megtalálható. Mivel a modellalapú megközelítésnél a szűrő előállításánál figyelembe vesszük az idősor viselkedését, ezért az alul- és túligazítás veszélye csökken.

[8] Ennek elkerülésére az X11-nél és az X12-nél is kötelező megvizsgálni, hogy az igazítani kívánt idősor szezonális-e.

Problémák az AMB-megközelítéssel

Az AMB-megközelítésnél az elméleti komponensek és a becsült komponensek különböző sztochasztikus tulajdonságokkal rendelkeznek.

Az elméleti komponensek egyik tulajdonsága az ortogonalitás, vagyis az egymással való korrelálatlanság. Ez a tulajdonság azonban a becsült komponensekre (illetve azok 1 időszakkal eltolt értékeire) már nem teljesül.

Továbbá a becsült komponensek a folyamat jövőbeli értékeitől is függenek. Nem könnyű megérteni, hogy a szezonalitásban vagy a trendben miért vannak a jövőre vonatkozó információk. A legjobb szemlélet talán az, hogy a jövőbeli értékek nem „generálják” a komponenseket, hanem csupán az optimális közelítésként kapott számításból adódik az összefüggés.

Az irreguláris komponens a modell alapján fehérzaj. A becsült irreguláris komponens spektruma viszont a szezonális frekvenciáknál nulla értéket vesz fel, tehát nem lehet fehérzaj. Mivel a becsült trend spektruma is zérus a szezonális frekvenciáknál, ezért a becsült szezonálisan kiigazított sor spektruma is nullává válik a szezonális frekvenciáknál. Úgy is mondhatjuk, hogy az elméleti szezonálisan kiigazított komponenssel ellentétben a becsült komponens túligazított.

2.9. Automatikus modellidentifikáció és paraméterbecslés (TRAMO)

Az eljárás célja, hogy automatikusan meghatározhassuk az idősorra illeszkedő modellt és megbecsüljük a modell paramétereit. Ebbe beleértjük az additív és multiplikatív modelltípus közötti választást, az outlierek automatikus felismerését és korrekcióját, esetleges hiányzó értékek interpolációját, a munkanaphatás tesztelését és számszerűsítését, felhasználó által definiált változókkal való regressziót, illetve az idősorra illeszkedő ARIMA-modell identifikációját (azaz a differenciálás rendjének és az autoregresszív és mozgóátlag paraméterek számának kiválasztását), és az ARIMA-modell paramétereinek becslését.

A TRAMO-nak az alábbiaknál matematikailag részletesebb ismertetése Gómez és Maravall (2001) cikkében megtalálható.

A felsorolt funkciókat az alább felsorolt négy részben (1.-4.) részletezzük, majd az ötödikben ismertetjük az ezek kombinációjával megvalósuló bonyolult algoritmust. A hatodik részben a munkanap- és húsvéthatással, mint kiemelten fontos speciális regresszorokkal foglalkozunk.

1. Log-teszt
2. Egységgyök-teszt
3. Az ARMA modell rendjének kiválasztása
4. Automatikus outlier felismerés és korrekció
5. A kombinált algoritmus
6. A munkanap- és húsvéthatás

1. Log-teszt

Célja a multiplikatív és az additív modelltípus közötti automatikus választás, azaz annak eldöntése, hogy az idősor nem megfigyelhető komponensei (trend, szezonális komponens, irreguláris komponens) összegszerűen vagy szorzatszerűen kapcsolódnak. Nem szabad összekeverni az itt használt additív illetve multiplikatív modelltípus fogalmát az ARIMA-modellekre vonatkozó multiplikatív illetve additív szezonális ARIMA-modellek fogalmával: ebben a kiadványban szezonális ARIMA-modell alatt mindig multiplikatív szezonális ARIMA-modellt értünk.

A multiplikatív esetben az idősor értékeinek logaritmusát vesszük, és erre illesztünk additív modellt. Valójában tehát az additív és a log-additív modell között döntünk.

Ennél és a következő tesztekben is, gyakran illesztjük az idősorra az úgynevezett Airline-modellt, azaz az ARIMA(0,1,1)(0,1,1)$_s$ modellt. Ennek az az oka, hogy ez a modell gyakran előfordul a gyakorlatban és jól közelít más modelleket is.

Tehát a logaritmálatlan illetve a logaritmált sorra az említett Airline modellt illesztjük, és a kisebb BIC-cel rendelkező esetet választjuk. A BIC (Bayes-i információs kritérium, szokás még SIC-nek: Schwarz információs kritériumnak is hívni) az ARMA-modellek illeszkedésének egy mértéke, amely figyelembe veszi a modell paramétereinek számát is. Az alábbiakban ismertetendő automatikus modellidentifikálásnál is a TRAMO a BIC-et alkalmazza.

2. Egységgyök-teszt

Célja az ARIMA-modell d reguláris (azaz nem szezonális) és D szezonális differenciálási rendjének meghatározása. Mivel az 1-B reguláris differencia operátor és az $1\text{-}B^s$ szezonális differencia operátor gyökei mind az egységkörre esnek, ezért a differenciálás rendjének meghatározását szokás egységgyök tesztnek hívni. Ha a szükséges differenciálások számát meghatároztuk, akkor a differenciált sorra már csak ARMAmodellt kell illeszteni az ARIMA-modell AR és MA tagjainak meghatározásához. A differenciálás rendjére a legtöbbször a következő korlátokat tételezzük fel: $0 \leq d \leq 2$, $0 \leq D \leq 1$.

A módszer a következő:[9]

Speciális szezonális AR majd ARMA-modellt (várható értékkel) illesztünk a z sorra, a paraméterbecsléseket a Hannan-Rissanen (HR) módszerrel (lásd később) vagy egzakt maximum likelihood (ML) becsléssel[10] kapjuk, majd a kapott egységgyökök adják a d reguláris illetve a D szezonális differenciálás rendjét. A lépéseket a kapott egységgyököktől függően megismételjük. A használt AR-modell az $AR(2)(1)_s$, amit Tiao és Tsay (1983)-ban szereplő eredmény indokol, amely szerint, ha legalább akkora rendű AR-modellt illesztünk az idősorra, mint a differenciálás rendje, akkor a becsült AR-modellben megjelenő egységgyökök konzisztensen becslik az ismeretlen egységgyököket. Az $AR(2)(1)_s$ modell választás megfelel a differenciálás rendjére feltett $0 \leq d \leq 2$, $0 \leq D \leq 1$ korlátoknak. A következő lépésben használt ARMA-modell az $ARMA(1,1)(1,1)_s$, ami az airline modell általánosításának

[9] A klasszikus egységgyök tesztek (Dickey-Fuller, Phillips-Perron) rosszul viselkednek bizonyos típusú autokorrelációk esetén (Gómez és Maravall, 2001), illetve a szezonális egységgyököket sem kezelik.

[10] Az egzakt ML-becslésnél a likelihood függvényt úgy maximalizáljuk, hogy a kezdeti értékeket is véletlennek tekintjük, szemben a feltételes ML-becsléssel, ahol ezek rögzítettek. A nehézséget az adja, hogy így nemlineáris egyenleteket kell megoldani. Ezért a megoldás csak numerikusan történhet, például Kalman-szűrővel.

tekinthető. Használatát rugalmassága indokolja, ami miatt olyan egységgyököket is találhat, amit az AR-modell nem.

Az utoljára becsült modell reziduuma alapján döntjük el, hogy várható értékes modellt fogunk-e használni vagy várható érték nélkülit.

A most bemutatott eljárás ARIMA hibatényezős regressziós modellre alkalmazva is érvényes eredményt ad az ARIMA-modell egységgyökeire és így a később ismertetendő regressziós becslést úgy alkalmazhatjuk, hogy a differenciálás rendjét ismerve a becslés már a differenciált soron történik.

3. Az ARMA-modell rendjének kiválasztása és a modell paramétereinek becslése

Az egyszerűség kedvéért tegyük fel, hogy az ARIMA-modell várható érték nélküli.

Ezután a modell rendjének kiválasztása a következőképpen megy: a 2.-es pontban leírt egységgyök teszt eredményének megfelelően differenciáljuk a sorunkat, majd a kapott sorhoz ARMA-modellt keresünk. A p AR és q MA rendek kiválasztásához a BIC-re épülő ún. büntető függvényes módszert használjuk, tehát azon ARMA(p,q) modellt keressük, amelyre a $BIC_{p,q}$ értéke a lehető legkisebb.

$BIC_{p,q} = \log(\hat{\sigma}^2_{p,q}) + (p+q)\log(n-d)/(n-d)$, ahol $\hat{\sigma}^2_{p,q}$ a fehérzaj szórásnégyzetének ML becslése.[11]

Az eljárás a következőképpen történik:

$BIC_{p,q}$ számítása különféle $p=(p_r, p_s)$ reguláris és szezonális autoregresszív rendekre és $q=(q_r, q_s)$ reguláris és szezonális mozgóátlag rendekre HR-módszerrel.

A szezonális rész p_s+q_s rendjét a BIC hajlamos túlbecsülni, ezért több legkisebb BIC-ű modell közül választjuk ki a „legtakarékosabb” szezonális részűt és lehetőség szerint kiegyensúlyozott (azaz $p_s+D=q_s$) modellt.

[11] A BIC-et más alakban is gyakran felírják, így az itt ismertetett alak exponenciális függvényét, vagy (n-d)-vel felszorzott verzióját. Ezek a modell szelekció szempontjából ekvivalens felírások.

Az algoritmus úgy lett kialakítva, hogy minél kevesebb számítást igényeljen. A lényege, hogy először egy speciális ARMA-modellt rögzítünk reguláris résznek, és a szezonális részre vonatkozóan minimalizáljuk a BIC-et, majd az így kapott szezonális részt rögzítjük, és a reguláris részre vonatkozóan minimalizáljuk a BIC-et. Végül az így kapott reguláris részt rögzítjük, és még egyszer minimalizáljuk a BIC-et a szezonális részre vonatkozóan.

A BIC minimalizálásánál figyelembe vesszük a rendekre vonatkozó korlátokat, ami általában $0 \leq p_r$, $q_r \leq 3$; $0 \leq p_s$, $q_s \leq 2$. (Ha a TRAMO-t a SEATS-cel együtt használjuk, akkor $0 \leq p_s$, $q_s \leq 1$.)

A fenti módszer azért megfelelőbb, mint a BIC minden egyes (a korlátozásnak megfelelő) p,q-ra való számolása, mert lényegesen kisebb a számítási igénye: nem csak kevesebb BIC értéket kell kiszámítani, de az egyes BIC értékek számolása során is kevesebb paramétert kell becsülni, hiszen vagy a szezonális vagy pedig a reguláris rész rögzítve van.

Az algoritmus során a BIC számolásához szükség van a fehérzaj ML varianciabecslésére. Ehhez egzakt ML becslés helyett az ARMA(p,q) modell paramétereit a Hannan-Rissanen (HR) módszer egy változatával becsüljük. Ez ugyanis kevésbé számítás igényes.

A HR-módszer lényege röviden, hogy egy invertálható ARMA-modell transzformálható végtelen rendű AR-modellé, amit közelíthetünk véges rendű AR-modellel. Tehát alkalmasan nagy rendű AR-modellt becslünk, majd a becsült reziduumokat visszaírva az eredeti ARMA-modell késleltetett innovációinak helyére, közönséges legkisebb négyzetes becslést alkalmazhatunk. Az így becsült paraméterek melletti modellel kapott reziduumokból számolható a reziduumok varianciája.[12]

Az ARMA-modell identifikációja után, a paraméterek számának ismeretében a paraméterek már becsülhetők feltételes vagy feltétel nélküli legkisebb négyzetes eljárással vagy egzakt maximum likelihood becsléssel. Az egzakt ML becslés gyorsan számolható Kalman-szűrővel, így ez a módszer az ajánlott. A numerikus számításhoz szükséges kezdő értékek beállítására a most ismertetett HR-módszer használatos.

[12] Az első néhány reziduumot ahelyett, hogy 0-val becsülnénk, inkább a Kalman-szűrővel számoljuk.

4. Automatikus outlier felismerés és korrekció

Célja az idősorban esetlegesen előforduló outlierek (kiugró értékek) automatikus felismerése és az idősor megtisztítása hatásuktól. Az outlierek kezelése különös körültekintést igényel, mert outlierek figyelmen kívül hagyása következtében helytelen modellt határozhatunk meg, a modell paramétereinek becslését erősen torzíthatják a nem kezelt outlierek, illetve több outlier együttes jelenléte zavarhatja egymás felismerését (maszk hatás).

Egy outlier hatását általánosan a következőképpen jellemezhetjük:

$$z_t^* = z_t + \omega \nu(B) I_t^T ,$$

ahol $\nu(B)$ B polinomjainak hányadosa és az outlier típusának jellemzésére szolgál; z_t az outliertől mentes idősor; z_t^* a megfigyelt idősor; az I_t^T az outlier megjelenési helyének indikátor függvénye: értéke 1 ha t=T, és 0 egyébként; az ω az outlier hatásának nagyságát adja meg. Az outlierek következő négy típusát vehetjük figyelembe:

AO (additív outlier): $\nu(B) = 1$,

TC (temporary /transitory/ change – csillapodó jellegű törés): $\nu(B) = 1/(1 - \lambda B)$,

LS (level shift – szinteltolódás): $\nu(B) = 1/(1 - B)$,

IO (innovációs outlier): $\nu(B) = \theta(B)/(\delta(B)\phi(B))$, ahol θ, ϕ és δ az idősorra meghatározott ARIMA-modell mozgóátlag, autoregresszív, illetve differenciálási polinomjai.

A TC-nél a λ értéke 0,7-re van beállítva. Az innovációs outliert a gyakorlatban nem használják, mivel hatására az idősor szintje lényegesen megváltozik: Gómez és Maravall (2001) tapasztalata szerint ARIMA-modellek identifikációjánál és illesztésénél nem segít inkább csak zavar az IO kezelése.

Tegyük fel először, hogy egy outlier van, mégpedig t=T-nél, és ismerjük az outlier mentes idősor ARIMA-modelljét paraméterekkel együtt. Ekkor a következő ún. regARIMA-modellt (regresszió ARIMA-hibataggal) írhatjuk fel:

$$z^* = Y\omega + z , \quad (1)$$

ahol z^* a megfigyelt idősor értékeiből képzett vektor, azaz $(z_1^*,\ldots,z_n^*)'$, hasonlóan z az outlier nélküli idősor vektora, és $Y=(\nu(B)I_1^T,\ldots,\nu(B)I_n^T)'$ az outlier vektora. Feltesszük, hogy z ARMA-folyamat, azaz $\delta(B)=1$; ha nem ez a helyzet akkor a továbbiakban a z^*, z és Y differenciálásával kapott vektorokkal dolgozunk.

Felírjuk z kovariancia mátrixát: Var(z)=$\sigma^2\Omega$, ahol σ^2 a (z-hez tartozó) fehérzaj szórásnégyzete, az Ω pedig az ARMA-modell autokovariancia függvényéből adódik, ami a modell és paraméterek ismeretében kiszámolható. Vegyük az Ω mátrix Cholesky felbontását[13] (ezt megtehetjük, mert σ^2 Ω kovariancia mátrix, így Ω pozitív definit), azaz $\Omega=LL'$, ahol L alsó háromszög mátrix, ami azt jelenti, hogy a főátló feletti összes elem 0. Az (1) egyenletet balról szorozva L^{-1}-zel egy közönséges legkisebb négyzetes (OLS) feladatot kapunk, mert az $L^{-1}z$ fehérzaj: $Var(L^{-1}z)=L^{-1}Var(z)(L^{-1})'=\sigma^2(L^{-1}L)(L'L'^{-1})=\sigma^2I_n$, ahol I_n az nxn-es egységmátrix. Tehát a következő egyenlethez jutottunk:

$$r^*=X\omega+r\,,\quad (2)$$

ahol $r=L^{-1}z$, $r^*=L^{-1}z^*$ és $X=L^{-1}Y$.

A TRAMO-program a számolást a Kalman szűrő segítségével végzi, és így kapja meg z^*-ból r^*-ot és Y-ból X-et. Majd (2) alapján OLS-becsléssel kapjuk ω becslését, valamint a próbastatisztikát az $\omega=0$ nullhipotézis vizsgálatához:

$$\widetilde{\omega}=(X'X)^{-1}X'r^*\quad (3)$$

$$Var(\widetilde{\omega})=(X'X)^{-1}\sigma^2\,,$$

a τ próbastatisztika $\omega=0$ teljesülése esetén (azaz amikor nincs outlier) standard normális eloszlású:

$$\tau=(X'X)^{1/2}\widetilde{\omega}/\sigma\,.\quad (4)$$

A gyakorlatban persze az ARIMA-modell paramétereit nem ismerjük ezért becsülni kell azokat. Ezt z^*-ra végezzük el egzakt ML-becsléssel (tehát úgy

---

[13] Lásd pl. Hamilton (1994, 91pp).

járunk el, mintha nem lenne outlier), majd ω és τ becslését a paraméterek becsléseit felhasználva kapjuk:

$$\widehat{\omega} = (\widehat{X}'\widehat{X})^{-1}\widehat{X}'\widehat{r}^*$$

$$\widehat{\tau} = (\widehat{X}'\widehat{X})^{1/2}\widehat{\omega}/\widehat{\sigma}\,.$$

Ha kiváncsiak vagyunk, hogy tetszőleges t=T-ben vajon van-e outlier, és milyen típusú, akkor kiszámoljuk az $\widehat{\omega}_{AO}^T$, $\widehat{\omega}_{TC}^T$, $\widehat{\omega}_{LS}^T$, $\widehat{\omega}_{IO}^T$ becsléseket és a hozzájuk tartozó $\widehat{\tau}$ próbastatisztikákat minden T-re, ezeket t-értékeknek hívjuk, annak ellenére, hogy nem t-eloszlású a próbastatisztika. Majd vesszük a próbastatisztikák maximumát minden T-re illetve tp={AO,TC,LS,IO}-ra: $\Lambda = \max_{T,tp} |\widehat{\tau}_{tp}^T|$. Ezután az így kapott t-értéket összehasonlítjuk egy előre definiált C értékkel, ami legtöbbször az idősor hosszától függ (minél hosszabb az idősor, annál nagyobb). Ha $\Lambda > C$, akkor van outlier; ennek az outliernek a helye az a T időpont, típusa pedig az a tp, amely (T,tp) pár maximalizálta a Λ-t. Ha $\Lambda \le C$, akkor nincs outlier az idősorban.

Több outlier becslésére két módszer kínálkozik. Az egyik, hogy a fenti, egy outlier detektációjára való módszert többször egymás után alkalmazva sorban kapjuk az outliereket. A másik módszer az outliereket egyszerre becsli többszörös regresszióval (ilyenkor ismerni kell a potenciális outlierek helyét, mert az összes lehetséges kombináció végignézése túl sok számítást igényelne). A két eljárás nem ugyanazt az eredményt adja. A TRAMO egy olyan lépésenkénti (stepwise) megközelítést alkalmaz, melynek során az outlierek bevonásához az egyszeres regressziós módszert, míg a kivételéhez a többszörös regressziót alkalmazza (részletesebben lásd később). A többszörös regresszió egyenlete:

$$z_t^* = z_t + \sum_{i=1}^{k}\omega_i \nu_i(B) I_t^{T_i}\,.$$

Hasonlóan járunk el mint az egyszeres esetben: (3) és (4) alapján számolunk, feltéve, hogy az ARIMA-modell paraméterei ismertek. Ekkor persze ω az ω_i-k vektora, Y és így X is mátrix. Emiatt az $\widetilde{\omega}$ számolása mátrix inverziót igényelne, ami helyett az ilyenkor szokásos módon QR-felbontást használ a program, ami hatékony és numerikusan stabil módszer. A QR-felbontást Householder transzformációval valósítja meg.

A számítások során szükség van a reziduumok σ szórásának egy olyan robusztus becslésére, amely nem érzékeny az outlierek jelenlétére. Ezért a MAD (mean absolute deviations) felhasználásával becsüljük σ-t:

$$\hat{\sigma} = 1{,}483 \cdot medián\{| r_t^* - medián(r_t^*) |\}$$

ahol a szorzó 1, 483-as értéke az r_t^* feltételezett normalitásán alapszik.

Nézzük most az automatikus outlier felismerő eljárás algoritmusát, abban az esetben, ha ismerjük az ARIMA-modell (p,d,q) rendjét, de paramétereit nem! Lépésenkénti outlier kezelés történik, hasonló ahhoz az eljáráshoz, amellyel egy regressziónál kiválaszjuk a legjobb regresszorokat. Az outliereket egyesével vonjuk be, mindegyik után újra becsüljük az ARIMA-modell paramétereit. Mikor már nem találunk több outliert, a fentebb ismertetett többszörös regressziót alkalmazzuk, és így t-értékeket kapunk a bevont outlierekre. Amennyiben van olyan outlier, amelynek t-értéke kisebb a C konstansnál – amely az outlierek egyesével való bevonásánál is szerepelt –, akkor a legkisebb t-értékűt kivesszük. Ha nincs ilyen outlier, akkor készen vagyunk.

Az eljárást csak kicsit kell módosítani, amennyiben adott változókkal való regressziót is beépítünk a modellbe:

$$z_t = y_t{}'\beta + \nu_t ,$$

ahol z_t a megfigyelt idősor, y_t a regressziós változók t időpontra vonatkozó értékeiből alkotott vektor, β a regressziós együtthatók vektora (aminek első eleme lehet a várható érték), ν_t pedig ARIMA-modell 0 várható értékkel.

Tehát az alábbi ilyenkor az eljárás (2. ábra).

A. Inicializálás

Amennyiben vannak regressziós változók, akkor a regressziós együtthatókat közönséges legkisebb négyzetes (OLS) becsléssel becsüljük (mintha ν_t fehérzaj lenne), majd az így kapott együtthatókkal számolt regressziós hatástól megtisztítjuk az idősort.

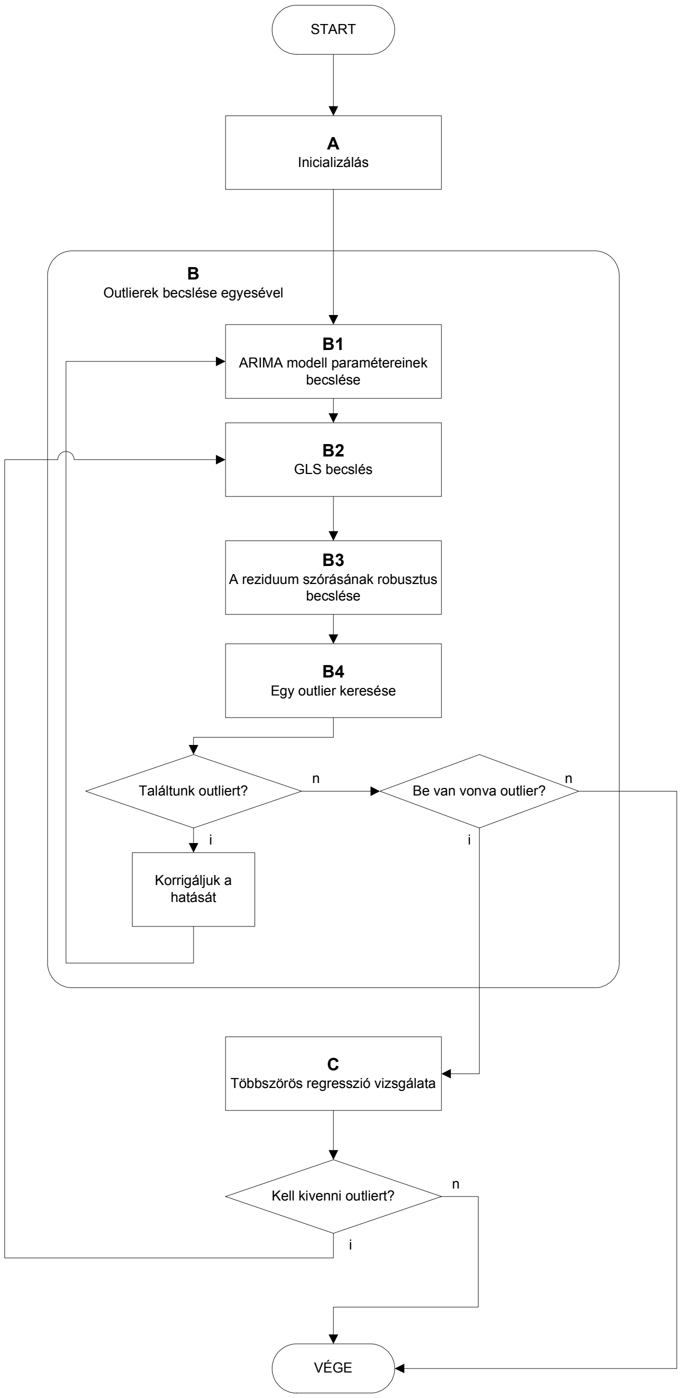
B. Outlierek egyesével való becslése

1. Az ARIMA-modell paramétereit megbecsüljük, HR-módszerrel vagy egzakt ML-dal.
2. Az ARIMA-modell paramétereinek becsléseit rögzítjük, a regressziós együtthatókat – beleértve a már bevont outlierek együtthatóit is – pedig GLS-becsléssel (általánosított legkisebb négyzetese becslés) becsüljük (azaz úgy, mint ahogy az outlierek regressziós becslésénél leírtuk, tehát figyelembe véve a hibatag korreláltságát). Így új becslést kapunk a reziduumokra.
3. A becsült reziduumokból becsüljük a reziduumok szórását a korábban ismertetett robusztus módon.
4. Most következik az egy outlier keresése, a korábban leírt módon, tehát t-értékek számolása minden lehetséges időpontra és outlier típusra. Ehhez felhasználjuk a 2. pontban kapott reziduumokat, illetve a 3. pontban kapott szórásbecslést. Ha nem találtunk outliert, és nincs is bevonva outlier, akkor megállhatunk, az idősorban nincs outlier. Ha nem találtunk outliert, de be van vonva outlier, akkor a C. pontra ugrunk. Ha találtunk egy outliert, akkor ennek a hatásától is tisztítsuk meg az idősort, és folytassuk az eljárást az 1.-es ponttól.

C. Többszörös regresszió vizsgálata

A B.2-nél kapott becslésekhez tartozó t-értékek alapján döntünk: ha van olyan outlier, amelynek t-értéke kisebb mint a C konstans, akkor a legkisebb t-értékű outliert kivesszük, és folytatjuk az eljárást a B.2.-es ponttól.

2. ábra: Outlierek felismerése és korrekciója



5. A kombinált algoritmus

A 4.résznél lévő eljárásnál feltettük, hogy ismerjük az ARIMA-modell (p,d,q) rendjét. Ha nem ez a helyzet, akkor a fenti algoritmust kombinálni kell a korábban a 3. résznél ismertetett automatikus identifikációs eljárással. Ez az algoritmus fogja össze az eddig tárgyalt összes becslési eljárást, és így a TRAMO számítási menetének teljes ismertetésének tekinthető (3. ábra):

A. Előzetes tesztek

Log-teszt (lásd 1.rész), munkanaphatás tesztelése, húsvéthatás tesztelése. A munkanap- és a húsvéthatás teszteléséhez most feltesszük, hogy az idősor modellje airline, azaz ARIMA(0,1,1)(0,1,1).

B. Inicializálás

Ha a felhasználó nem kér outlier kezelést, akkor ugrás a C.1-re.

Ha kér, és adott meg modellt, akkor azt elfogadjuk, különben az airline modellt választjuk. C.3-ra ugrunk.

C.1

Ha a felhasználó megadta a differenciálás rendjét és hogy kell-e várható értéket használni a modellben, akkor ugrás a C.2-re. Különben ha van regressziós hatás akkor az idősort megtisztítjuk a hatásuktól. Ezután automatikusan döntünk a differenciálás rendjéről és a várható érték használatáról (lásd 2. rész). Ugrás a C.2-re.

C.2

Outlierektől és más regressziós hatásoktól – ha van – megtisztítjuk az idősort, majd ARMA(p,q) modell automatikus identifikációja a differenciált sorra (lásd 3. rész). A munkanap- és húsvéthatás szignifikanciáját ellenőrizzük az új modellre.

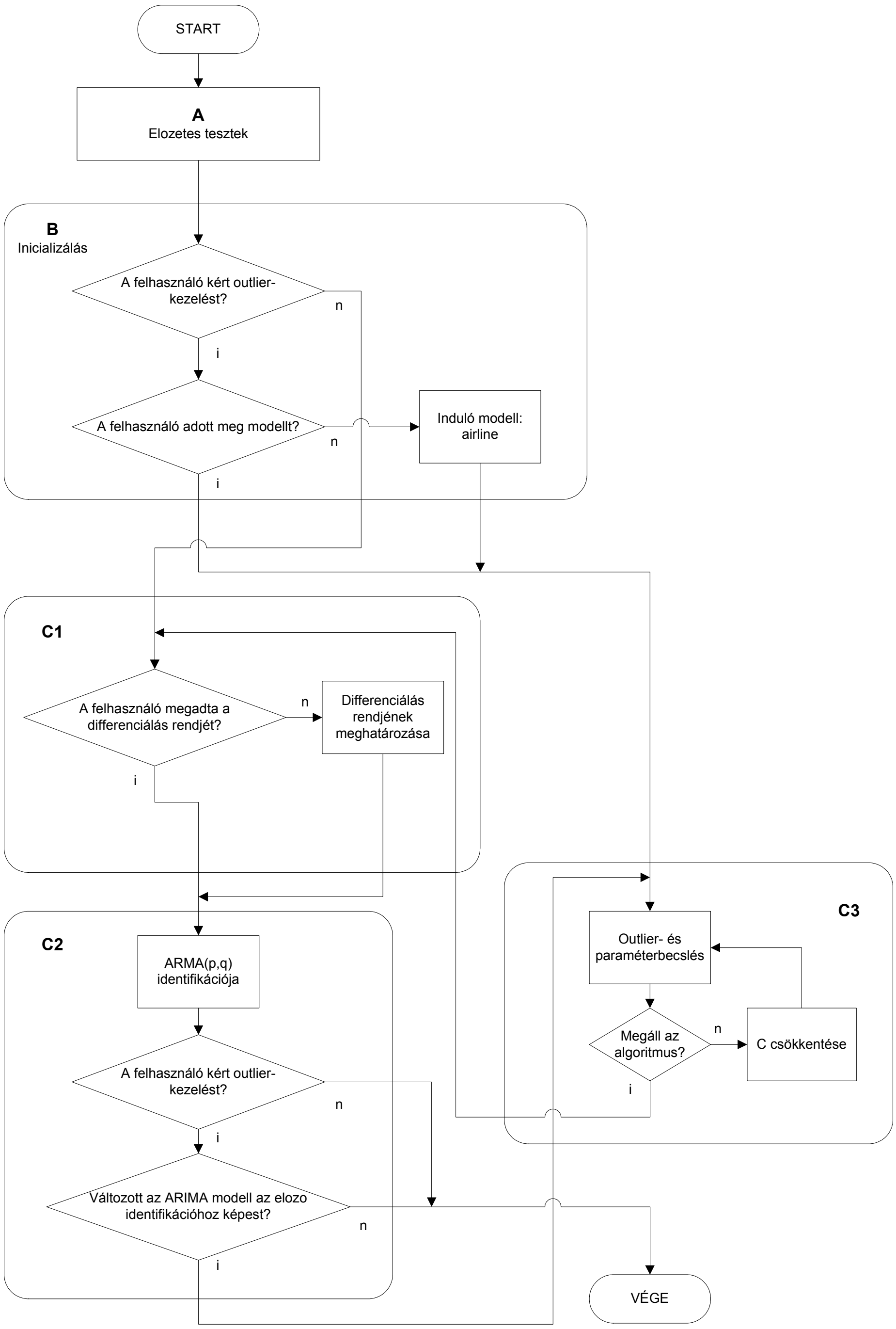
Ha a felhasználó akar outlier kezelést, akkor ugrás a C.3-ra, különben kész vagyunk és megállunk.

C.3

Feltéve, hogy ismerjük a modellt, a 4. részben ismertetett módon automatikusan detektáljuk az outliereket. Ha nem áll meg az algoritmus, akkor csökkenthető a C kritikus érték. Visszamegyünk C.1.-re.

A fenti algoritmusnál három kör után, ha még mindig nem jó az illeszkedés, akkor a differenciált sorra ARMA(3,1)(0,1) modellt állítunk be, a differenciálás rendje pedig az utolsó körben kapott érték.

3. ábra: Kombinált algoritmus



6. A munkanap- és a húsvéthatás kezelése

Mint azt korábban említettük, a munkanaphatás az a jelenség, amikor az egy időszakra vonatkozó megfigyelés értékét befolyásolja az időszak munkanapjainak száma.

A program a munkanaphatást regressziós változókkal jellemzi, és az előző pontban leírt ciklus során szűri ki a TRAMO az idősorból. Attól függően használhatunk 1 vagy 6 regresszort, hogy a munkanaphatásnál csak a munkanapokat és a szabadnapokat (szombat, vasárnap és ünnepnap) különböztetjük meg, vagy a hét minden egyes napjának különböző hatást tulajdonítunk. Az 1 és a 6 regresszoros esetben is a változók értékei úgy lettek kialakítva, hogy az egy hétre vonatkozó hatás nulla legyen. Az 1 regresszoros esetben a regressziós változó értéke minden egyes időszakra az időszak munkanapjainak száma mínusz a szabadnapok száma szorozva 5/2-del. A 6 regresszoros esetben a regressziós változók: a hétfők száma mínusz vasárnapok száma, keddek száma mínusz vasárnapok száma, …, szombatok száma mínusz vasárnapok száma, ahol a vasárnapokhoz számítanak az esetleges ünnepnapok is.

A fentiekből látható, hogy a munkanappal való korrekció nem valamiféle átlagos munkanap számhoz történő igazítást jelent.

Lehetőség van a szökőéveknél jelentkező februári szökőnap, azaz a hosszabb február hónap hatásának figyelembe vételére is. Ez még egy regresszort jelent. Értéke februárok kivételével nulla, szökőévben lévő (29 napos) februárnál 0,75, nem szökőévben lévő februárok esetén −0,25 (a változó értékeinek kialakítása mögött megint az a logika, hogy egy olyan időintervallumban, ahol már nem érvényesülhet a hatás, nulla legyen: itt négy évre összegezve kell nullát kapnunk).

A TRAMO lehetőséget nyújt a húsvét hatásának kiemelt kezelésére is. A húsvétot egyszer figyelembe vesszük a munkanaphatás kezelésekor mint munkaszüneti napot (1 regresszornál hétvégi napnak, 6 regresszornál vasárnapnak számít), másrészt a megelőző d nap hatását figyelembe véve alakítunk ki egy külön regressziós változót a húsvéthatás kezelésére. Havi gyakoriságú idősort feltételezve: ennek március és április hónapok kivételével az összes hónapra nulla az értéke. Márciusra a változó értéke $p_M - m_M$, ahol p_M a d napon belül azon napok aránya, amelyek márciusra esnek, m_M pedig a sok éven át vett átlaga az ilyen p_M értékeknek. Az áprilisi hónapokhoz rendelt érték $p_A - m_A$, ahol p_A és m_A hasonlóképpen van definiálva. Jó közelítéssel $m_M = m_A = 1/2$. Ily módon a március és április hatása együttesen nulla, hiszen $p_A = 1 - p_M$.

A figyelembe vett napok száma, azaz d értéke állítható, alapbeállítás a $d = 6$.

2.10. A SEATS

A program az idősorok nem megfigyelt komponensekre bontásának úgynevezett ARIMA-modell alapú módszerei közé tartozik. Az itt szereplő ismertetés Maravall (1995) alapján készült, hasznos referencia még Planas (1997). A SEATS-et Burmannek az Angol Nemzeti Bankban szezonális kiigazításra írt programjából (1982-es verzió) fejlesztették ki. A programot a havi vagy ritkább gyakoriságú adatokkal való munkára szánták.

A SEATS feltételezi, hogy a szezonálisan kiigazítandó idősor lineáris, normális fehérzaj innovációkkal. Amikor ez a feltételezés nem teljesül, a SEATS képes együttműködni a TRAMO-val, ami eltávolítja a sorból a különleges hatásokat, azonosítja és eltávolítja a különféle típusú outliereket, és interpolálja a hiányzó megfigyeléseket a már korábban leírtak szerint. Ekkor az ARIMA-modellt is a TRAMO-tól veszi át.

A program a sort több különféle komponensre bontja. A dekompozíció lehet multiplikatív, vagy additív. Mivel az előbbi logaritmálással az utóbbiba alakítható, ezért jelen tárgyalásban az additív modellel foglalkozunk, azaz:

$$x_t = \sum_i x_{it},$$

ahol x_{it} jelöl egy komponenst. A SEATS által figyelembe vett komponensek a következők:

$x_{pt} =$ a TREND komponens,

$x_{st} =$ a SZEZONÁLIS komponens,

$x_{ct} =$ a CIKLIKUS komponens,

$x_{ut} =$ az IRREGULÁRIS komponens.

A szezonálisan kiigazított idősor a trend komponens, a ciklikus komponens és az irreguláris komponens összege, azaz a megfigyelt idősor mínusz a szezonális komponens. Mivel gyakori, hogy ciklikus komponens nincs az idősor felbontásánál (a megfigyelt idősorokra leggyakrabban illesztett airline modell esetén például ez a helyzet), ezért sokszor előfordul, hogy egyszerűen azt mondjuk, hogy a szezonálisan kiigazított idősor a trend és az irreguláris komponens összege.

A komponenseket a spektrum (illetve nemstacionárius esetben a pszeudospektrum) alapján karakterizáljuk: a trend komponens képviseli az idősor hosszú távú fejlődését, és spektrális csúcsként jelenik meg a nulla frekvenciánál. Úgy is mondhatjuk, hogy a trend egy olyan ciklus, amelynek végtelen nagy a periódus ideje. A szezonális komponens hatását a szezonális frekvenciáknál jelentkező spektrális csúcsok mutatják. Az irreguláris komponens képviseli a szabálytalan, fehérzaj viselkedést, és így lapos (konstans) spektruma van. A ciklikus komponens képviseli a szezonálisan kiigazított sor trendjétől való, a tiszta fehérzajtól különböző eltéréseket.

A komponensekre bontás a következő feltételezésekre épül:

1. Mindegyik komponens egy-egy ARIMA-folyamat realizációja. Képlettel:

$$\phi_i(B)x_{it} = \theta_i(B)a_{it},$$

ahol a_{it} fehérzaj változó, és a ϕ_i polinomok már tartalmazzák a differenciaképzés polinomját is (azaz egységgyökeik is lehetnek).

2. A különböző komponensek ARIMA-modelljeihez tartozó innovációk egymással korrelálatlanok, azaz $Cov(a_{it}, a_{ju}) = 0$, ahol $i \neq j$. Ez persze azt is jelenti, hogy a komponensek korrelálatlanok.

A feltételezés mögött az a meggondolás áll, hogy a különféle komponenseket egymástól különböző tényezők alakítják.

3. A komponensek autoregresszív polinomjainak, azaz a ϕ_i-knek nincs közös gyökük.

A feltétel megtehető az általánosság csorbítása nélkül, mivel különböző autoregresszív gyökök különböző frekvenciáknál okoznak csúcsot a spektrumban, és a különféle komponensek pedig különböző frekvenciájú csúcsokhoz rendelhetők.

Ennek a feltételnek és az előzőeknek az alapján a megfigyelt idősor olyan ARIMA-modellel írható le, ahol az ARIMA-modell autoregresszív polinomja a komponensek autoregresszív polinomjainak szorzata.

Képlettel:

$$\phi(B)x_t = \theta(B)a_t,$$

ahol a_t fehérzaj innováció, a ϕ autoregresszív és θ mozgóátlag polinomok pedig a következőképpen fejezhetők ki a komponensek autoregresszív illetve mozgóátlag polinomjainak segítségével:

$$\phi(B) = \prod_{i=1}^{k} \phi_i(B),$$

$$\theta(B)a_t = \sum_{i=1}^{k} \theta_i(B) \prod_{\substack{j=1 \\ j \neq i}}^{k} \phi_j(B) a_{it} .$$

A felírás alapja az a tény, hogy korrelálatlan mozgóátlag folyamatok összege is mozgóátlag folyamat.

4. A komponensek mozgóátlag polinomjainak, azaz a θ_i-knek nincs közös egységgyökük.

Ez a feltétel biztosítja, hogy a megfigyelt idősor ARIMA-modellje invertálható legyen, míg a komponenseknél megengedjük a neminvertálhatóságot. A megfigyelt idősor ARIMA-modelljének invertálhatósága pedig ahhoz kell, hogy a később bemutatott Wiener-Kolmogorov szűrő konvergens legyen.

5. A megfigyelt idősor ARIMA-modellje, azaz a ϕ autoregresszív polinom, a θ mozgóátlag polinom és az a_t innovációk varianciája ismert.

Ez a feltétel az ARIMA-modell becslésével teljesíthető.

A fentebbiekben felírt feltételek még nem elegendőek, hogy egyértelműen meghatározzuk a komponensek ARIMA-modelljeit. Annak érdekében, hogy egyértelműen azonosítsuk a komponenseket, feltesszük, hogy azok az irreguláris komponens kivételével zajtól mentesek. Ezt hívják kanonikus tulajdonságnak, és ez azt jelenti, hogy az irreguláristól különböző komponensekből nem lehet kinyerni hozzáadott fehérzajt, azaz a trend, a szezonális komponens és a ciklikus komponens spektruma is felveszi a minimális nulla értéket (ily módon ezek a komponensek neminvertálható ARIMA-modellel rendelkeznek). Az irreguláris komponens szórása ily módon maximalizált, ezzel szemben a trend, szezonális és ciklikus komponens annyira stabil, amennyire csak lehetséges. Tetszőleges spektrumfelbontásból kapható kanonikus felbontás: egyszerűen az irreguláris komponenstől különböző komponensek spektrumából kivonjuk az adott komponens spektrumának minimum értékét, és az így kivont számok összegével megnöveljük az irreguláris komponens spektrumát.

Most nézzük, hogy konkrétan milyen módon dolgozik a SEATS!

A SEATS először felbontja a megfigyelt idősor ARIMA-modelljét, azaz megállapítja a komponensek ARIMA-modelljeit. Ez frekvencia-tartományban történik. A spektrumot – ami

$$\frac{\sigma^2}{2\pi}\frac{\left|\theta(e^{-i\omega})\right|^2}{\left|\phi(e^{-i\omega})\right|^2}$$

alakban írható fel – felosztjuk a különböző komponensekhez tartozó spektrumok összegére. Mivel tudjuk, hogy a komponensek AR-polinomjainak szorzataként áll elő a megfigyelt idősor AR-polinomja, ezért az első feladat az idősor AR-polinomjának felbontása. A megfigyelt idősorra vonatkozó ARIMA-modell autoregresszív polinomjának gyökeit a program ezért hozzárendeli valamelyik komponens autoregresszív polinomjához a gyököknek megfelelő spektrális csúcsok frekvenciáját figyelembe véve, mert a spektrális csúcs a frekvenciának megfelelő periodikus viselkedést jelez. A SEATS a gyakorlatban a gyökök argumentuma alapján dönt, ami legtöbbször közel esik a spektrális csúcs frekvenciájához.[14]

[14] A gyök(ök) alapján a spektrum maximumának frekvenciája a következő: valós gyök esetén 0 vagy π, komplex konjugált gyökpár esetén 0, π vagy egy olyan ω frekvencia, amelyre $\cos\omega = \frac{(1+r^2)}{2r}\cos\alpha$, ahol a gyökpár abszolút értéke r, argumentuma pedig $\pm\alpha$. Egység abszolút értékű gyökpár esetén végtelenhez tart a spektrum értéke a gyök argumentumával egyező frekvenciánál.

A 0 frekvenciához tartozó nagy abszolút értékű gyököket a trendhez, a szezonális frekvenciákhoz (pl. negyedéves idősornál: $\left[\frac{\pi}{2}-\varepsilon,\frac{\pi}{2}+\varepsilon\right]$ és $\left[\pi-\varepsilon,\pi\right]$, ahol ε a programban rögzített kicsi szám) tartozó gyököket a szezonális komponenshez, a kis abszolút értékű 0 frekvenciához tartozó és a ciklikus (a 0 és az első szezonális frekvencia között) valamint a szezonális frekvenciák közötti frekvenciákhoz tartozó gyököket a ciklikus komponenshez rendeljük. Az irreguláris komponenst mindig fehér zajnak tekintjük.

Nézzük meg egy konkrét ARIMA-modell esetén, hogy hogyan zajlik az AR-polinom felbontása! A vizsgált modell legyen az egyik legegyszerűbb és a gyakorlatban leggyakrabban előforduló modell, az airline modell, azaz az ARIMA(0,1,1)(0,1,1), és tegyük fel, hogy havi gyakoriságú idősorunk van. Ekkor a megfigyelt idősor modellje a következőképpen írható fel:

$$(1-B)(1-B^{12})x_t=(1-\theta_1 B)(1-\Theta_1 B^{12})a_t.$$

Az AR-polinom: $(1-B)(1-B^{12})=(1-B)^2(1+B+B^2+\cdots+B^{11})$. Ezért a polinom gyökei:

– az 1 kétszeres gyök és a 0 frekvenciához tartozik,
– 10 db egység abszolút értékű komplex és 1 db valós gyök: $\cos(k\cdot 2\pi/12)+i\sin(k\cdot 2\pi/12)$, ahol k 1-től 11-ig fut, és k=6-ra kapjuk a -1-et gyöknek. Ezek a gyökök a szezonális frekvenciákhoz tartoznak, azaz $\frac{k\pi}{6}$-hoz ($k=1\ldots6$).

Az eljárás alapján a kétszeres 1 gyököt a trendhez, a szezonális frekvenciákhoz tartozó 11 db gyököt pedig a szezonális komponens AR-polinomjához rendeljük.

Tehát a trend modellje:

$$(1-B)^2 x_{pt}=\theta_p(B)a_{pt}. \quad (1)$$

Ehhez hasonló trend modelleket gyakran használnak ökonometriai modellekben a trend jellemzésére. A modell a $\beta_0+\beta_1 t$ determinisztikus trend általánosításaként is felfogható, hiszen teljesül, hogy

$$(1-B)^2(\beta_0+\beta_1 t)=(1-B)\beta_1=0,$$

azaz az általánosítás abban áll, hogy nem 0-át, hanem egy mozgóátlag folyamatot feltételezünk a (1) egyenlet jobb oldalán.

A szezonális komponens modellje az AR-polinomjához rendelt gyökök alapján:

$$(1 + B + B^2 + \cdots + B^{11})x_{st} = \theta_s(B)a_{st} \,. \quad (2)$$

Ismét gyakran használt modellről van szó, és a determinisztikus szezonális komponens általánosításának tekinthető. A determinisztikus szezonális komponensre jellemző, hogy a szezonális eltérések értékeinek (egy évre vonatkozó) összege 0, azaz

$$x_{st} + x_{s(t-1)} + x_{s(t-2)} + \cdots + x_{s(t-11)} = 0 \,.$$

A (2) egyenletben a „jobb oldal = 0" feltétel lett kicserélve mozgóátlag folyamatra.

Most térjünk vissza a SEATS ismertetéséhez!

A trend, szezonális és ciklikus komponensekre vonatkozó kanonikus feltétel egyértelmű felbontást határoz meg, amiből a komponensek ARIMA-modelljei megkaphatóak (beleértve a komponensek innovációinak varianciáit). A számolás úgy történik, hogy a megfigyelt idősor és a komponensek spektrumára egyenletet oldunk meg, kihasználva, hogy a komponensek spektrumainak összege egyenlő a megfigyelt idősor spektrumával. Bizonyos tagokat már ismerünk az egyenletben, így a megfigyelt idősor spektrumát és a komponensek spektrum felírásának nevezőjét, mert azt a komponens ARIMA-modelljének AR-polinomjából számoljuk ki. Az ismeretlenek a komponensek MA-polinomjainak együtthatói, amik a spektrum felírásánál a számlálót határozzák meg. A spektrumokra vonatkozó egyenletből egy egyenletrendszert lehet felírni, annak alapján, hogy az egyenletünk bal oldalán és jobb oldalán szereplő $\cos(k\omega)$ tagok együtthatóinak meg kell egyezniük (itt k egész szám és ω függvényében vannak felírva a spektrumok). Néhány ismeretlen együtthatónak tetszőleges értéket adhatunk (mert alulhatározott az egyenletrendszer, ezért kell a kanonikus feltétel), célszerűen 0-át, majd az így megoldott egyenletből kapott komponenseket kanonikussá tesszük a már leírt módon, azaz a minimum értékek levonásával.

Ha az ARIMA-modell nem tesz lehetővé elfogadható dekompozíciót, akkor a SEATS helyettesíti egy elfogadható közelítéssel. Egy dekompozíció nem elfogadható, ha valamelyik komponens spektruma negatív értéket vesz fel.

Egy konkrét $[x_1, x_2, \ldots, x_T]$ idősor-realizációra a program kiszámítja a komponensek minimum átlagos négyzetes hiba (Minimum Mean Square Error /MMSE/) becslését, amit idősoroknál optimális becslésnek tekintünk. Ezt úgy számolja ki, hogy egy ún. Wiener-Kolmogorov-típusú szűrőt alkalmaz a véges idősorra a komponensek ARIMA-modellje alapján, az idősort előbb kiterjesztve (előrejelezve illetve „visszajelezve”) a „jövő” illetve a „múlt” felé. Minden $t = 1, \ldots, T$-re és minden i komponensre kiszámítja az $\widehat{x}_{it|T}$ becslést, ami egyenlő az $E(x_{it} | x_1, \ldots, x_T)$ feltételes várható értékkel.

A Wiener-Kolmogorov-szűrő most következő levezetése stacionárius idősor esetére alkalmazható, az eljárást később nemstacionárius idősorok esetére kiterjesztették (ezt nem részletezzük). Az általánosság kedvéért nem egy konkrét komponens becslésére ismertetjük az eljárást, hanem egy tetszőleges, ARMA-modellel jellemzett jelre, amit s_t-vel jelölünk. A maradékot, azaz $x_t - s_t$-t n_t-vel jelöljük, és feltesszük az s és az n korrelálatlanságát (ez megfelel a komponensekre tett feltevésnek). Feltesszük, hogy a múlt és a jövő felé is végtelen hosszú az idősor. A végtelen hosszú idősor alapján kapott becslést jelöljük $\hat{s}_t$-pal. Az $\hat{s}_t$-ot az idősor megfigyelt értékeiből időinvariáns lineáris szűrő segítségével becsüljük, azaz

$$\hat{s}_t = \sum_{k=-\infty}^{\infty} \nu_k x_{t-k} = \nu(B) x_t, \qquad (3)$$

ahol $\nu(B) = \sum_{k=\infty}^{\infty} \nu_k B^k$ a szűrő. Az MMSE-becslés az s_t-nek a megfigyelt x értékekre vett ortogonális vetítése, ami azt jelenti, hogy $\mathrm{cov}(s_t - \hat{s}_t, x_{t-j}) = 0$ tetszőleges j egész számra. Ebbe behelyettesítve a (3)-ban kapott felírást $\hat{s}_t$-ra, kapjuk az

$$\mathrm{cov}(s_t, x_{t-j}) - \sum_{k=-\infty}^{\infty} \nu_k \,\mathrm{cov}(x_{t-k}, x_{t-j}) = 0 \qquad (4)$$

összefüggést. Mivel az s és az n korrelálatlanságát feltettük, ezért $\mathrm{cov}(s_t, x_{t-j}) = \mathrm{cov}(s_t, s_{t-j} + n_{t-j}) = \mathrm{cov}(s_t, s_{t-j})$ teljesül.

Most áttérünk frekvenciatartományba, azaz az autokovarianciákat az x és s spektruma (g_x és g_s) segítségével írjuk fel, és így kapjuk (4)-ből:

$$\int_{-\pi}^{\pi} e^{i\omega j}\left[g_s(\omega) - \nu(e^{-i\omega})g_x(\omega)\right]d\omega = 0 .$$

Ez minden egész j-re igaz, így következik az alábbi azonosság:

$$g_s(\omega) - \nu(e^{i\omega})g_x(\omega) = 0 .$$

Átrendezéssel kapjuk, hogy

$$\nu(e^{-i\omega}) = \frac{g_s(\omega)}{g_x(\omega)} .$$

Ennek az összefüggésnek a segítségével a szűrő közvetlenül (nem frekvencia tartományban) az x, s és n ARIMA-modelljével a következőképpen írható fel ($F = B^{-1}$, az előretolás operátor):

$$\nu(B) = \frac{V_s}{V_x}\frac{\theta_s(B)\theta_s(F)\phi_n(B)\phi_n(F)}{\theta_x(B)\theta_x(F)}$$

Ezután az s_t MMSE becslése, $\hat{s}_t$ a következőképpen számolható x_t-ből:

$$\hat{s}_t = \left[\frac{V_s}{V_a}\frac{\theta_s(B)\theta_s(F)\phi_n(B)\phi_n(F)}{\theta(B)\theta(F)}\right]x_t .$$

A szögletes zárójelben álló kifejezés a Wiener-Kolmogorov-szűrő. Ez szimmetrikus és konvergens. A konvergenciát a megfigyelt idősor ARIMA-modelljének invertálhatósága (4. feltétel) biztosítja. A konvergencia miatt a szűrő csonkítható, így nem kell végtelen sok megfigyelést figyelembe venni előre illetve hátrafelé. Mivel a szűrő függ az idősor ARIMA-modelljétől, ezért mondhatjuk, hogy a Wiener-Kolmogorov-szűrő alkalmazkodik a sorok különféle sztochasztikus struktúráihoz, ellentétben például az X család rögzített szűrőivel. Így elkerülhető a szezonalitás alul- vagy túlbecslése.

Amikor $T \to \infty$, akkor az $\hat{x}_{it|T}$ becslésből lesz az $\hat{x}_{it}$ „végső” becslés. $t = T$-re a párhuzamos becslést, $\hat{x}_{iT|T}$-t vesszük, ami a sor utolsó megfigyelésének becslése. A gyakorlatban rendszerint az idősor két végétől legalább három év távolságra lévő kiigazított adatok már végső becslésnek vehetők (amennyiben az idősor ARIMA-modellje változatlan marad).

Alkalmazási nézőpontból a végső és a párhuzamos becslések a leginkább érdekesek.

Amikor a TRAMO-t és a SEATS-et együtt futtatjuk, akkor a hatásokat, amiket a dekomponálás érdekében a TRAMO eltávolított az idősorból, a végső komponensekbe visszateszi. Így például a szinteltolódás outliereket a trendhez rendeli, míg a csillapodó jellegű törés és az additív outlierek az irreguláris tagba kerülnek; a szezonális komponens tartalmazni fogja a munkanap- és húsvéthatást. Az outlierekre vonatkozó választást az indokolja, hogy a szinteltolódás hosszú távon hat, míg a másik két típusú kiugró érték rövid távon. Tehát a szezonálisan kiigazított sor tartalmazza az összes outliert. Így a végső komponensek együttese – azaz additív esetben az összegük, multiplikatív esetben a szorzatuk – megegyezik az eredeti, megfigyelt idősorral.

A TRAMO-SEATS módszernek különféle szoftveres implementációi vannak: az eredeti DOS alapú TRAMO-SEATS program, a Windows-alapú TRAMO-SEATS for Windows (TSW) és az Eurostat által kifejlesztetett, a szintén Windows-alapú, de az X12-ARIMA-hoz is kezelőfelületet adó Demetra.

A közeljövőben az X12-ARIMA és a TRAMO-SEATS módszerek (szoftver szintű) integrálódása várható, ezt mutatja, hogy az X12-ARIMA 1.3-as béta verziójában az idősor linearizálását és az ARIMA-modell kiválasztását az eddigi regARIMA eljárás helyett alapbeállítottként a TRAMO végzi, míg a regARIMA használata megmaradt lehetőségnek. Az erőfeszítések jelenleg arra irányulnak, hogy minél inkább egységesítsék a kiigazítás eredményét értékelő diagnosztikákat.

2.11. A TRAMO/SEATS diagnosztikái

A TRAMO/SEATS tesztjei azt mérik, hogy a megbecsült és aztán felhasznált ARIMA-modell mennyire illeszkedik az idősorra. Azt a hipotézist teszteljük, hogy a becsült reziduumok vajon tényleg normális fehérzajként viselkednek-e. Egyrészt azt vizsgáljuk meg, hogy van-e autokorreláció a reziduumokban. Ehhez ún. portmanteau típusú próbákat alkalmazunk, a Ljung-Box- és a Box-Pierce-féle teszteket a reziduumokra. A reziduumok linearitásának teszteléséhez ugyanezen próbákat a reziduumok négyzetére alkalmazzuk. Másrészt a normalitást ellenőrizzük a reziduumokra vonatkozó ferdeség és csúcsosság próbákkal és a kettő kombinálásával kapott normalitás teszttel. Az említett próbák a következőképpen néznek ki:

A Ljung-Box-próbastatisztika:

$$Q(h) = N(N+2)\sum_{j=1}^{h}\frac{\hat{\rho}^2(j)}{N-j},$$

ahol $\hat{\rho}$ a reziduumok becsült autokorreláció függvénye. Ha a reziduumok egy ARMA(p, q) folyamat reziduumai, akkor a $Q(h)$ eloszlása aszimptotikusan $\chi^2_{h-(p+q)}$. h értékét az idősor megfigyeléseinek gyakorisága alapján választják meg, havi idősoroknál h=24, negyedéves idősoroknál h=16 a használatos.

A szezonális autokorreláció vizsgálatára való a Box-Pierce-teszt[15]:

$$Q_s = N(N+2)\sum_{j=1}^{3}\frac{\hat{\rho}^2(js)}{N-js},$$

ahol s az évenkénti megfigyelések száma (azaz 12 a havi és 4 a negyedéves idősoroknál). A nullhipotézis melletti eloszlás közelítőleg χ^2_2.

A ferdeség tesztelésére:

$$m_3 = \frac{\sum \hat{a}_t^3}{N\hat{\sigma}_a^3},$$

[15] A Box-Pierce-féle próbán gyakran nem a fenti szezonális autokorrelációt vizsgáló tesztet értik, hanem a Ljung-Box próbánál gyengébb autokorreláció tesztet, amelynek tesztstatisztikája: $N\sum_{j=1}^{h}\hat{\rho}^2(j)$, és aszimptotikus eloszlása ugyanaz, mint a Ljung-Box- tesztnél.

ahol $\hat{a}_t$ a reziduumok, N a reziduumok száma, és $\hat{\sigma}_a$ a reziduumok becsült szórása. Ha normális eloszlású reziduumokról van szó (azaz teljesül a nullhipotézis), akkor m_3 eloszlása aszimptotikusan N(0, 6/N).

A csúcsosság tesztelésére:

$$m_4 = \frac{\sum \hat{a}_t^4}{N\hat{\sigma}_a^4}.$$

Normális eloszlású reziduumok esetén m_4 eloszlása aszimptotikusan N(3, 24/N).

A ferdeség és a csúcsosság tesztből kombinált normalitás teszt (Jarque-Bera-teszt):

$$M = \frac{N}{6}m_3^2 + \frac{N}{24}(m_4 - 3)^2.$$

A nullhipotézis teljesülése, azaz normális reziduumok esetén, az M eloszlása aszimptotikusan χ_2^2.

3. NEMZETKÖZI ELVÁRÁSOK, GYAKORLAT

A nemzetközi elvárásoknál elsősorban az Európai Unió illetve annak statisztikai szervezete, az Eurostat elvárásait, ajánlásait kell a KSH-nak figyelembe vennie. Az Eurostat feladata nem csak a statisztikai adatok begyűjtése, hanem azok összehasonlíthatóságának és minőségének biztosítása, így a szezonális kiigazítást illetően is olyan igényeket fogalmaz meg, amelyek ezeket biztosítják. Ezen kívül saját hatáskörben is végez szezonális kiigazítást, publikál szezonálisan kiigazított adatokat általában a teljes Európai Unióra vonatkozóan.

3.1. A szezonális kiigazítás szabályozása

A szezonális kiigazításra nem vonatkozik EU-jogszabály, így az EUROSTAT elvárásait ajánlásokként fogalmazza meg. Az ajánlások most következő leírásának forrása: Eurostat (1998b).

Minden kérdéskörnél az ajánlások után leírjuk az Európai Uniós tagországok gyakorlatát, aminek forrása az Európai Unió és az OECD közös felmérése, amely 2001 első félévében zajlott, és a 2001-ben publikált adatok szezonális kiigazítási módszertanát vizsgálja. (OECD (1998), Eurostat (2003))

A felmérést az Európai Unió Monetáris, Pénzügyi és Fizetésimérleg-statisztikák Bizottsága (Committee on Monetary, Financial and Balance of Payments Statistics, CMFB) kezdeményezte, céljuk a szezonális kiigazítási módszer harmonizálási lehetőségének vizsgálata volt. Az OECD javaslatára az uniós tagországok mellett az (akkori) csatlakozó országok és az OECD nem európai tagországainak gyakorlatát is vizsgálták. A vizsgálatban nem csak az adott ország statisztikai hivatalainak gyakorlatát, hanem a központi bankok gyakorlatát is felmérték. Hazánkat mind a Magyar Nemzeti Bank, mind a KSH képviselte. Ugyanakkor fontos hangsúlyozni, hogy a felmérés idején az MNB már a TRAMO/SEATS módszert alkalmazta, míg a KSH még az X11-ARIMA-t.

Kiigazításra használt módszer

Az első és legfontosabb kérdés a szezonális kiigazításra használt módszer kiválasztása. Az Eurostat jelenleg két módszert támogat: a TRAMO/SEATS-et mint teljesen modell alapú eljárást és az X12-ARIMA-t, amely számos funkcióval kibővült a korábbi, már nem támogatott X11-ARIMA-hoz képest.

Az EU-ban a statisztikai intézetek több mint 70%-a használ X11-családba tartozó szezonális kiigazító eljárást és közel 50%-uk használja a TRAMO/SEATS módszert. (Az intézmények több mint 40%-a egyszerre több módszert is alkalmaz.) A TRAMO/SEATS és az X12-ARIMA kombinációját 23%, a TRAMO/SEATS-et egyedül 19% használja. Az X12-ARIMA egyedüli módszerként nem használatos. Egyéb módszert (is) használ 8%.

A csatlakozó országok intézményei közül 72% használja az X11 családba tartozó módszer valamelyikét, 67% a TRAMO/SEATS módszert. A TRAMO/SEATS és az X12-ARIMA kombinációját 33%, a TRAMO/SEATS-et egyedül 17% használja.

Aggregátumok kezelése

Az aggregátumokat érintő direkt vagy indirekt kiigazítás közötti választást illetően az Eurostat azt az álláspontot képviseli, hogy a direkt módszert célszerű alkalmazni annak érdekében, hogy az aggregátumok szezonális kiigazításának magasabb legyen a minősége. Egyes idősoroknál a felhasználók részéről erős lehet az igény a konzisztenciára, így ebben az esetben a kiigazítás után az eltérést szét kell osztani az aggregálandó rész idősorok között azok nagyságának arányában. Ugyanígy az időbeli konzisztencia biztosítására (pl. negyedéves, illetve éves adatok között) is van lehetőség. A konzisztencia ára a részidősorok szezonális kiigazításának alacsonyabb minősége.

Az EU-országok intézeteinek 30%-a figyelembe veszi az aggregációs problémát, de nincs domináns módszer. A maradék 70%-nál a problémát tanulmányozzák és/vagy a jelenlegi szoftver nem támogatja.

Revíziós politika

A revíziós politikát illetően az Eurostat azt javasolja, hogy évenként egyszer történjen az ARIMA-modell és paramétereinek rögzítése, egy éven belül a modell ne változzon. A túl gyakori újrabecslés semmiképp sem javasolt, mivel a gyakori nagymértékű revízió megzavarhatja a felhasználókat.

A modell paramétereit fix gyakorisággal (általában évente) frissítik az EU-országok intézeteinek 60%-ában, míg magát a fix szűrőt vagy az ARIMA-modellt az intézetek 70%-a frissíti évente. Ezen kívül az intézmények 38%-a az alapadatokban bekövetkezett változás esetén is frissíti a paramétereket és az egyéb beállításokat.

A csatlakozó országok esetében az intézmények közel fele frissíti a modell-beállításokat fix gyakorisággal, és ugyanennyien jelezték azt is, hogy az alapadat-revízió esetén is frissítenek.

Munkanaphatás kezelése

Az Eurostat nagy hangsúlyt fektet arra, hogy a munkanapkorrekciónak regressziós módszerrel kell történnie, a korábban elterjedt arányosítás nem használható, mivel az túl igazításhoz vezethet. Elvárás továbbá, hogy a munkanappal kiigazítás és a szezonális kiigazítás módszere harmonizáljon. Ha a szezonális kiigazítás első lépése a munkanappal kiigazítás, akkor a harmonizáció biztosított.

2004-től további igénye az Eurostatnak, hogy a munkanappal kiigazított adatsoroknál a tagállamok biztosítsák, hogy a bázisévben az indexek átlaga 100% legyen. Ezzel elérhető, hogy a tagállamok adataiból képzett aggregátumok esetén konzisztens, összehasonlítható eredményt kapjanak. Amennyiben a statisztikai hivatalok ezt nem teljesítik, akkor az Eurostat saját hatáskörben elvégzi ezt a korrekciót, aminek eredményeképp a nemzeti hivatal által közölt adat különbözhet az Eurostat által közölt adattól.

Demetra

A két támogatott módszerrel, az X12-ARIMA-val és a TRAMO/SEATS módszerrel kapott eredmények összehasonlíthatóságának érdekében az EUROSTAT kifejlesztetett egy felhasználóbarát szoftvert, a Demetra-t, amelyben a két módszer ugyanabban a programkörnyezetben használható.

Amikor arról beszélünk, hogy a KSH-ban a TRAMO/SEATS módszert használjuk, akkor ezen a Demetra szoftver TRAMO/SEATS részének alkalmazását értjük.

3.2. Szezonális kiigazítás az Eurostatnál

Az Eurostat – a tagországok számára általánosan megfogalmazott ajánlások után – kialakította az STS (Short-Term Business Statistics) mutatókra vonatkozó egységes szezonális kiigazítási gyakorlatát is, amelyet 2004-ben az STS munkacsoport ülésen fogadtak el. (Eurostat (2004a.), (2004b.), (2004c.))

Szezonális kiigazítás és munkanappal kiigazítás alapelvei

Az STS-mutatók igazításánál is elvárás a regressziós módszer alkalmazása, az összhang biztosítása a szezonálisan kiigazított és a munkanaphatástól megtisztított adatsorok között, és a nemzeti ünnepek figyelembe vétele. Az Eurostat részére megküldött adatok mellett, a metaadatok között a tagországoknak szerepeltetni kell az igazításhoz használt módszer megnevezését, a kiigazításhoz használt regresszorokat, és az alkalmazott aggregálási eljárást. A modellt és beállításait (pl. transzformáció) – lehetőség szerint – legalább egy évig fixen kell tartani, ugyanakkor a paramétereket legalább évente, illetve az alapadatokban bekövetkezett nagyobb változások esetén (pl.: bázisváltás, módszertani váltás) frissíteni kell.

Az európai aggregátumok szezonális kiigazítását direkt módszerrel képzik az Eurostatnál, azaz a tagállamok alapadatait, illetve néhány mutatószám esetén[16] a munkanappal kiigazított adatait adják össze, és ezt igazítják szezonálisan.

Abban az esetben, ha egy tagállam az Eurostatnak nem az előírtaknak megfelelően szolgáltat adatot, azaz az alapadat mellett nem küldi ki a szezonálisan kiigazított adatsorokat, a trendet, illetve szükség esetén a

[16] Például: ipari termelés, építőipar, kiskereskedelmi forgalom

munkanappal kiigazított adatokat, az Eurostat – előre meghatározott elveknek megfelelően – maga igazítja ki az adatsort. Ilyenkor a nemzeti statisztikai hivatal közölt adatai eltérhetnek az Eurostat által publikált adatoktól. Természetesen az Eurostat publikációiban közli, hogy ki végezte az igazítást.

Az Eurostat kiigazítási eljárása

Az előzőekben leírtak alapján az Eurostat nem csak az európai aggregátumok képzése esetén, hanem egyes tagországok adatai alapján is végez szezonális kiigazítást. A paramétereket lehetőség szerint évente egyszer rögzítik. Természetesen nagyobb revízió esetén, illetve az európai aggregátumok képzési módjának megváltozása esetén a modellt újrabecslik.

A kiigazításhoz a Demetra 2.04 szoftvert, azon belül a TRAMO/SEATS módszert alkalmazzák. Az automatikus futtatás előtt néhány paramétert (ARIMA-modell típusa, transzformáció, munkanapregressziók és további regressziók száma és típusa, outlierek) előre rögzítenek, a beállítások sokszor az idősor hosszától függnek.

A **munkanappal kiigazított adatsor** képzésekor nemzeti szinten – amennyiben a kiigazítást a tagállam nem végzi el – a tagállamok alapadataiból indulnak ki, az európai aggregátum esetén a már munkanappal kiigazított nemzeti adatokból.

Hosszú idősornak tekintik azokat az idősorokat, amelyek 1995 előtt indulnak, azaz 10 évnél hosszabbak. Ezeknél 7 regresszoros munkanaphatást és húsvéthatást tesztelnek, lehetőséget adva arra, hogy a regresszorok számát a program automatikusan csökkentse. A rövid idősorok esetén az alapbeállítás: 2 regresszoros munkanaphatás és húsvéthatás tesztelése.

Bár ajánlásában az Eurostat kéri a kiigazítás során a nemzeti ünnepek hatásának figyelembe vételét, mégis mikor saját hatáskörben igazítja a nemzeti idősorokat, ezt nem veszi figyelembe. Ennek oka, hogy a Demetrába beépített nemzeti ünnepek ellenőrzésre szorulnak az időközben bekövetkezett változások miatt. Várhatóan a Demetra új verziójában ez ki lesz javítva.

Az outlierek 3 típusát különböztetik meg a kiigazítás során: additív outlier, csillapodó jellegű törés, szinteltolódás. Az outlierek becslése automatikusan történik. Az ARIMA-modell típusát azonban nem automatikusan határozzák meg, hanem a (011)(011) modellt veszik alapbeállításnak, és logaritmikus transzformációt alkalmaznak. Ha ez a modell nem illeszkedik, akkor más modellt keresnek (erre vonatkozóan nincs előírás).

Ha a tagállam a saját aggregált idősorából nem szűri ki a munkanaphatást, akkor az Eurostat indirekt igazítási módszert választ. Ugyanezt alkalmazza a munkanappal kiigazított európai aggregátum képzése esetén is.

A **szezonális kiigazítást** a munkanappal kiigazított adatsorból kiindulva végzi, így a munkanaphatást már nem teszteli. Az outliereket a munkanappal kiigazított adatsorhoz hasonlóan teszteli és ebben az esetben is az Airline-modellt (011)(011) alkalmazza. A szezonálisan kiigazított európai aggregátumot direkt módszerrel képzi, ennek alapját az indirekt módszerrel létrejött munkanappal kiigazított európai aggregátum adja. A szezonális és munkanappal kiigazított adatok további konzisztenciáját azzal biztosítja, hogy a fontosabb és tartalmilag összefüggő adatsorokra (például az ipari termelés teljes nemzetgazdaságra vonatkozó adatai, és a legnagyobb súlyú alágazataira) ugyanazokat a beállításokat alkalmazza: ugyanaz az ARIMA-modell, transzformáció, regresszorok száma stb.

A fentieket összevetve az 5. fejezetben leírt KSH gyakorlattal, látszik, hogy a KSH jelenlegi gyakorlata szinte teljes mértékben megfelel az Eurostat elvárásainak, illetve gyakorlatának.

4. Módszertani váltás a KSH-ban

2002 előtt a KSH-ban csak néhány területen végeztek szezonális kiigazítást, többnyire az X11 és X11-ARIMA módszerrel. Általában az év utolsó adatával, automatikus futtatással elvégezték a szezonális kiigazítást, majd a program által a következő időszakra előrejelzett szezonális tényezők alapján határozták meg a szezonálisan kiigazított idősor értékeit. A munkanapok és a húsvét hatását általában nem vették figyelembe, ahol azonban a munkanaphatást is számszerűsítették, ott ezt havonta határozták meg az előre becsült szezonális tényező leválasztása után. A trendet havonta újrabecsülték a program segítségével. Sok területen alkalmazták a szezonális kiigazítás mellett vagy helyett az eredeti idősorból képzett „időszak / előző év azonos időszaka” mutatót is.

A Hivatalon belül a szezonálisan kiigazított adatok revíziója tekintetében sem volt egységes a gyakorlat. A legtöbb területen visszamenőleges korrekció nem történt, csak ha jelentősen módosult a szezonálisan kiigazított érték. Néhány főosztályon azonban az év utolsó adatával kapott kiigazítási eredmények alapján visszamenőleg is módosították az adatokat, de csak egy évre vonatkozóan.

A 1990-es években nemzetközi szinten egyre inkább elterjedtek az újabb szezonális kiigazító módszerek, mint az X12-ARIMA vagy a TRAMO/SEATS. Ehhez az elméleti és gyakorlati fejlődéshez kapcsolódóan 2000–2001 táján a KSH-ban is megnőtt az igény egy új, egységesen használható módszer és gyakorlat bevezetésére, és jóval több szezonálisan kiigazított adat publikálására. A Hivatalra több oldalról nehezedett nyomás: egyrészt a negyedéves GDP kiigazítását is el kellett kezdeni, másrészt a hazai felhasználók, különösen a Magyar Nemzeti Bank is kritikát fogalmazott meg a KSH szezonális igazításával kapcsolatosan, illetve az Európai Unió harmonizációs igénye, amely a tag és tagjelölt országok szezonális kiigazításának egységesítését támogatta, is indokolta egy új módszer bevezetését.

Az Eurostat már jóval a csatlakozás előtt igényelte a leendő tagországok adatait is, nem csak alapadat formájában, hanem munkanappal kiigazítva, illetve szezonálisan és munkanappal kiigazítva is. Annak érdekében, hogy az egyes országok adatai összehasonlíthatóak, és aggregátumok képzésére alkalmasak legyenek, egységes elvárásokat fogalmazott meg a jelenlegi és leendő tagországok számára (Eurostat, 1998b). Két kiigazítási módszert ajánlott: az X12-ARIMA és a TRAMO/SEATS módszert. A könnyebb használhatóság kedvéért kifejlesztetett egy kezelőfelületet (Demetra), amely mindenki számára ingyenesen hozzáférhető és mindkét módszert tartalmazza.

2001 második felében a módszertani osztály munkatársaiból és az egyes szakstatisztikák szakértőiből megalakult *„A szezonális kiigazítás harmonizációja"* munkacsoport. A csoport összetételének kialakításánál alapvető szempont volt, hogy minden olyan főosztály részt vegyen a munkában, ahol évesnél nagyobb gyakoriságú adatgyűjtések feldolgozása folyik. Így az életszínvonal-statisztika, az iparstatisztika, a külkereskedelem-statisztika, a mezőgazdaság-statisztika, a nemzeti számlák, a pénzügystatisztika és a szolgáltatásstatisztika szakértői kapcsolódtak be a megbeszélésekbe.

További szempont volt, hogy majd a kialakításra kerülő gyakorlatot szakstatisztikától független szakértők koordinálják, akik minden szakterület szempontját egyformán figyelembe tudják venni, hiszen a szezonális kiigazítás módszertana független a szakstatisztikáktól (bár a kiigazítás végeredményét nagyban befolyásolja a szakstatisztikusok véleménye, például az outlierek magyarázhatóságáról, a munkanap- és húsvéthatás meglétéről), speciális matematikai statisztikai ismereteket igényel, és az Eurostat előírásai a fő irányelvek tekintetében minden területen azonosak. Ezért lehetővé vált, hogy ne valamelyik szakfőosztály koordinálja a szezonális kiigazítást, hanem erre specializálódott munkatársak végezzék el ezt a feladatot. Így erre a szerepre – a nemzetközi gyakorlathoz hasonlóan – a legalkalmasabbnak a 2001-ben létrejött Statisztikai mintavételi és módszertani osztály[17] munkatársai tűntek.

A munkacsoport munkája során egy elemzésekre épülő, megalapozott javaslatot kívánt tenni a szezonális kiigazítás új rendszerére vonatkozóan, amely mind a Hivatal, mind a partnerintézmények és a felhasználók számára elfogadható. A vizsgálat során a munkacsoport célja egy olyan egységes kiigazítási politika kialakítása volt, amely minden olyan idősorra, amelyet a KSH-ban ki kell igazítani, megfelelő minőséggel alkalmazható, ugyanakkor 2002 elejétől bevezetésre kerülhet. Az egységes szezonális kiigazítási

[17] Az osztály neve megalakulásakor Statisztikai mintavételi és módszertani osztály volt, de 2004-től Mintavételi és módszertani osztályra módosult.

politika az egységes szezonális kiigazítási módszertan alkalmazását, az eredmények elfogadásának egységes kritériumrendszerét, az évközi kiigazítási eljárást, a szezonális kiigazításra vonatkozó publikációs és revíziós rendszert foglalja magába. A fő cél közös alapelvek, egy keretrendszer meghatározása, amelyet a KSH a külső felhasználók felé kommunikálhat, és amellyel összhangban az egyes főosztályok kialakíthatják saját revíziós politikájukat és publikációs gyakorlatukat. Ugyanakkor nem célja, hogy az egyes főosztályokon ugyanazokat a mutatókat és grafikonokat publikálják.

Az egységesítésnél további szempont volt az Eurostat ajánlásaival való összhang is.

Az egységes gyakorlat kialakításának alapvető feltétele volt, hogy minden szakterületen egyformán alkalmazható legyen. Ezért a munkacsoport első feladat az volt, hogy megállapítsa létezik-e ilyen eljárás. Ehhez először felmérték, hogy az egyes szakfőosztályokon alkalmaznak-e szezonális kiigazítást, milyen szoftverrel, milyen módszerrel. Milyen revíziós politikát követnek, milyen formában publikálják az eredményeket? Van-e az Eurostatnak az adott területre vonatkozóan valamilyen speciális előírása, ajánlása? Milyen egyéb felhasználói igényeket ismernek?

A korábbi gyakorlat feltérképezése mellett az Eurostattal való összhang érdekében a módszertani osztály munkatársai megvizsgálták az Eurostat által ajánlott két módszert: az X12-ARIMA és a TRAMO/SEATS módszert.

A 176 idősoron a Demetra segítségével elvégzett vizsgálatok, tesztelések során a munkacsoport 11 szempont alapján értékelte a módszereket:[18]

Értékelési szempont	TRAMO/SEATS	X12-ARIMA
Nemzetközi ajánlásoknak megfelelő módszer	+	+
Tudományosan elfogadott, korszerű módszer	++	+
Rövid idősorok kezelése	-	--
Magyar ünnepnapok kezelése	+	-
Stabil eredmények az idősor végén	++	+
Szezonalitás szűrésének hatásossága	++	+
Eredmények statisztikai diagnosztikája	+	++
Automatikus futtatás lehetősége	+	+
Nagyszámú idősor kezelése	++	+
Könnyen kezelhető input és output fájlok	+	+
Felhasználóbarát kezelőfelület	+	+

A Demetra környezetéből adódóan az automatikus futtatás lehetősége, a felhasználóbarát kezelőfelület és a könnyen kezelhető input és output fájlok tekintetében nem volt különbség a két módszer között. Szintén mindkét módszer megfelel a nemzetközi ajánlásoknak, hiszen az Eurostat ajánlásának megfelelően választotta ki a munkacsoport ezt a két szóba jöhető módszert.

A TRAMO/SEATS módszertan elmélete az idősorelemzés korszerűbb, sztochasztikus módszerein alapszik.

---

[18] A „+” és „-” jelek a két módszer közötti különbséget tükrözik egy szempontra vonatkozóan, de nem alkalmasak egy módszer különféle szempontok szerinti összehasonlítására, pl. a táblázatból nem következik, hogy a TRAMO/SEATS tudományosan jobban elfogadott, mint amennyire nemzetközileg ajánlott.

A rövid idősorok kezelése mindkét módszer esetében problematikus kérdés. Sajnos a hazai körülmények között, különösen a negyedéves bontású adatsorok esetében gyakran előfordult, hogy csak igen rövid idősorok álltak rendelkezésre a rendszerváltozás, illetve az azt követő, számos területen bekövetkezett gyökeres gazdasági-, illetve módszertani változások következtében. A rövid idősorok problémája azonban az ad hoc filterek sajátosságainak következtében érzékenyebben érinti az X12-ARIMA eljárást.

A TRAMO/SEATS módszertan továbbá fel van készítve a magyar ünnepnapok kezelésére is, valamint stabilabb eredményeket szolgáltat, mivel az újabb adatok megjelenése a becslés során mérsékeltebb változást okoz az idősor végén, mint az X12-nél használt ad hoc filterek esetén.

A módszer vizsgálata mellett sor került a szezonálisan kiigazítandó idősorok matematikai statisztikai szempontból történő vizsgálatára, szakmailag adekvát modellezésére, a kiugró értékek, trendváltozások, esetleges strukturális törések, egyéb befolyásoló tényezők (pl. munkanapok, ünnepnapok) hatásának kimutatására, szignifikanciájuk vizsgálatára, a főosztályon való megvitatásra, illetve szakstatisztikai, valamint közgazdasági szempontból történő magyarázatára.

A kísérleti elemzések során a módszertani osztály munkatársai mintegy 176 idősort vizsgáltak meg, illetve igazítottak ki szezonálisan. Minden idősorra elvégezték a Demetra szoftver alapértelmezett beállításokkal történő futtatását az automatikus modulban, mind a TRAMO/SEATS, mind az X12-ARIMA módszer kiválasztása mellett. Azon idősoroknál, ahol ez nem adott megfelelő eredményt, részletes analízist is végeztek.

A futtatások tapasztalatai alapján a TRAMO/SEATS módszertan mellett szólt az is, hogy a 176 futtatás során csak 22 alkalommal nem tudott adekvát modellt illeszteni a vizsgált idősorra a diagnosztikai eredmények kiértékelése után. Ez az esetek csupán 12,5%-a. Az X12-ARIMA ugyanakkor 60 esetben vetette el az általa illesztett modellt a kapott diagnosztikák alapján. Ez már valamivel több, mint 34%-a az összes idősornak. Mivel a KSH-ban a szezonálisan kiigazított idősorok számának folyamatos növekedésére lehetett számítani, és a rendszeres havi munkánál fontos szempont, hogy a futtatás a lehető legkevesebb időt vegye igénybe, így az automatizáltság fontos szerepet játszott a döntés során.

A főosztályoktól összegyűjtött korábbi gyakorlatra vonatkozó anyagokat és a kísérleti számítások eredményeit a Statisztikai mintavételi és módszertani osztály munkatársai összesítették, és ezek alapján a úgy ítélték meg, hogy a hazai körülmények között a TRAMO/SEATS jól, illetve jobban alkalmazható, mint az X12-ARIMA, így a Statisztikai mintavételi és módszertani osztály a munkacsoport következő ülésén a TRAMO/SEATS egységes alkalmazását javasolta a Hivatalon belül. A program matematikai-statisztikai háttere jobb eredményeket tesz lehetővé, mint az alternatív program, az X12-ARIMA, azáltal, hogy a mozgóátlagolással végzett becslések csak kellően hosszú idősorok esetén képesek jó hatásfokkal kiszűrni a véletlen ingadozások hatását. Így a Hivatal által elemzett, többnyire meglehetősen rövid idősorok esetén az X12-ARIMA a tartós tendenciát és a szezonálisan kiigazított értékeket is csak nagyobb ingadozással, illetve nagyobb bizonytalansággal képes becsülni. Továbbá a TRAMO/SEATS által generált trend és a szezonálisan kiigazított idősor vége sokkal kevésbé érzékeny egy újabb adat megjelenésére, ezért az időszak végére lényegesen stabilabb eredményeket produkált.

A munkacsoportban a TRAMO/SEATS programmal szemben legtöbbször megfogalmazott kérdés az volt, hogy az átállásnál a program eredményei mennyire fognak eltérni az eddig publikáltaktól. A csoport résztvevői jelezték, hogy ezt szeretnék részletesen elemezni a főosztályokon, mielőtt elköteleznék magukat a javasolt módszer alkalmazása mellett.

Ugyanakkor az aggregátumok kezeléséről sikerült megállapodni. Mivel a KSH által igazított idősorokból mindig az aggregátum bír a legnagyobb jelentőséggel, ezért az aggregátumokra a direkt kiigazítás tűnt a legmegfelelőbbnek, azzal a kiegészítéssel, hogy a direkt és az indirekt módszer által szolgáltatott értékek közötti differenciákat ajánlott szétosztani az alágazatok nagyságrendjének arányában a részek szezonálisan kiigazított értékei között.[19]

A belső egyeztetések után 2002 I. negyedévében külső szakértők (Pénzügyminisztérium, Magyar Nemzeti Bank) részvételével szakmai fórumon kerültek bemutatásra és megvitatásra az eredmények.

A belső és külső egyeztetések után elkészült a döntési javaslat a KSH vezetősége részére, amely tartalmazta a munka főbb eredményeit, ennek alapján javaslatot tett az alkalmazandó módszerre, rögzítette a további munka kereteit, a Hivatalon belüli munkamegosztást. Nemzetközileg

[19] A próbaév tapasztalatai alapján később ezt úgy módosítottuk, hogy lehetőség szerint nem osztjuk szét a differenciát. Részletesebben lásd a gyakorlatnál.

elfogadott és szakmailag is megalapozott az olyan típusú munkamegosztás, ahol az adatsorok szakmai felelősei végzik el a szezonális kiigazítást, a szezonális kiigazítás elméleti és módszertani felelőseivel évente egyszer egyeztetett modell alapján, az előzetesen elfogadott irányelveknek megfelelően, a felhasználók részletes tájékoztatása mellett.

Így 2002-től kezdődően új, egységes módszertant vezetett be a KSH, amely összhangban van a hazai- és az Eurostat-elvárásokkal, és teljesíti a kiigazítással szemben támasztott követelményeket. A módszertan alkalmazásához a Demetra szoftver került kiválasztásra.

Az új gyakorlatot a 2002. február 5-i Gazdaságstatisztikai Felhasználói Fórum elfogadta, majd a módszertani váltást a 2002. február 18-i Elnöki Értekezlet jóváhagyta. Ezután a KSH sajtóközleményben (2002. március 14.) tájékoztatta a közvéleményt a Hivatal szezonális kiigazítási módszertanában történt változásról. (A kialakított gyakorlat részletes leírását a következő fejezet tartalmazza.)

2002 áprilisában az idősorok elméletéről (AR-, MA-, ARIMA-folyamatokról, kapcsolódó becslésekről, statisztikai próbákról), a szezonális kiigazítás módszereiről, a TRAMO/SEATS módszerrel történő kiigazítás elméleti és gyakorlati kérdéseiről és a szoftverhasználatról – a majdani alkalmazók és minden érdeklődő munkatárs részére – a Statisztikai mintavételi és módszertani osztály munkatársai 3 részből álló tanfolyamot tartottak, annak érdekében, hogy a szezonálisan kiigazítás a szakfőosztályokkal egyeztetett módon, koordináltan kerüljön elvégzésre. Ez a tanfolyam kiegészítette a korábbi oktatásokat, például a Kanadai Statisztikai Hivatal szakértői által tartott tanfolyamot.

A 2002-ben és 2003-ban a későbbiekben leírt munkamegosztással, TRAMO/SEATS módszerrel történt a kiigazítás. 2002 novemberében a szakfőosztályokon elkészültek a kiigazításra kerülő fontosabb idősorok dokumentációi is. Az egyes dokumentumok tartalmazzák az idősorok főbb jellemzőit (hossz, bázisév, gyakoriság, publikálás), illetve az aktuális kiigazítás eredményeit (outlierek, munkanap- és húsvéthatás), és lehetőleg ezek gazdasági értelmezését.

A 2002-es „próbaév” tapasztalatai alapján 2003 szeptemberében belső szakértői értekezlet keretében a szakfőosztályok képviselői és a Statisztikai mintavételi és módszertani osztály munkatársai megtárgyalták a felmerült problémákat és azok megoldásait. Vezetői Kollégiumi Előterjesztési

Javaslatban rögzítették az alapelveket és a kialakult gyakorlatot, amelynek alapján a jövőben is szeretnék folytatni a szezonális kiigazítást.

A Statisztikai Szemle hasábjain került sor a KSH által alkalmazott szezonális kiigazítással, a szezonálisan kiigazított adatok publikálásával kapcsolatos néhány kérdés megvitatására (Friss, 2003, illetve Bauer és Földesi, 2003).

2003. december 9-én az MTA Statisztikai Bizottságának Módszertani albizottsága is napirendre tűzte a szezonális kiigazítás témáját, ahol az elméleti előadás után a résztvevők megismerkedtek az MNB és a KSH szezonális kiigazítási gyakorlatával.[20] A vita során a KSH gyakorlatával szemben nem merült fel kritika.

Szintén decemberben (11-én) a Gazdaságstatisztikai Felhasználói Fórumon is bemutatásra került a KSH gyakorlata.

Végül 2004 áprilisában lépett életbe a szezonális kiigazítás egységes gyakorlatáról szóló szabályzat, amely a KSH-ban kiigazításra kerülő összes idősorra vonatkozik.

[20] Az elméleti bevezetőt Sugár András (a BKÁE adjunktusa), a korreferátumokat az MNB részéről Montvai Beáta, a KSH részéről Bauer Péter tartotta.

5. A KSH SZEZONÁLIS KIIGAZÍTÁSI GYAKORLATA 2002-TŐL

2002 elejétől tehát a szezonális kiigazításhoz a TRAMO/SEATS módszert alkalmazzuk, és a módszert a Demetra programmal használjuk. Mivel a program futtatásánál, a módszer alkalmazásánál számos lehetőség adódik, így a kialakított gyakorlatról szóló szabályzatnak nem csak a módszert kell rögzítenie, hanem a paraméterrögzítés, a revízió, a munkamegosztás kereteit is.

A TRAMO/SEATS havi, negyedéves vagy féléves gyakoriságú idősorokat tud kezelni, de a KSH-ban jelenleg csak havi és negyedéves idősorokat igazítunk szezonálisan.

A KSH-ban általában bázisévhez viszonyított indexeket igazítunk – jelenleg a legtöbb idősornál a 2000. év adatainak átlaga a bázis –, de előfordul abszolút számokból álló idősor és adott év adott hónapjához viszonyított indexek igazítása is (ez utóbbira példa a fogyasztóiár-index idősora, amelynél 1994. december = 100%). Nem igazítunk „előző év azonos időszaka = 100%" típusú indexeket vagy „bázisév azonos időszaka = 100%" típusú indexeket, mert ezeknél az eredeti idősorhoz képest információt vesztünk és megváltozik az idősor dinamikája is. Korábban ezeket a mutatókat a szezonális hatás egyszerű kiszűrésére használták, de ezekkel például nincs lehetőség a mozgó szezonalitás kezelésére, ezért kevesebb információt nyújtanak, mint a szezonálisan kiigazított idősor.

A szezonális kiigazításhoz szükséges, hogy az igazítandó idősor havi gyakoriság esetén legalább 36 megfigyelésből álljon, azaz legalább 3 év hosszú legyen, negyedéves gyakoriság esetén pedig legalább 16 megfigyelés, azaz 4 éves hosszúság szükséges. Ezek természetesen a minimum értékek ahhoz, hogy a program diagnosztikákat számoljon, az idősorra illesztett ARIMA-modelltől függően ennél hosszabb idősorokra lehet szükség. Rövid idősorok szezonális kiigazítása nem ad megbízható, jó minőségű eredményt, ezért törekedni kell arra, hogy minél hosszabb idősorok álljanak rendelkezésre. Jelenleg a KSH-ban a havi és negyedéves szezonálisan kiigazított idősorok többsége 1998-tól áll rendelkezésre, így a minimális követelményének ugyan eleget tesznek, de relatív rövidségük a kiigazítás eredményeinek stabilitását jelentősen befolyásolja.

Önmagában azonban a hosszú idősor sem garantálja a jó eredményt: az idősornak homogénnek is kell lennie, azaz olyannak, melynek előállítása időben változatlan módszertan szerint történik, és hasonló viselkedésű az idősor teljes hosszában. Kisebb változások megengedhetőek, például a szezonális tényező lassú változását, azaz a mozgó szezonalitást valamint a kiugró értékként kezelhető változásokat a módszer figyelembe veszi. Ha azonban az idősor viselkedése, elsősorban a szezonalitás lényegesen megváltozik és egységes modell már nem illeszthető rá teljes hosszúságában, akkor az idősort csonkítani kell. Nyilvánvaló, hogy ez a lehetőség elsősorban hosszú idősorok esetében adott, a KSH-ban jelenleg igazított idősorok még csonkítás nélkül is rövidnek tekinthetők.

Szezonálisan kiigazított adatokat és trendet a statisztikai hivatalok, így a KSH is ugyanazon idősorra ismételten közölnek, ahogy az idősor kiegészül az újabb és újabb időszakok megfigyeléseivel. Mivel a több megfigyelésből álló idősor több információt hordoz a sor komponensekre bontásához, emiatt a szezonálisan kiigazított adatok és a trend is visszamenőlegesen, az idősor egész hosszára nézve változik a korábbi eredményekhez képest, azaz revízióra kerül sor. Ennek a változásnak a mérséklésére, illetve ütemezésére van lehetőség, mégpedig az idősorra illesztett ARIMA-modell és paramétereinek korlátozott újrabecslésével. A következő lehetőségek adottak:

– Az ARIMA-modell és paramétereinek újrabecslése minden egyes alkalommal, amikor új időszakra vonatkozó megfigyeléssel egészül ki az idősor. Ez a megközelítés teljes mértékben felhasználja az összes rendelkezésre álló információt, és így a lehető legjobb eredményt adja, de nagy mértékű és minden egyes publikációnál jelentkező revízióval jár.

– Az ARIMA-modell rögzítése, de a paraméterek újrabecslése minden egyes új megfigyelésnél. Az ARIMA-modell frissítésére évente egyszer kerül sor. Ilyenkor a rendelkezésre álló információkat korlátozottan használjuk fel, de még így is viszonylag jó eredményt kapunk. A revízió kisebb mértékű, mint ha a modellt is újrabecsülnénk minden új megfigyelésnél. Nagyobb revízió csak az évenkénti modell-frissítésnél következik be.

– Az ARIMA-modell és a paraméterek rögzítése, újrabecslésük csak évente egyszer. Ilyenkor még korlátozottabb a rendelkezésre álló információk felhasználása, de még ez is elég jó eredményt ad. Az év közbeni revízió viszonylag kis mértékű, nagyobb revízió következik be az évenkénti modell- és paraméterfrissítésnél.

– A szezonális kiigazító módszer évenkénti egyszeri használata, amikor a következő évi szezonális tényezőt előre jelezzük, a már ismert időszakokra pedig újrabecsüljük. Az év során, az új megfigyelések adatainak beérkezésekor az előre jelzett szezonális tényezővel számolva állítjuk elő a szezonálisan kiigazított adatokat. Év közben egyáltalán nem használjuk fel az új megfigyelésekből a szezonalitásra vonatkozó információt. Revízió csak a szezonális kiigazító módszer évi egyszeri használatakor jelentkezik, de akkor nagymértékben.

A revízió mérséklésekor két egymásnak ellentmondó szempontot is figyelembe kell venni. Egyrészt ha az adatokat egy szűk szakértői réteg használja, akkor az eredmények frissessége a legfontosabb, tehát az ő szempontjaiknak a paraméterek folyamatos újrabecslése felel meg legjobban. Másrészt ha az eredmények széles körben kerülnek publikálásra, akkor a nem szakértő felhasználók szemében az adatok stabilitása lesz a legfontosabb, ezért az ő igényeiknek a szezonális tényezők előrejelzése lenne a megfelelő, mert igy az adatok visszamenőleg csak egyszer változnának.

A nemzetközi gyakorlatban mindegyik módszer alkalmazására találhatunk példát. A KSH számára fontos a lehetőleg kis mértékű revízió, de az eredmények jó minősége is, ezért a harmadik lehetőséget, tehát az évi egyszeri modell- és paraméterfrissítést választottuk. Ekkor – bár korlátozottan, de – figyelembe vesszük az új adatból származó új információt, ugyanakkor a visszamenőleges revízió kisebb, mint a folyamatos újrabecslésnél. Így tehát a modell és a paraméterek rögzítése rendszerint az év utolsó adatainak beérkezésekor történik, az új modell és paraméterek szerinti kiigazítás eredménye pedig a következő év első adatával jelenik meg.

A modellnek és a paramétereknek a tervezettől eltérő, év közbeni újrarögzítésére akkor kerülhet sor, ha alapadatokat érintő revízió történik azokra a megfigyelésekre, amelyeket felhasználtunk a korábbi modell- és paraméterbecslésnél, hiszen így az igazításhoz korábban használt modell már nem illeszkedik megfelelően az új adatsorra. Újrarögzítés történhet akkor is, ha a kiigazító program diagnosztikái alapján az újabb időszakokra vonatkozó megfigyelésekkel kiegészült idősor lényegesen más viselkedést mutat, mint ami a rögzített modellnek és paramétereknek megfelel. Erre a program által adott diagnosztikákból, vagy az outlierek magas számából következtethetünk. A revízióról és az aktuális beállítások megváltozásáról a KSH a kiadványok módszertani részében tájékoztatja a felhasználókat.

A szezonális kiigazítás során felmerülő egyik probléma a munkanaphatás kezelése.

A munkanaphatás kezelésekor figyelembe vesszük a magyar ünnepnapokat, ami a Demetrában beépített lehetőség a TRAMO/SEATS módszer használata esetén. A húsvéthatás vizsgálatakor az alapbeállítás szerinti 6 napos hatással számolunk.

A TRAMO képes automatikusan tesztelni, hogy az idősornál jelentkezik-e munkanap- vagy húsvéthatás, azonban célszerű figyelembe venni a rendelkezésre álló szakértői információkat is a hatások meglétéről, vagy arról, ha az adott idősorban nem értelmezhető a munkanapok eltérő számának hatása vagy a húsvéthatás. Ilyenkor az automatikus teszteléstől eltekintük, vagy az eredményét felülvizsgáljuk, és az információnak megfelelő beállításokat használunk. A szakértői információk különösen fontosak a rövid idősorok esetében, ahol a statisztikai tesztek nem elég megbízhatóak.

A paraméterrögzítés során azt is figyelembe vesszük, hogy rövid idősor esetén nem célszerű a hét minden munkanapját külön-külön figyelembe venni, azaz sok (6 vagy 7) regresszort használni, mert ekkor fennáll a veszélye, hogy az illesztett modelltől való eltérést minimalizálva az idősort túlzottan megtisztítjuk a munkanaphatástól, azaz túligazítjuk, és az idősor lényegi viselkedése elvész.

Arra is figyelni kell, hogy a regressziós együtthatók előjele értelmezhető-e, azaz például ha az idősornál feltételezhető, hogy a több munkanap növeli az időszak értékét, akkor a munkanapok regressziós együtthatója pozitív előjelű legyen.

A kiugró értékeket is képes automatikusan kezelni a TRAMO: mind ezek helyét, mind típusát automatikusan detektálja, majd minden egyes talált kiugró értéknek egy-egy regressziós változót feleltet meg.[21] Három fajta kiugró értéket veszünk figyelembe: additív kiugró értéket, csillapodó jellegű törést és szinteltolódást.

---

[21] Itt lényegesen egyszerűsítettük az eljárás menetét, a pontos leírás megtalálható a Tramo-ról szóló részben.

A kiigazítás során figyelembe vesszük az esetleges szakértői információkat arról, hogy egy adott időszakban lehetett-e kiugró érték, illetve a TRAMO által automatikusan talált kiugró értékeknek mi lehet a közgazdasági, társadalmi, időjárási vagy egyéb külső tényezőkben keresendő magyarázata. Különösen fontos a szakértői információ az idősorok végén megjelenő kiugró értékeknél, mert ezek típusa matematikai szempontból sokáig bizonytalan lehet, a típus későbbi módosulása pedig nagymértékű revízióhoz vezethet. Például ha az új outlier additív outlier vagy csillapodó jellegű törés, akkor a hatás csak a szezonálisan kiigazított idősorban jelenik meg. Ugyanakkor, ha ez az outlier szinteltolódás, akkor a hatás a szezonálisan kiigazított idősoron kívül a trendben is megjelenik.

Mint a munkanaphatásnál, itt is elmondható, hogy a túl sok regresszor használatát lehetőleg kerülni kell, mert megfelelően sok outlier alkalmazásával tetszőleges modell kiválóan illeszkedhet, emiatt az egy idősorhoz rendelt kiugró értékek számát a Demetra alapbeállítása szerint 5%-ban maximáljuk az idősor összes megfigyeléseinek számához képest. Amennyiben mégis több kiugró érték kezelésére van igény, amire rövid, különösen negyedéves idősoroknál lehet szükség, az 5%-os érték megváltoztatható.

A KSH-ban gyakori, hogy aggregált idősorokat és azok alágazatait is ki kell igazítani szezonálisan. A szezonális kiigazítási módszereket nem ismerő felhasználó azt várná, hogy a szezonálisan kiigazított adatokra is teljesül az, hogy az alágazatok együttese kiadja az aggregátumot, azaz abszolút számok esetén az alágazatok összege megegyezik az aggregátummal, illetve indexek esetén az alágazatok súlyozott átlaga egyezik meg az aggregátummal. Ez sajnos nem feltétlenül teljesül automatikusan, sem a TRAMO/SEATS módszernél, sem számos más szezonális kiigazító módszernél, így az X12-ARIMA módszernél sem. A probléma kezelésére kézenfekvő a következő lehetőségek egyikét választani:

– **Direkt igazítás**. Ekkor az alágazatokat és az aggregátumot külön-külön igazítjuk. Az eredményül kapott szezonálisan kiigazított idősorokra nem teljesül az aggregációs feltétel, viszont mind az alágazatokra, mind az aggregátumra a legjobb minőségű eredményt kapjuk.

– **Indirekt igazítás**. Ilyenkor az alágazatokat kiigazítjuk, és a kiigazított idősorok aggregátumát fogadjuk el az aggregátum kiigazítottjának.

Ilyenkor az aggregációs feltétel definíció szerint teljesül, viszont az aggregátum kiigazításának minősége romlik.

– **Szétosztásos módszer**. A direkt igazítás egy változata, amikor az aggregátum kiigazítottja és az alágazatok kiigazítottjainak aggregációjával kapott sorok közötti eltérést az alágazatok között szétosztjuk. Ezzel az alágazatok kiigazításának minőségét némileg rontjuk, viszont az aggregációs feltétel teljesül és az aggregátum szezonális kiigazítása is jó minőségű marad.

E három lehetőség közül a KSH-ban rendszerint az elsőt, azaz a direkt igazítást választjuk, tekintve, hogy ez adja a legjobb eredményt mind az alágazatokra, mind az aggregátumra. Az inkonzisztencia a KSH-ban többségben lévő bázisévhez viszonyított indexek igazításánál általában nem szembetűnő, de abban az esetben nyílvánvalóvá válhat, amikor két alágazat van és az aggregált kiigazított indexe nem esik a két alágazat kiigazított indexe közé. Amennyiben előírás vagy határozott igény van az aggregációs feltétel teljesülésére, akkor a szétosztásos módszert alkalmazzuk, mivel az aggregátum minősége fontosabb, mint az alágazatoké.

Szintén probléma jelentkezik az időbeli aggregációnál: a szezonálisan kiigazított adatok éves összege nem feltétlenül adja ki az eredeti, kiigazítatlan adatok éves összegét. Amennyiben igény van arra, hogy ez a két összeg megegyezzen, akkor a különbség szétosztható az év szezonálisan kiigazított adatai között, de ezzel rontjuk a kiigazítás minőségét, éppen ezért ezt a megoldást lehetőség szerint nem alkalmazzuk.

Gyakori, hogy a KSH a szezonálisan kiigazított adatok mellett (vagy helyett) – elsősorban az Eurostat igényeit kielégítve – úgynevezett munkanappal kiigazított adatokat is előállít. Ezek az idősorok úgy állnak elő, hogy az eredeti idősorokból csak a munkanap- és húsvéthatást szűrjük ki, a szezonális hatást nem. Amennyiben az idősornál jelentkezik munkanap- vagy húsvéthatás, akkor a munkanappal kiigazított adatok alkalmasabbak lehetnek az „időszak / előző év azonos időszaka” típusú összehasonlításra, mint az eredeti idősor adatai. Abban az esetben, ha sem munkanap-, sem húsvéthatás nem jelentkezik az idősornál, akkor a munkanappal kiigazított idősor megegyezik az eredeti idősorral.

A munkanappal való kiigazítás és a szezonális kiigazítás között – az előírásoknak megfelelően – összhang van, mert a szezonális kiigazításnál becsült munkanap- és húsvéthatást szűrjük ki a csak munkanappal való kiigazításnál is. A munkanappal kiigazított idősort a Demetra nem állítja elő közvetlenül, de az eredményül kapott idősorokból kiszámítható. A munkanappal kiigazított idősoroknál is felmerül az alágazatok aggregációs problémája és az időbeli aggregáció problémája, a választott megoldások itt is ugyanazok.

Az Eurostat rövid távú mutatókra (STS) vonatkozó legfrissebb ajánlása alapján a munkanappal kiigazított idősoroknál biztosítjuk, hogy a bázisév indexeinek átlaga a kiigazítatlan és a munkanappal kiigazított adatokra megegyezzen, azaz 100% legyen. Ezt a munkanappal kiigazított idősor átskálázásával oldjuk meg, ami azt jelenti, hogy a munkanappal kiigazított idősor bázisévi átlagával leosztjuk a kiigazított sor minden egyes értékét. Ez az átskálázás nem változtatja az idősor dinamikáját, azaz az egyes időszakok értékeinek egymáshoz képesti aránya nem változik, csupán a bázisévhez viszonyított indexek értékei változnak meg az. Hasonló előírás a szezonálisan kiigazított adatokra nincs, ezért azokat nem skálázzuk át.

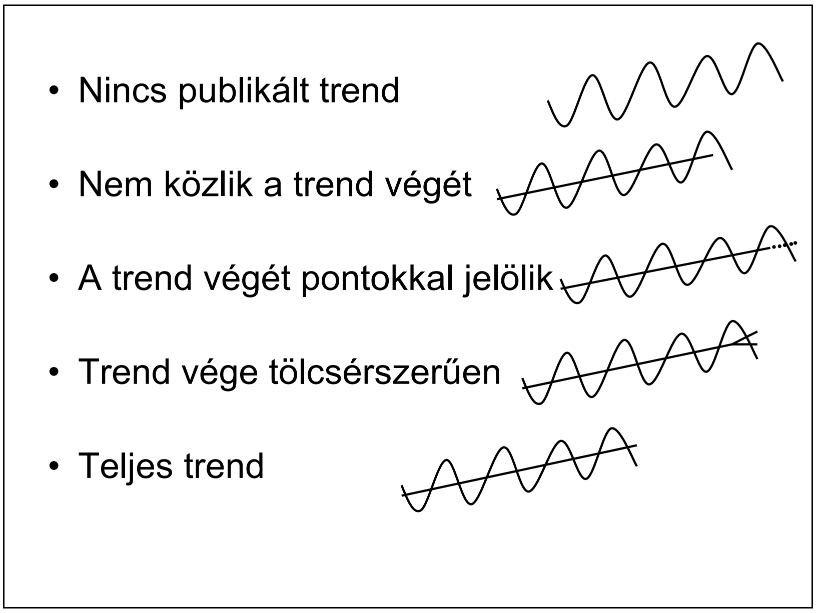
Publikációs szempontból problémát jelent, hogy a trend utolsó néhány értéke meglehetősen bizonytalan, az újabb megfigyelésekkel kiegészült idősorra végrehajtott kiigazítás során nagymértékben változhat visszamenőlegesen. Vannak, akik fontosnak tartják, hogy ez a bizonytalanság valamilyen módon a trend grafikonján is jelezve legyen, mások azon az állásponton vannak, hogy mivel a trend utolsó néhány adata szezonális kiigazítási módszertől függetlenül mindenképpen bizonytalan, ezért ezt felesleges külön jelezni. Éppen ezért a trend publikációjára többféle gyakorlat létezik, amelyek alkalmazása nemhogy statisztikai hivatalonként, de gyakran egy hivatalon belül témakörönként is változik.

– Az egyik lehetőség az, hogy egyáltalán nem publikálunk trendet. Ez az eljárás nem igazán szerencsés, véleményünk szerint a trend publikációjára szükség van, mert a kevésbé hozzáértő olvasó számára többet mond, mint a szezonálisan kiigazított adatsor.

– Lehetséges, hogy a trend végét nem publikáljuk. Igaz, hogy így kiiktatjuk a trend végének bizonytalanságát, de a trend információ tartalmát is csökkentjük, mert az a tény, hogy a trend végének adatai változhatnak, nem jelenti, hogy ezek az adatok nem hordoznak információt, és a trend végének változásából (az ún. csapkodó farok) is lehet következtetéseket levonni. Különösen hasznos lehet, ha a trendet a szezonálisan kiigazított adatokkal együtt vizsgáljuk. Itt érdemes megjegyezni, hogy ha a trendet egyszerűen az eredeti adatsorra illesztett mozgóátlaggal számoljuk, akkor a trend végének kiszámítására nincs is lehetőség (hacsak nem készítünk előrejelzést az eredeti idősorra).

– Elképzelhető, hogy a trend végének néhány adatára nem hosszabbítjuk meg a trendvonalat, hanem csak ponttal jelöljük értéküket. Ezzel a megoldással nem vesztünk információt, miközben a trend végének bizonytalanságát is jeleztük.
– Lehetséges olyan megközelítés is, amikor a trend végét tölcsérszerű ábrával rajzoljuk, azaz megadjuk azt a tartományt, amelyben a trend nagy valószínűséggel haladni fog.
– Végül pedig a legegyszerűbb (a trend elhagyásán kívül), hogy a trendet teljes hosszúságában külön jelölések nélkül közöljük.

4. ábra: A trend publikálásának lehetőségei



A KSH-n belül a trend publikációjának kérdésében nem született konszenzus, ezért a főosztályok saját hatáskörben, a Tájékoztatási főosztállyal egyeztetve döntenek a trend publikációjának módjáról. Jelenleg a trend teljes hosszúságú publikációját, a trend végének elhagyását és a trend közlésének elhagyását is alkalmazzák a KSH-ban.

2002 második felében elkészültek a kiigazításra kerülő fontosabb idősorok belső dokumentációi. Az egyes dokumentumok évente frissülnek, tartalmazzák az idősorok főbb jellemzőit (hossz, bázisév, gyakoriság, publikálás), az aktuális kiigazítás eredményeit (paraméterek, kiugró értékek, munkanap- és húsvéthatás) és ezek gazdasági értelmezését. Ezek a dokumentumok biztosítják az elvégzett munka folyamatos nyomon követhetőségét, a hozzáférhetőséget és az átláthatóságot.

A KSH-ban a szezonális kiigazítás elvégzésére munkamegosztást alakítottunk ki. A szakfőosztályok végzik saját idősoraik évközi kiigazítását, ők rendelkeznek a szükséges szakértői információkkal és felelősek az eredmények publikálásáért. A Mintavételi és módszertani osztályon történik az évenkénti (és az esetleges évközi) modell- és paraméterrögzítés, az évközi igazítás eredményeinek matematikai statisztikai ellenőrzése, ezen kívül itt történik a szezonális kiigazítás módszertani koordinációja és a témakörben jelentkező tudományos eredmények követése is. A munkamegosztást és a főosztályok közötti kommunikáció folyamatát az 5. ábra mutatja.

5. ábra: A munkamegosztás a szakfőosztályok és a Mintavételi és módszertani osztály között

Összefogalalóan tehát az év eleji paraméterrögzítés folyamata az alábbiak szerint alakul:

– Az előző év decemberi adatának publikálása után – erre a havi idősorok többségénél februárban, negyedéves idősoroknál március-áprilisban kerül sor[22] – a szakfőosztályok az alapadatokat tartalmazó idősorokat megküldik a Mintavételi és módszertani osztály munkatársainak.

– A Mintavételi és módszertani osztályon megtörténik a modell és a paraméterek rögzítése. A rögzítés során a matematikai statisztikai szempontokon túl figyelembe vesszük a korábbi időszakokra vonatkozó szakértői információkat a munkanap- és húsvéthatás meglétéről, az outlierek helyéről, típusáról.

– A rögzített paramétereket és az új paraméterekkel történő kiigazítás várható eredményeit Mintavételi és módszertani osztály munkatársai elküldik a szakfőosztálynak.

– A szakfőosztály szakértője átnézi az eredményeket, különös tekintettel a munkanap- és húsvéthatásra, és az újonnan bejövő outlierekre vonatkozóan. A problémákat, észrevételeket a szakértők megbeszélik, és a megbeszélésnek megfelelően a Mintavételi és módszertani osztály munkatársai módosítják a paramétereket, ahol szükséges.

– A Mintavételi és módszertani osztály elküldi a végleges modell- és paraméter-beállításokat a szakfőosztály munkatársának, aki a következő évben ezen beállítások alapján végzi az idősorok szezonális kiigazítását. A tárgyidőszak januári, illetve I. negyedéves adata már az új beállítások alapján kerül publikálásra.

Az új paraméterekkel történő év közbeni igazítás lépései a következők:

– A szakfőosztály a rögzített paraméterek alapján a Demetra program segítségével elvégzi a kiigazítást.

– A kiigazítás eredményeit elküldi a Mintavételi és módszertani osztály munkatársainak, akik matematikai statisztikai szempontokat figyelembe véve megvizsgálják az eredményeket.

---

[22] Az adatok közlésének pontos dátumát mindig az aktuális tájékoztatási naptár tartalmazza, amely a KSH honlapjáról elérhető (www.ksh.hu).

- Amennyiben a kiigazítás során nem merül fel probléma, akkor az eredmények publikálhatóak.
- Ha valamelyik adatsornál probléma merül fel (például az outlierek száma az új outlierek miatt meghaladja a megengedett mértéket, vagy a program diagnosztikáinak eredménye alapján a rögzíztett modell nem megfelelő az idősorra), akkor a szakfőosztály és a Mintavételi és módszertani osztály közösen megvizsgálják az idősort (minden outlier közgazdaságilag indokolható-e, milyen folyamatok okozhatják az idősornak a korábbi jellegétől eltérő viselkedését, ami miatt a rögzített modell már nem megfelelő). Ilyen esetekben általában sor kerül a paraméterek újrabecslésére, és a publikálásra már módosított beállításokkal kerül sor.

Év közben az alapadatok revíziója miatt is előfordulhat a paraméterek újrarögzítése. A folyamat ebben az esetben az év eleji lépéseknek megfelelően történik, az új beállítások a szakfőosztály és a Mintavételi és módszertani osztály szakértőinek egyeztetésével alakulnak ki.

A szezonális kiigazítás során alkalmazott paramétereket, beállításokat jelenleg nem publikálja a KSH, de – amennyiben a felhasználók részéről erre igény mutatkozik – a szakfőosztályok elérhetővé teszik.

A KSH kiadványaiban az eredeti adatok mellett általában a szezonálisan kiigazított adatok és a trend, és ritkábban a munkanappal kiigazított adatok kerülnek publikálásra, illetve az ezen adatokból képzett indexek: eredeti adatoknál az „időszak/előző év azonos időszaka" mutató, a csak munkanappal- és a szezonálisan kiigazított adatok esetén az időszak/előző időszak" mutatók. A szezonálisan kiigazított adatokból származó „időszak/előző év azonos időszaka" mutatót nem közli a KSH – amennyiben ez a mutató az alapadatnál rendelkezésre áll –, hiszen ez az érték az eredeti adatsorból számított ugyanezen mutatótól általában csak kis mértékben tér el (az eltérés leginkább a mozgó szezonalitásból ered), és általában nem hordoz a felhasználók számára többletinformációt.

IRODALOMJEGYZÉK

Bauer P. – Földesi E. (2003). Észrevételek az idősorelemzési módszerek alkalmazásával kapcsolatos kérdésekhez. *Statisztikai Szemle*, 81. évf. 9. szám, szeptember, 826-831. p.

Bauer P. – Földesi E. (2004). A szezonális kiigazítás harmonizációja a Központi Statisztikai Hivatalban. *Statisztikai Szemle*, 82. évf. 8. szám, augusztus, 691-704. p.

Brockwell, P. J. – Davis, R. A. (1996). Introduction to Time Series and Forecasting. New York: Springer.

Eurostat (1998a). Seasonal Adjustment Methods – A Comparison for Industry Statistics, revised version. Luxembourg.
[web: http://forum.europa.eu.int/Public/irc/dsis/eurosam/library?l=/documents_methodological&vm=detailed&sb=Title. Utolsó megtekintés: 2005. július 12.]

Eurostat (1998b). Seasonal Adjustment Policy – Some Eurostat Proposals. SAM 98 Seminar, 22-24 October, Bucharest.
[web: http://europa.eu.int/en/comm/eurostat/research/noris4/documents/policy/index.htm. Utolsó megtekintés: 2005. július 12.]

Eurostat (2003). Feasibility study about Demetra in Eurostat/NSIs and ECB/NCBs. *Rendelkezésünkre bocsátotta Jean-Marc Museux az Eurostat részéről.*

Eurostat (2004a). Recommendations for Seasonal Adjustment in STS. Working Party, 12 November, Luxembourg.

Eurostat (2004b). Recommendations for Working-Day Adjustment in STS. Working Party, 12 November, Luxembourg.

Eurostat (2004c). Seasonal Adjustment in Eurostat: new parameters. Working Party, 12 November, Luxembourg.

Fischer, B. (1995). Decomposition of Time Series – Comparing Different Methods in Theory and Practice. Luxembourg.
[web: http://europa.eu.int/comm/eurostat/research/index.htm?http://europa.eu.int/en/comm/eurostat/research/noris4/&1. Utolsó megtekintés: 2005. július 12.]

Freschl, Gy. – Kotász, Gy. – Nyáry Zs. (1982). Bevezetés az idősori módszerek gyakorlatába. *KSH Statisztikai módszertani füzetek*. 1. szám.

Friss P. (2003). Kérdések az idősor-elemzési módszerek alkalmazásáról. *Statisztikai Szemle*, 81. évf. 7. szám, július, 588-595. p.

Gómez, V. – Maravall, A. (1996). Programs TRAMO (Time series Regression with ARIMA noise, Missing observations, and Outliers) and SEATS (Signal Extraction in ARIMA Time Series). Instructions for the User. Working Paper 9628, Servicio de Estudios, Banco de España.

Gómez, V. – Maravall, A. (2001). Automatic Modelling Methods for Univariate Series. In: Peña, D. – Tiao, G. C. – Tsay, R. S. (eds.): A Course In Time Series Analysis. New York: J. Wiley and Sons.

Halabuk, L. – Hrubos, I. – Hulyák, K. – Nyáry, Zs. (1964). Az idényszerű változások mérése és kiküszöbölése. *KSH Nemzetközi módszertani füzetek*. 6. szám.

Hamilton, J. D. (1994). Time Series Analysis. Princeton: Princeton University Press.

Hrubos, I. – Hulyák, K. – Paizs, J. – Theiss, E. (1968). Szezonális kiigazítási eljárások összehasonlítása. *KSH Ökonometriai füzetek*. 9. szám.

Hulyák, K. (1977). Idősorok sztochasztikus modelljei. *KSH Ökonometriai füzetek*. 13. szám.

Hunyadi L. – **Mundruczó Gy**. – **Vita L.** (1996). Statisztika. Budapest: Aula Kiadó.

Maravall, A. (1995). Unobserved Components in Economic Time Series. In: Pesaran, H. – Schmidt, P. – Wickens, M. (eds.): The Handbook of Applied Econometrics. Vol. 1, Oxford: Basil Blackwell.

Maravall, A. (1999). Short-Term Analysis of Macroeconomic Time Series. In: Kirman, A. P. – Gérard-Varet, L-A. (eds.): Economics Beyond the Millennium. Oxford: Oxford University Press.

OECD (2002). Harmonising Seasonal Adjustment Methods in European Union and OECD Countries. Short-term Economic Statistics Expert Group – Meeting, 24-25 June, Paris.
[web: http://www.oecd.org/dataoecd/1/9/1933606.doc. Utolsó megtekintés: 2005. július 12.]

Planas, C. (1997). Applied Time Series Analysis: Modelling, Forecasting, Unobserved Components Analysis and the Wiener-Kolmogorov Filter. Luxembourg.
[web: http://forum.europa.eu.int/Public/irc/dsis/eurosam/library?l=/documents_methodological&vm=detailed&sb=Title. Utolsó megtekintés: 2005. július 12.]

Sugár A. (1999a). Szezonális kisimító eljárások összehasonlítása. Gazdasági Minisztérium, Gazdaságelemző Intézet.

Sugár A. (1999b). Szezonális kiigazítási eljárások (I.). *Statisztikai Szemle*, 77. évf. 9. szám, szeptember, 705-721. p.

Sugár A. (1999c). Szezonális kiigazítási eljárások (II.). *Statisztikai Szemle*, 77. évf. 10-11. szám, október-november, 816-832. p.

Tusnády G. – Ziermann M. (1986). Idősorok analízise. Budapest: Műszaki Könyvkiadó.

A KSH módszertani kiadványsorozataiban eddig megjelent kötetek

Nemzetközi Módszertani Füzetek

1. A polgári nemzeti jövedelemstatisztikák módszertana, 1956
2. A ráfordítás-kibocsátás (input-output) rendszer vázlatos ismertetése, 1957
3. A polgári beruházási statisztika egyes módszertani kérdései, 1959
4. Volumenindexek számítása tőkés országokban, 1959
5. Az életszínvonal nemzetközi összehasonlításánál alkalmazott mutatószámok kiválasztása, 1960
6. Az idényszerű változások mérésének és kiküszöbölésének statisztikai módszerei, 1964
7. A magyar népgazdaság M-1. statisztikai makromodellje, 1965
8. Szimuláció statisztikai makromodellekkel, 1966

Ökonometriai Füzetek

9. Szezonális kiigazítási eljárások összehasonlítása, 1968
10. Az időjárás és a mezőgazdasági termelési eredmények, 1968
11. Információelméleti mérőszámok alkalmazása a gazdasági elemzésben, 1970
12. Az M-4. modell: input-output összefüggéseket tartalmazó ökonometriai modell, 1973
13. Idősorok sztochasztikus modelljei, 1977
14. Az aggregáció problémája a gazdasági elemzésben, 1977
15. Bevezetés a spektrálanalízisbe, 1978
16. Termelési függvények a gazdasági elemzésben, 1979

A KSH módszertani kiadványsorozataiban eddig megjelent kötetek
Módszertani Füzetek

1. Az ágazati kapcsolatok mérlege (A dinamikai összehasonlítás problémái), 1966
2. A termelékenység dinamikájának mérése az iparban, 1967
3. A magyar háztartás-statisztikai megfigyelés, 1968
4. Az ipari termelés indexei, 1968
5. A népgazdasági energiamérleg, 1968
6. A külkereskedelmi áruforgalmi statisztika módszertana, 1969
7. Az ipariár-indexek számítási módszere, 1970
8. A piaciár-indexek számítási módszere, 1971
9. A népgazdasági mérlegrendszer módszertana, 1973
10. A külkereskedelmiár-statisztika módszere, 1972
11. A beruházásiár-indexek számítási módszere, 1972
12. A ráfordítások hatékonyságának mutatószámrendszere az iparban, 1973
13. A kiskereskedelmiár-statisztika módszere, 1974
14. Építőipariár-statisztika, 1975
15. Közhasználatú közlekedési statisztika, 1976
16. A magánkisipar értéki mutatóinak számítási módszere, 1976
17. A külkereskedelmiár-statisztika módszere, 1977
18. Az ipar gyártási ágainak rendszere, 1976

Statisztikai Módszertani Füzetek

1. Bevezetés az idősori módszerek gyakorlatába, 1982
2. Az állatállomány számbavételének módszere Magyarországon, 1982
3. A népességtovábbszámítás módszere, 1982
4. A piaci statisztika módszere, 1983
5. A nemzetközi idegenforgalom határstatisztikai megfigyelésének módszertana, 1983
6. Módszertan és esettanulmányok a regressziószámítás köréből, 1983
7. Egységes Lakossági Adatfelvételi Rendszer (ELAR), 1984

8. Az ipari rendelésállomány alakulásának statisztikai vizsgálati módszerei, 1984
9. A külkereskedelmi forgalom statisztikájának módszertana, 1984
10. A beruházás ár- és költségindexek számítási módszere, 1984
11. Az időmérleg-vizsgálat módszertana az 1976–77. évi felvétel alapján, 1985
12. Kereseti függvény meghatározásának módszere, 1985
13. A gépállomány korának, műszaki színvonalának és kapacitáskihasználásának mérési módszerei az iparban, 1985
14. A mezőgazdasági termelési érték számításának módszere, 1985
15. Az építőiparár-statisztikai módszertana, 1985
16. Többváltozós matematikai-statisztikai módszerek együttes alkalmazásának bemutatása egy esettanulmány segítségével, 1985
17. Az ipariár-indexek számítási módszere, 1985
18. A lakosság kereseti struktúrájának összehasonlító vizsgálata LES-, AIDS- és ROTTERDAM-modellekkel, 1985
19. A mezőgazdaságiár- és az erdőgazdálkodásiár-statisztika módszere, 1986
20. A magyar háztartás-statisztikai megfigyelés módszertana, 1986
21. A külkereskedelmiár-statisztika módszere, 1986
23. A környezetstatisztika módszertana (Az ENSZ környezetstatisztikai keretrendszere), 1986
24. Az ELAR-minta és az 1984. évi mikrocenzus mintájának kiválasztási eljárása, 1987
25. Az Ágazati Kapcsolatok Mérlege szerkesztésének és mutatórendszerének módszertana, 1987
26. Módszertani ajánlás az újszülött populáció vizsgálatára, a hátrányos helyzetet előidéző okok feltárására, 1987
27. A környezetstatisztika módszertana II. Az ENSZ EGB környezetstatisztikai osztályozásai, 1987
28. A környezetstatisztika módszertana III. (A területfelhasználás statisztika környezeti vonatkozásai. Finnország–Magyarország—Svédország), 1987
29. A környezetstatisztika módszertana IV. (A KGST környezetstatisztikai mutatószámrendszere.)
30. A kiskereskedelmiár-statisztika módszere, 1989
31. A kisipariár-statisztika módszertana, 1989
32. Árindexek mintavételi hibájának számítása; alkalmazás a kiskereskedelmiár-indexre, 1990

33. A hozzáadottérték-súlyozású ipari termelési index számításának módszertana, 1994
34. A külkereskedelmi statisztika módszertana, 1994
35. Iparstatisztika. Rövid távú mutatók: termelés, értékesítés, rendelésállomány, 1996
36. A külkereskedelmiár-index számításának módszertana, 1997
37. A háztartási költségvetési felvétel módszertana, 1997
38. A munkaerő-felmérés módszertana, 1998
39. A fogyasztóiár-statisztika módszere, 2000
40. Munkaerő-felmérés módszertana, 2002
41. A negyedéves bruttó hazai termék (GDP) számítási módszere Magyarországon
42. A K+F-statisztika módszertana

CONTENTS