Hulyák Katalin

Idősorok rövidtávú előrejelzése ARIMA modellekkel[1]

1. Bevezetés

Míg a statisztikai elemzések egy csoportjánál feltételezzük, hogy megfigyeléseink egymástól függetlenek, a gazdasági, társadalmi stb. jelenségek megfigyeléséből nyert idősoraink értékei általában egymástól függőek, s éppen e belső függőség jellemzi a vizsgált jelenséget. Az ARIMA modellek a belső függőséget megragadva, olyan modellekkel írják le a statisztikai idősorokat, amelyekkel az idősorok jövőbeli értékei általában sikeresen előrejelezhetők.

A hetvenes évek elején G. E. P. *Box* és G. M. *Jenkins* [1], [2] egységes rendszerbe foglalták e modellek elméleti tulajdonságait és kidolgozási folyamatukat. Az említett szerzők kidolgozták az ARIMA modellek egységes modellrendszerét, s e rendszerből való választás stratégiáját. E modellekkel a legkülönfélébb idősorok viszonylag kevés paraméter felhasználásával leírhatók és előrejelezhetők. Az előrejelzett értékekhez megbízhatósági intervallumokat is rendelhetünk, és új információk beérkezésével az előrejelzések folyamatosan korszerűsíthetők a számítások ismételt elvégzése nélkül.

A modellek lényegében nem térnek el a sztochasztikus modellek (azok közül is az egy-egyenletes regressziós modellek) ismert típusától, s becslésük és felhasználásuk is ismert módszerek és tesztek szerint történik. Módszertani nehézséget csupán a nem-lineáris összefüggések paraméterbecslése okozhat, amely problémák a könyvtári programok alkalmazásával fokozatosan leküzdhetők.

Míg a korábbi idősorelemzési módszerek általában csak a módszer kiválasztását hagyták a felhasználóra, alkalmazásuk pedig már automatikus; a Box—Jenkins-féle módszerek alkalmazása nem teljesen automatizálható. Mind a modell kiválasztásánál, mind becslésénél és ellenőrzésénél a felhasználónak nagyobb szabadsága van a beavatkozásra az alkalmazás összes fázisában.

Az ARIMA modellek kedvező tulajdonságai mellett meg kell említenünk a módszer hátrányait is. Éppen a felhasználó nagyobb döntési szabadsága miatt fennáll a rossz döntések valószínűsége, annak a veszélye, hogy rosszul választunk. Mivel a módszer nem teljesen automatizálható, csak a módszert értő és ismerő végezheti az előrejelzést. Mindebből az is következik, hogy ugyanazokat az adatokat elemezve különböző felhasználók, különféle modelleket specifikálhatnak, s így előrejelzésük is különböző lehet. Végül meg kell említenünk, hogy csak viszonylag hosszú idősorok birtokában célszerű ARIMA modellek specifikálása.

[1] A VII. Magyar Operációkutatási Konferenciára (Pécs 1977) benyújtott dolgozat módosított változata.

Box és Jenkins alapművének [2] 1970-ben történt publikálása óta, egyre többen végeznek alkalmazási kísérleteket az ARIMA modellekkel, pl.: [3], [4], [5]. A Központi Statisztikai Hivatal Ökonometriai Laboratóriumában folyó idősorelemzési munka egyik részeként mi is kísérletet tettünk e módszer hazai ismertetésére és alkalmazására [6]. Jelen dolgozat e munkát ismerteti nagy vonalakban.

Először az ARIMA modellek fajtáit és jellemző tulajdonságaikat tárgyaljuk. Utána bemutatjuk a modellszerkesztés és előrejelzés folyamatát. Áttérve az alkalmazásokra, először magyar idősorokkal végzett számítások következnek, majd ismertetjük azokat az eredményeket, amelyeket NDK és lengyel gazdasági idősorok ARIMA modellekkel történő előrejelzésével nyertünk.

2. Az ARIMA modellek típusai

Az ARIMA modellek azon az alapvető elméleti feltevésen nyugszanak, hogy minden olyan idősor, amelynek egymásutáni értékei függenek egymástól, független véletlen ingadozásokból (fehér zajból) generálható [7]. Ezeket a véletlen ingadozásokat $a_t, a_{t-1}, \ldots$-vel fogjuk jelölni, s feltesszük, hogy várható értékük 0 és varianciájuk $\sigma^2_{a_t}$. Az a_t fehér zaj változó valamilyen szűrőn keresztül haladva generálja a z_t idősort. A modell éppen ezt a szűrőt jelenti, azaz azt a formulát szolgáltatja, amely a_t-ből a z_t idősort állítja elő. A modell változói között tehát csak az $a_t, a_{t-1}, \ldots$ és a $z_t, z_{t-1}, \ldots$ változók szerepelnek. Röviden ismertetjük azt az ötféle modelltípust, amelyekkel, feltevések szerint, midenféle idősor közelíthető. Ezek a modellek, rövidített nevükön, az AR-típusú, a MA-típusú, az ARMA-típusú, az ARIMA és a szezonális ARIMA modellek.

Az ARIMA modellek tárgyalásához szükséges néhány alapvető operátor:

Az idősorok sztochasztikus modelljeinek leírásában támaszkodunk az operátoros jelölésmódra, illetve az operátorokkal végzett műveletekre: az operátorszámításra. E helyen csak a tárgyaláshoz elengedhetetlenül szükséges operátorokat definiáljuk. (Bővebben foglalkozik az operátorokkal [6].)

A B *késleltető operátort* a következőképpen definiáljuk:

$$Bz_t = z_{t-1}.$$

A B inverz operátora B^{-1}:

$$B^{-1} z_t = z_{t+1}.$$

A ∇ *differencia operátor* a következő:

$$\nabla z_t = z_t - z_{t-1}.$$

A B és a ∇ operátor közötti összefüggés:

$$\nabla z_t = (1 - B) z_t.$$

Meg kell említenünk az operátorok *polinomjait*, mert a továbbiakban általában azok szerepelnek. Ezekkel a polinomokkal műveletek ugyanúgy végezhetők, mint a közönséges algebrában a polinomokkal.

E bevezetés után térjünk rá az ARIMA modellek különböző típusaira.

a) *Autoregresszív (AR) modellek*

Igen sok időbeli folyamat reprezentálására hatékonyan alkalmazhatók a p-ed rendű autoregresszív modellek. Ezekkel a modellekkel valamely idősor t-időszaki értékét saját, korábbi $t-1$, $t-2, \dots t-p$. időszaki értékeinek, valamint az a_t változónak lineáris függvényeként fejezzük ki:

$$z_t = \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \dots + \varphi_p z_{t-p} + a_t$$

vagy:

$$\varphi(B) z_t = a_t,$$

ahol $\varphi(B)$ p-ed rendű autoregresszív operátorpolinom:

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p.$$

Különösen gyakran használjuk az elsőrendű AR(1) és a másodrendű AR(2) modelleket:

$$z_t = \varphi_1 z_{t-1} + a_t,$$

$$z_t = \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + a_t.$$

b) *Mozgóátlagolású (MA) modellek*

Ebben a formában a z_t változót csupán az a_t (fehér zaj) változó: a_t, a_{t-1}, $\dots, a_{t-q}$ értékeinek véges aggregátumaként állítjuk elő:

$$z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q},$$

vagy

$$z_t = \theta(B) a_t,$$

ahol $\theta(B)$ a q-ad rendű mozgóátlagolás operátora:

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q.$$

Ezzel az operátorpolinommal lényegében a_t-ből mozgó átlagokat képezünk, de úgy, hogy a mozgó átlag súlyainak sem összegére, sem előjelére, semmiféle kikötést nem teszünk.

Az elsőrendű MA(1) folyamat modellje:

$$z_t = a_t - \theta_1 a_{t-1}.$$

c) *Vegyes típusú, autoregresszív és mozgóátlagolású (ARMA) modellek*

Az idősorok mind rugalmasabb modellekkel való közelítése érdekében fejlesztették ki az autoregresszív és mozgóátlagolású tagokat egyaránt tartalmazó, vegyes jellegű ARMA modelleket. Ezeknek a modelleknek a magyarázó változói között az idősor megelőző értékeihez nem egy egyszerű véletlen változó járul, hanem annak tetszésszerinti mozgóátlagolású aggregátuma. Ez utóbbi változó szerepeltetése a klasszikus regressziós modellek véletlen változójára tett szigorú feltételeket lényegesen fellazítja. Ezért a modell:

$$z_t = \varphi_1 z_{t-1} + \dots + \varphi_p z_{t-p} + a_t - \theta a_{t-1} - \dots - \theta_q a_{t-q},$$

vagy:

$$\varphi(B) z_t = \theta(B) a_t.$$

Az ARMA modellek már kétféle dimenzióval, p-vel és q-val jellemezhetők, amelyek az autoregresszivítás, ill. a mozgóátlagolás rendjét jelentik.

A legegyszerűbb, elsőrendű autoregresszív és elsőrendű mozgóátlagolású ARMA (1,1) modell a következő:

$$z_t - \varphi_1 z_{t-1} = a_t - \theta_1 a_{t-1}.$$

d) *ARIMA modellek nem stacionárius idősorokra*

Az a), b), c) pontban ismertetett modellek elsősorban stacionárius idősorokra alkalmazhatók. Igen sok idősor – különösen az erős fejlődési tendenciát mutató gazdasági jelenségeket leíró sorok – nem stacionárius jellegűek. Ezek a nem stacionárius sorok általában bizonyos fokú differenciáikra nézve már stacionáriusak. Ha ezekre a differenciákra írunk fel ARMA modellt, akkor nyerjük az ARIMA modellt.

Ha $w_t = \nabla^d z_t$-vel jelöljük a z_t idősor d-edik differenciáinak képzését, akkor a modell:

$$\varphi(B)\, w_t = \theta(B)\, a_t.$$

Az eredeti z_t adatokból:

$$\eta(B)\, z_t = \varphi(B)(1 - B)^d z_t = \theta(B)\, a_t,$$

ahol a $\eta(B)$ általánosított autoregresszív operátor a d-ed fokú differenciák képzését is magába foglalja. Így az ARIMA modell háromféle dimenzióval (p, d, q-val) jellemezhető, ahol p az autoregresszivitás rendje, d a differenciaképzés foka, q pedig a mozgóátlagolás rendje. Az ARIMA modellek speciális esetként, mind az AR (p, o, o) modellt, mind a MA (o, o, q) modellt, mind az ARMA (p, o, q) modelleket tartalmazzák. Gyakorlati alkalmazások szerint a legtöbb idősor közelíthető az ARIMA modellek valamelyikével p, d és q alacsony (1–2) értékei mellett. Kivételt képeznek a szabályos ingadozásokat tartalmazó, szezonális idősorok.

A leggyakrabban felhasznált ARIMA modelltípusok:

1. A (0,1,1) modell:

$$\begin{aligned} \nabla z_t &= a_t - \theta_1 a_{t-1} & \qquad p &= 0 \\ &= (1 - \theta_1 B)\, a_t. & d &= 1 \\ & & q &= 1 \end{aligned}$$

2. A (0,2,2) modell:

$$\begin{aligned} \nabla^2 z_t &= a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} & \qquad p &= 0 \\ &= (1 - \theta_1 B - \theta_2 B^2)\, a_t. & d &= 2 \\ & & q &= 2 \end{aligned}$$

3. Az (1,1,1) modell:

$$\begin{aligned} \nabla z_t - \varphi_1 \nabla z_{t-1} &= a_t - \theta_1 a_{t-1} & \qquad p &= 1 \\ (1 - \varphi_1 B) \nabla z_t &= (1 - \theta_1 B)\, a_t. & d &= 1 \\ & & q &= 1 \end{aligned}$$

e) *Szezonális ARIMA modellek*

Az idősorelemzés témakörében a szabályos ingadozások, a szezonális és ciklikus hullámzások vizsgálata igen elterjedt. A módszerek túlnyomó többsége az idősorokat előbb komponenseire: a trend-, a szezonális és a véletlen tényezőre bontja, majd ezeket külön-külön határozza meg. Előrejelzésnél azután a külön-külön előrejelzett tényezőket építik egymásra, s így származtatják az idősor előrejelzett értékét. Az ARIMA modellek szezonális idősorok közelítésére is kiterjeszthetők és jól alkalmazhatók anélkül, hogy az idősort komponenseire bontanák. Sőt e modellek alkalmazásánál nem kell előre eldöntenünk egy idősorról, hogy az szezonális vagy nem-szezonális, hanem a módszer alkalmazása során dől el, hogy a konkrét sor milyen modellel írható le optimálisan.

A szezonális ARIMA modellel az idősorokban érvényesülő belső függőséget két irányban vizsgáljuk. Az egyik, az egymásután következő havi (vagy negyedévi) megfigyelések egymásutáni folyamatos függősége, amire az előző típusú modellek is épülnek. A másik hatás, a különböző évek azonos hónapjai (negyedévei) között fennálló függőség; amely éppen a szabályos ingadozásokat tartalmazó sorok jellemző sajátsága. Mindkét függőségi rendszerre felírunk egy ARIMA modellt, majd ezeket egymásra építve kapjuk a következő szezonális (multiplikatív) ARIMA modellt:

$$\varphi(B)\,\Phi(B^s)\,\nabla^d\,\nabla^P z_t = \theta(B)\,\Theta(B^s)\,a_t,$$

ahol $\varphi(B)$ a modell „nem-szezonális" p-ed rendű autoregresszivitását kifejező operátorpolinom,

$\Phi(B^s)$ a modell „szezonális" P-ed rendű autoregresszivitását kifejező operátorpolinom,

$\theta(B)$ a modell „nem-szezonális" q-ad rendű mozgóátlagolását kifejező operátorpolinom,

$\Theta(B^s)$ a modell „szezonális" Q-ad rendű mozgóátlagolását kifejező operátorpolinom.

Ebben a modellben egy (p, d, q) dimenziójú és egy (P, D, Q) dimenziójú modellt építettünk egymásra s periódusú szezonalitást feltételezve. Ez a legáltalánosabb forma, amelyet $(p, d, q) \cdot (P, D, Q)_s$-ad rendű szezonális ARIMA modellnek nevezünk. A legtöbb szezonális idősor ennek az általános modellnek egyszerű változataival leírható, ahol p, d, q és P, D, Q értékei 0 vagy 1 értéket vesznek fel.

Elvégzett számításainkban (lásd a 4. pontban) igen nagy szerepe volt a szezonális idősorok előrejelzésének. Csaknem mindegyik választott idősorunk tartalmazott szabályos szezonális ingadozást. Ezekre a sorokra viszont minden esetben egy egyszerű szezonális modellt a $(0,1,1) \cdot (0,1,1)_{12}$ típusú ARIMA modellt tudtuk alkalmazni, amely explicit formában a következő:

$$z_t - z_{t-1} - z_{t-12} + z_{t-13} = a_t - \theta\, a_{t-1} - \Theta\, a_{t-12} + \theta\,\Theta\, a_{t-13},$$

azaz

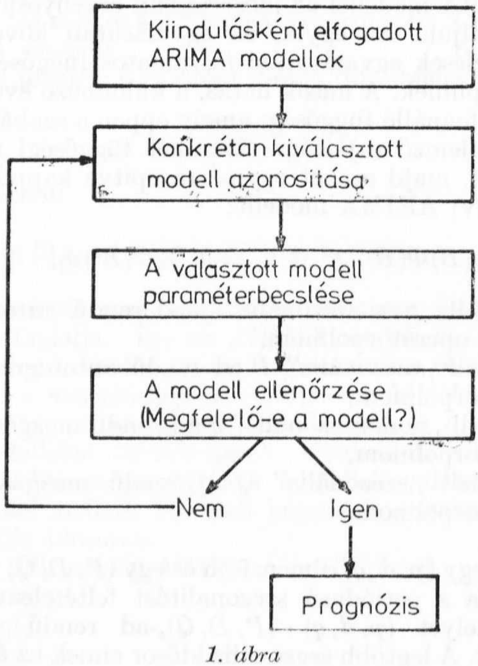
$$\nabla\,\nabla_{12} z_t = (1 - \theta\, B)\,(1 - \Theta B^{12})\, a_t.$$

Ez a modell olyan speciális szezonális ARIMA modell, amelynek mindössze két paramétere van: θ és Θ.

3. A modellszerkesztés folyamata

A jelenségekről, gazdasági folyamatokról rendelkezésre álló információnk nem elégséges ahhoz, hogy meg tudnánk szerkeszteni a folyamatok egzakt matematikai modelljeit. Ezért arra vagyunk utalva, hogy hiányos elméleti ismereteinket az empirikus információkkal kiegészítve és alátámasztva, ökonometriai modelleket szerkesszünk. Statisztikai idősorok elemzésénél maguk az idősorok tartalmazzák azt az empirikus ismeretanyagot, aminek alapján, ismerve a sztochasztikus folyamatok főbb típusainak jellemző vonásait, a megfelelő ARIMA modellt fel tudjuk írni.

A konkrét idősorra megfelelő modellt egy iterációs eljárással választjuk ki és számszerűsítjük, amelynek lényegét az 1. ábra blokk-diagramja szemlélteti.



1. ábra

Első lépésként a szóbajöhető modellek körét leszűkítjük. Ilyenkor a vizsgált idősorral végzett előzetes számítások alapján megállapítjuk a legjobban indokolt ARIMA modell (vagy modellek) p, d, és q értékeit, majd kezdőértéket számolunk a paraméterekre. Ezt a szakaszt nevezzük a *modell azonosításának* (vagy identifikációnak). Az azonosítás során az idősorból becsült autokorrelációkra és parciális autokorrelációkra támaszkodunk. Ezek a becsült mutatószámok az elméleti függvények közelítései, s ezért viselkedésükből következtethetünk az elméleti függvények tulajdonságaira. A különböző típusú ARIMA folyamatok elméleti autokorreláció és parciális autokorreláció függvényeinek jellemző sajátosságai viszont ismeretesek. A kétféle információ összevetése képezi a modellazonosítás alapját. Az identifikáció szempontjából nem közömbös, hogy a becsült mutatószámoknak mekkora a standard hibája. Ezért ebben a szakaszban a becsült autokorrelációk standard hibáját is megvizsgáljuk. Az azonosítás lezárul, ha a modelltípus(ok) kiválasztása megtörtént, a paraméterekre induló értékeket számoltunk és a becsült standard hibák elképzeléseink relatív helyességét támasztják alá.

A második szakaszban határozzuk meg a paraméterek végleges értékét. Ahhoz, hogy a paraméterek jó (hatásos) becslését nyerjük, rendelkezésre álló adatainkat is hatékony módon kell felhasználnunk a *becslés* érdekében. E hatékony módszerek közül (mint általában a statisztikai modellek illesztésénél) kiemelkednek a maximális esélyességen alapuló, és a legkisebb négyzetek elvén alapuló módszerek. Problémát jelenthet, hogy a linearitás feltétele nem áll fenn az ARIMA modellek összes paraméterére. Míg a φ autoregresszív paraméterekre lineáris függvényeket nyerünk, a mozgó-átlagolás θ paramétereinek normál-egyenletei nem-lineáris függvények. Ezért a mozgó átlagolású tagot is tartalmazó általános ARIMA modellek becslését a nem lineáris legkisebb négyzetek valamelyik iterációs módszerével becsüljük. Gyakorlati számítások szerint általában 4—6 iterációs lépés biztosítja a konvergenciát.

A becsült paraméterek hibáinak ellenőrzése már a modellszerkesztés következő szakaszának a *modell ellenőrzésének* a része. A modell felhasználása előtt meg kell vizsgálnunk az illesztett modell jóságát. Sokféle statisztikai vizsgálat ismeretes a modellek jóságának elemzésére. E tesztek közül az ARIMA modellek ellenőrzésére a reziduumokon alkalmazott eljárások a legelterjedtebbek. Ezek a módszerek nemcsak a modell viszonylagos megfelelő, ill. nem-megfelelő voltáról tájékoztatnak, hanem feltárják a hiányosságok jellegét is, s így útmutatást adnak a modell javítására is. Ha a modell ellenőrzése olyan súlyos hiányosságokat hoz felszínre, hogy a modell tovább nem tartható, akkor visszatérünk a modell-típus újabb azonosítására. Ebben az esetben a modell inkorrektségéről tanúskodó reziduumokat használjuk fel a második identifikációs szakaszban. Erre a második iterációs lépésre ritkán kerül sor.

Ha az ellenőrzés elfogadható eredményeket hoz, rátérünk az ARIMA modellek tulajdonképpeni felhasználási területére az *előrejelzésre*. A modell olyan rekurzív formulát szolgáltat, amellyel az idősor múltbeli értékeit felhasználva a jövőbeli értékeket előrejelezhetjük. Az előrejelzéshez szükségünk van a modell paramétereire, az z_t és az a_t múltbeli ismert értékeire. A még nem ismert értékek helyére várható értéküket helyettesítjük.

Az előrejelzéshez kapcsolódva megbizhatósági intervallumokat számolunk. Amennyiben a prognózist folyamatosan végezzük, lehetőségünk van az új beérkező információk alapján az előrejelzések folyamatos korszerűsítésére is, anélkül, hogy az összes számítást újra el kellene végeznünk.

4. Gyakorlati alkalmazások

A következőkben bemutatjuk az ARIMA modellek felhasználását illusztráló számításainkat és azok eredményeit.

a) Először *tizenöt hazai idősor ARIMA modelljeit* szerkesztettük meg. Az idősorokat úgy választottuk, hogy közöttük különféle — gazdasági, demográfiai, szezonális és nem-szezonális — típusú idősorok legyenek. Egységesen 120 megfigyelésből álló idősorokat vizsgáltunk, az 1964—1973. évi időszak havi adatainak alapján. Azt tűztük ki célul, hogy e tizenöt sorra alkalmazzuk az ARIMA modelleket, s előrejelezzük a sorok havi értékeit az 1974—1975.

évekre. A modell-azonosítás stádiumában úgy jártunk el, hogy maximálisan figyelembe vettük a szóbajöhető modelleket, s így a tizenöt sorra negyven modellt specifikáltunk. A kiválasztott modellek paramétereinek induló értékét a becsült autokorrelációkból számoltuk. A választott modellek között szezonális és nem-szezonális modellek, autoregresszív és mozgó átlagolású tagot tartalmazó és vegyes modellek is szerepeltek. Ezért a modellek becslését a legáltalánosabb ARIMA modellre kidolgozott programmal kellett végezni. A *Marquardt*-féle algoritmust [8] felhasználva, a legkisebb négyzetek nem lineáris modellre alkalmazott iteratív módszerével becsültük a paramétereket. Minden modell becslésénél legfeljebb tíz iterációs lépést próbáltunk ki. E tíz lépés alatt vagy elértük a konvergenciát (azaz a paraméterértékeket megközelítettük 1%-os pontossággal) vagy kiírattuk a tizedik lépés eredményeit. Huszonnyolc modell esetében általában 3—4, de legfeljebb 6 iterációs lépés után az eljárás konvergált. A továbbiakban csak ezekkel a modellekkel dolgoztunk. Bár ennek a huszonnyolc modellnek a becslési eredményei jónak mondhatók, mégsem sikerült minden egyes idősorra legalább egy „jó" modellt előállítani. Voltak viszont olyan idősorok, amelyekre 2—3 egymáshoz közeleső eredményeket hozó „jó" modellt illesztettünk. A becslést követően megállapítottuk, hogy kiválasztott idősorainak csaknem kivétel nélkül szezonális ingadozást tartalmaztak, s a szezonális modellek közül kiemelkedett fontosságban a (0,1,1) $(0,1,1,)_{12}$ típusú multiplikatív modell. Ezért az előrejelzést nem végeztük el az összes modellel, hanem egységes program alapján csak a (0,1,1) $(0,1,1)_{12}$ típusú modellekkel. Amire a számítások elkészültek előrejelzésünk ex-post jellegű lett, így össze tudtuk vetni az előrejelzett értékeket a tényadatokkal.

Az előrejelzések pontosságát az 1974. és 1975. évi havi tényadatok alapján egy δ-val jelölt mutatóval mértük. Ez a mutatószám az eltérések abszolút értékének a tényadatokhoz viszonyított átlagos százalékos értékét fejezi ki:

$$\delta = \frac{\sum\limits_{l} |z_{t+l} - \hat{z}_t(l)|}{\sum\limits_{l} z_{t+l}}.$$

Az előrejelzési intervallum (L = 24) elég hosszú ahhoz, hogy elemezhessük az intervallum hosszának és az előrejelzés pontosságának összefüggését. Így kiszámítottuk a δ mutatót L = 6, L = 12 és L = 24 esetében, vagyis a félévvel, egy évvel és két évvel előre számított prognózisokat elemeztük.

A következő táblázat a δ mutatókat tartalmazza[2]

A félévvel előre számított előrejelzések általában jól sikerültek. A tíz idősor közül nyolc esetében az előrejelzés átlagos hibája 5% alatt maradt; négy idősornál δ még a 3%-ot sem érte el. Az 5. sz. idősornál δ közel 7%-os értékű, a 10. sz. idősornál pedig a 10%-ot is meghaladta.

Az egy évvel előre számított előrejelzéseknél határozottan romlottak az eredmények. Már csak négy idősornál maradt 5% alatt az előrejelzés hibája. Másik négy idősornál 5% és 10% közé esett δ értéke. A maradék két idősornál a 10%-ot is meghaladta a hiba.

[2] Az előrejelzések részletes eredményeit és a konfidencia intervallumokat a [6] sz. forrás B függeléke tartalmazza.

	δ(%)		
	L=6	L=12	L=24
1. Élveszületések száma	4,4	11,7	12,6
2. Halálozások száma	4,3	6,6	8,0
3. Építőipari foglalkoztatottak	2,1	2,9	3,7
4. Gépipari munkások átlagbére	2,9	2,8	3,7
5. Széntermelés	3,6	5,7	6,0
6. Teherszállítás teljesítménye	6,9	6,6	6,9
7. Távolsági személyszállítás teljesítménye	3,7	3,0	2,5
8. Élelmiszer-forgalom	1,9	1,9	2,1
9. Ruházaticikk-forgalom	2,8	5,9	5,8
10. Magyarországra látogató külföldiek	12,8	25,2	26,7

A két évvel előre számított értékeknél az előrejelzések pontossága vagy nem változott, vagy csak igen csekély mértékben romlott az egyéves számításokhoz képest.

A táblázat alapján megállapítható, hogy az előrejelzések pontossága érzékenyen reagál az előrejelzési intervallum hosszának változtatására, de nem arányosan. Míg a félévhez viszonyítva az egy éves előrejelzések pontossága erősen lecsökkent, az egy éves és a két éves előrejelzések pontossága között nem mutatkozott lényeges különbség. Ugyancsak szembetűnő, hogy míg az előrejelzési intervallum kiterjesztése a viszonylag pontos ($\delta < 3\%$) előrejelzések pontosságát alig változtatja, addig a kevésbé jó vagy rossz előrejelzések fokozott mértékben romlanak.

Mivel a hosszabb intervallumra előre készített előrejelzések csak részben tekinthetők sikeresnek, s illusztrációképpen is, megvizsgáltuk az előrejelzések korszerűsítésének kérdését is.

A korrekcióknál az előrejelzésekkel egy időben számított ψ súlyokra (lásd [6]-ban a 105. oldalon), valamint az időközben beérkező tényadatokra támaszkodtunk.

A következő táblázatban a teljes kétéves (1974—75) időszak eredeti és korszerűsített előrejelzéseinek hibáit hasonlítjuk össze:

	δ(%)	
	eredeti	korszerűsített
	előrejelzések alapján	
1. Élveszületések száma	12,6	3,8
2. Halálozások száma	8,0	3,7
3. Foglalkoztatottak az építőiparban	3,7	0,5
4. Gépipari munkások átlagbére	3,7	2,4
5. Széntermelés	6,0	4,2
6. Teherszállítás teljesítménye	6,9	5,2
7. Személyszállítás teljesítménye	2,5	2,3
8. Élelmiszer-forgalom	2,1	—
9. Ruházaticikk-forgalom	5,8	4,1
10. Magyarországra látogató külföldiek	26,7	9,8

Az előrejelzések korszerűsítése utáni eredmények egyértelműen igazolják e lépés fontosságát, mivel lényegesen csökkentettük az előrejelzési hibákat. Összefoglalóan megállapítottuk:

1. Amennyiben csak fél évre készítettünk előrejelzést, a tíz sorból nyolc sorra jó ($\delta < 5\%$) eredményeket kaptunk.
2. A hosszabb, egy- és kétéves előrejelzéseknél csak a korszerűsített előrejelzésekkel tudtunk hasonlóan jó ($\delta < 5\%$) eredményeket elérni.
3. A tíz sorból egy idősornál a hiba minden változatnál magas volt, s ezt még a korrekció sem csökkentette elfogadható szintre, ezért e sor előrejelzésére nem volt alkalmas módszerünk.

b) A hazai idősorokon elvégzett számítások a módszer illusztrációját szolgálták, s ahogy említettük, az idősorokat is úgy választottuk, hogy a mintában sokféle tartalmú idősor szerepeljen. Az eredmények alapján megállapíthattuk, hogy a gazdasági jellegű idősorok (foglalkoztatottság, kereskedelmi forgalom stb.) előrejelzése jobban sikerült az ARIMA modellekkel, mint az egyéb jellegű (pl. demográfiai, meteorológiai) idősoroké. Ugyanezt támasztotta alá az a kísérlet is, amelynek során néhány szocialista ország gazdasági idősoraira végeztünk rövidtávú előrejelzéseket ARIMA modellekkel. Néhány szocialista ország gazdasági jelenségeinek dinamikai elemzése céljából, különböző országok hasonló jellegű idősoraira végzünk előrejelzést különböző idősorelemzési módszerek felhasználásával. E kutatási téma kapcsán előrejelzéseket végeztünk a $(0,1,1) \cdot (0,1,1)_{12}$ típusú szezonális ARIMA modellel is. A kísérlet azért is érdekes a módszer kipróbálása szempontjából, mert ezeknél a számításoknál csak viszonylag rövid (60 megfigyelésből álló) idősorokra támaszkodhattunk, amelynek alapján 12 megfigyelésre előre készítettünk prognózisokat. Az előrejelzések és a tényadatok összevetése tájékoztat az egyes módszereknek, s így az alkalmazott ARIMA modellnek az előrejelzési képességéről. A munka jelenlegi stádiumában, az NDK és Lengyelország néhány gazdasági idősorán elvégzett összehasonlító elemzés után azt mondhatjuk, hogy az ARIMA modellel, a többi alkalmazott módszerhez viszonyítva jobb vagy legalább olyan jó eredményeket értünk el.

A következő táblázatban bemutatjuk az NDK és lengyel idősorok előrejelzésének néhány eredményét. Az előrejelzések pontosságának elemzésére itt is a δ mutatót használjuk.

NDK gazdasági idősorok	$\delta(\%)$
Ipari termelés indexe	1,97
Elektromos energia	2,03
Építőipari termelés	2,86
Teherszállítás — autó	4,30
Teherszállítás — vasút	1,87
Személyszállítás	4,34
Felvásárlás — tej	2,17
Felvásárlás — tojás	3,36
Kiskereskedelmi áruforgalom	1,99
Kiskereskedelmi élelmiszer-forgalom	2,91
Kiskereskedelmi iparcikkforgalom	3,05

Lengyel gazdasági idősorok	$\delta(\%)$
Ipari termelés és szolgáltatások nettó értéke	3,04
Teherszállítás — vasút	3,03
Személyszállítás — vasút	2,04
Kiskereskedelmi élelmiszer-forgalom	2,96
Nem élelmiszer-cikkek forgalma	2,12
Felvásárlás — tej	3,03

(Beérkezett: 1977. november 30.)

IRODALOMJEGYZÉK

1. Box, G. E. P.—Jenkins, G. M.: *Some recent advances in forecasting and control I.* Applied Statistics 17. 91. 1968.
2. Box, G. E. P.—Jenkins, G. M.: *Time series analysis forecasting and control.* San Francisco, 1970. Holden Day.
3. Bray, J.: *Dynamic equations for economic forecasting with the G.D.P. — unemployment relation and the growth of G.D.P. in the U.K. as an example.* J. R. Statist. Soc. A. 134, 167. 1971.
4. Newbold, P.—Granger, C. W. J.: *Experience with forecasting univariate time series and the combination of forecasts.* J. R. Statist. Soc. A. 137.131. 1974.
5. Leskinem, E.—Teräsvirta, T.: *Forecasting the consumption of alcoholic beverages in Finland: A Box-Jenkins approach.* European Economic Review, Vol. 8. No. 4. 1976.
6. Hulyák, K.: *Idősorok sztochasztikus modelljei,* Ökonometriai Füzetek 13. sz. Budapest, 1977.
7. Yule, G. U.: *On a method of investigating periodicities in disturbed series with special reference to Wölfer's sunspot numbers.* Phil. Trans. A 226, 267. 1927.
8. Marquardt, D. W.: *An algorithm for least squares estimation of non-linear parameters.* Journal Soc. Ind. Appl. Math. 11, 431, 1963.
9. Goldfeld, S. M.—Quandt, R. W.: *Non-linear methods in econometrics.* Amsterdam, 1972. North Holland.

SHORT-TERM FORECAST OF TIME SERIES BY ARIMA MODELS

The article reports on the work done while constructing ARIMA models of individual time series and making short-term forecasts of time series on the basis of the models. The systematization of theoretical characteristics and elaborating processes of ARIMA (Autoregressive Integrated Moving Average) models is due to G.E.P. *Box* and G.M. *Jenkins*. According to the assumption of this method many different kinds of time series can be well described by some type of ARIMA models using a relatively small number of parameters and relying on the interdependence between consecutive elements of the time series. As a matter of fact, these models do not deviate much from the well-known type of stochastic, simple equation regression models and their estimation, evaluation and predictive application is also based on well-known techniques and tests.

In our experimental computations ARIMA models of fifteen time series were constructed at first. Time series were chosen in such a way that time series of various types — economic, demographic, seasonal and non-seasonal — could be found among them. Models were quantified on the basis of monthly data for the years 1964—1973 of the time series, i.e. on 120 observations, each. The best approximation was obtained most frequently by using a simple seasonal model of type ARIMA $(0,1,1) \cdot (0,1,1)_{12}$. With this model forecasts of the monthly data of ten time series for the years 1974 and 1975 were made then in the framework provided by this method updated forecasts for the same data were made.

Following the computations made with domestic time series short-term forecasts were made with the model ARIMA $(0,1,1) . (0,1,1)_{12}$ also for some economic time series of the GDR and Poland.

КРАТКОСРОЧНОЕ ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ С ПОМОЩЬЮ МОДЕЛЕЙ «АРИМА»

В данной статье описывается работа, в ходе которой составлялись модели «АРИМА» по индивидуальными временным рядам и на основании моделей осуществлялось краткосрочное прогнозирование временных рядов. Систематизация теоретических свойств и процессов разработки моделей «АРИМА» (Autoregressive Integrated Moving Average) связано с именами Г. Э. Бокс и Г. М. Дженкинс. В соответствии с предположением, выдвигаемым в рамках этого метода, посредством использования внутренних взаимосвязей следующих друг за другом элементов временных рядов большая часть самых различных временных рядов может быть хорошо описана каким-либо типом модели «АРИМА» при наличии относительно небольшого числа параметров. Эти модели, по существу, не отличаются от известного типа стохастической регрессивной модели с одним уровнением и оценка, использование при прогнозировании происходит также в соответствии с известными методами и тестами.

При проведении экспериментальных расчетов вначале были составлены модели «АРИМА» по 15 временным рядам. Ряды времени отбирались так, чтобы среди них были различные по типу временные ряды (экономические, демографические, сезонные и несезонные). В цифровом виде модели составлялись на основании месячных данных по рядам времени за 1964—1973 гг., т. е. 120 наблюдений. Наилучшее приближение при наибольшей частоте достигалось посредством одной простой сезональной модели, т. е. модели «АРИМА» тип $(0,1,1)\cdot(0,1,1)_{12}$ С помощью этой модели по 10 временным рядам прогнозирование проводилось по данным за 1974 и 1975 гг., а в последующем с помошью этого метода проводилось модернизированное прогнозирование.

После проведения расчетов по отечественным временным рядам краткосрочные прогнозы с использованием модели «АРИМА» тип $(0,1,1)\cdot(0,1,1)_{12}$ составлялись для ГДР и ПНР в отношении некоторых экономических временных рядов.