
Idősor analízis gyakorlati alkalmazása egészségügyi területen

Fazekasné Kis Mária

Debreceni Egyetem Agrártudományi Centrum,
Agrárgazdasági és Vidékfejlesztési Intézet,
Agrárinformatikai és Alkalmazott Matematikai Tanszék, Debrecen

ÖSSZEFOGLALÁS

A magyarországi lakosság körében a krónikus májbetegség és májzsugor, valamint a légcső-, hörgő-, tüdőrák okozta halálozási arányok idősoranalízis módszereivel történő vizsgálatait ismerteti a közlemény. WHO adatokra támaszkodva részletesen elemezésre kerülnek az adatok idősor modelljei és ezek megbízhatóságai. Az autoregresszív integrált mozgóátlag (ARIMA) modellekkel és autokorrelációs- és keresztkorrelációs függvényekkel azt vizsgáljuk, hogy a halálozási okok előfordulásaiban külső tényezőknek (pl. környezeti hatásoknak) lényeges szerepe van-e. Az autoregresszív (AR) együttható megbízhatóságát összehasonlítjuk a normális eloszlás, a White féle becslés és a folytonos idejű modelltől nyert konfidencia intervallumokkal.

Az elemzések eredményei alapján a krónikus májbetegség és májzsugor okozta halálozások esetében külső tényezők hatására lehet következtetni, a légcső-, hörgő-, tüdőrák okozta halálozásoknál viszont ezt nem lehetett kimutatni. Az AR(1) együtthatókra minden esetben a folytonos idejű becsléssel adódtak a legjobb eredmények.

A közlemény utal az ismertetett módszerek agrártudományi területeken történő alkalmazhatóságaira is.

SUMMARY

This article reviews the mortality data of chronic liver disease and cirrhosis, as well as tracheal, bronchial and lung cancers, in Hungary by methods of time series analysis. The methods of time series of death rates of chronic liver disease and cirrhosis as well as tracheal, bronchial and lung cancers and their reliability, are analysed from data available from WHO. The author used ARIMA models (autoregressive and integrated moving average models) and auto- and cross-correlation functions to study the substantial role an exogenous environmental factor has on incidences of death. The confidence intervals of autoregressive (AR) coefficient are compared to the standard normal distribution, the estimation of White's theory and the continuous time estimation model.

On the basis of the analysis, it may be concluded that chronic liver disease and cirrhosis can be influenced by an exogenous environmental factor, however, this relation cannot be demonstrated for deaths due to tracheal, bronchial or lung cancers. In each case, the continuous estimation of the AR(1) coefficients give the best results.

The paper demonstrates how the presented methods can be applied to agricultural science.

BEVEZETÉS

Gyakran előfordul, hogy közel egyenlő időközönként keletkezett adatokat kell elemezni, és a megfigyelési adatok között összefüggések lehetnek, az ilyen adatok képezik az idősorelemzés alapjait. Az empirikus idősorok jellemzését általában az adatok ábrázolásával kezdjük. A koordináta rendszer

vízszintes tengelyén az idő, az y tengelyén a megfigyelt adatok szerepelnek, alkalmas mértékegységben. Az ábrázolás alapján tájékozódhatunk arról, van-e valamilyen jellegzetes tendencia, azaz trend, vagy szezonális ingadozás az idősorban. Az empirikus idősor értékei egy olyan sztochasztikus folyamatot jellemeznek, amely három alapvető komponensre bontható: hosszú távon átlagosan érvényesülő hatásra, trendre, ciklikus vagy szezonális változásokat leíró komponensre és véletlen sztochasztikus hatásra. Az idősorok ilyen típusú felbontása Yule nevéhez fűződik. Az idősorokat jellemezhetjük egyrészt determinisztikus modellekkel, mely a determinisztikus komponensek függvény formájában történő meghatározását jelenti, másrészt sztochasztikus modellekkel, melyek az idősor véletlen komponensének viselkedését fejezik ki. A sztochasztikus folyamatok vizsgálata azért fontos, mivel a természeti jelenségek leírására szokásosan alkalmazott független valószínűségi változók sorozatai, ideális szemlélet. A jelenségeket jellemző véletlen mennyiségek időbeli egymás utáni megfigyelései statisztikailag nem függetlenek az esetek nagy hányadában. Az empirikus idősor valamely elméleti idősornak, a megfigyelt folyamatnak, egy parciális realizációja, mivel a megfigyelt adataink száma véges. A véletlen szerepét is figyelembe vevő idősorok, a sztochasztikus idősorok, az 1970-es években kezdtek ismertté válni. **Box** és **Jenkins** ebben az időszakban népszerűsítette az autoregresszív integrált mozgóátlag (ARIMA) modellek alkalmazásait [9-10, 28]. Magyar matematikusok is számos szakkönyvben foglalkoztak ezzel az egyre jelentősebbé váló témával [2-6, 8, 12-13, 25]. A számítógépek teljesítményének növekedésével fokozatosan lehetővé vált e modellek paramétereinek becslése.

Az ARIMA modelleket számos esetben közgazdasági folyamatok elemzésére alkalmazták.[7, 33, 38, 40, 42, 46]. Később egészségügyi felhasználások is ismertté váltak. Több érdeklődésre számot tartó problémát elemeztek ilyen módon. Nemzetközi viszonylatban vizsgálták, hogy nem fertőző betegségek, illetve tumoros betegségek, okozta halálozási arányok milyen idősori modellekkel adhatók meg. Az utóbbi időben az egyre fontosabbá váló külső környezeti tényezők hatásának kimutatásához is felhasználták. Elemezték, vajon a környezeti tényezőknek van-e hatása bizonyos betegség csoportok okozta halálozási arányokra [21]. A levegőszennyeződés és a halálozások számának alakulását is modellezték sztochasztikus idősorokkal [41]. Idősori módszerekkel keresték az összefüggést az influenza vírus terjedése és a szövődményes

esetek halmozott előfordulása között [48], valamint a balesetek számának és az időjárási viszonyoknak a kapcsolatára. Az említett modelleket a fentiekén kívül még több egészségügyi területen alkalmazták. [1, 11, 16-18, 22-23, 35, 39, 44]. ARIMA modelleket hazánkban főként közgazdasági elemzésekhez használták [33], ettől eltérő alkalmazásokról illetve kutatásokról is lehet magyar szerzők közleményeit olvasni [15]. Halálozási adatok elemzéséről természetesen számos magyar publikáció jelent meg, de ezek nem az említett elemzési eljárásokkal készültek [27, 29-31].

A közlemény a krónikus májbetegség és májzsugor valamint a légcső-, hörgő-, tüdőrák okozta halálozási adatok időszori modelljei paramétereinek és megbízhatóságainak becslését ismerteti. Továbbá ARIMA modellekkel és autokorrelációs- és keresztkorrelációs függvények felhasználásával elemzésre kerül, hogy a halálozási okok arányainak változására külső tényezők (pl. környezeti hatás) hatása lényeges-e. A standard normális eloszlás, a White féle becslés és a folytonos eloszlás alkalmazásával nyert konfidencia intervallumokkal az AR(1) folyamatok paramétereinek megbízhatóságára is becslést nyújt a közlemény. Az elsőrendű autoregresszív folyamatok tanulmányozásával több kutató is foglalkozott [2, 4-6, 8, 15, 24-25, 32, 34, 36-37, 43, 45, 47].

SZTOCHASZTIKUS IDŐSORI MODELLEK

Legyen a $z_t, z_{t-1}, z_{t-2}, \dots$ az egyenlő időközönként megfigyelt adatok sorozata, azaz egy idősor értékei a $t, t-1, t-2, \dots$ időpontokban. Legyen $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots$ független, azonos eloszlású véletlen változók sorozata. Ha a z_t megfigyelési érték lineárisan az előző z_{t-1} értéktől és a véletlentől függ, akkor $z_t = \phi z_{t-1} + \varepsilon_t$, ahol ϕ paraméter. Az ilyen modellt *elsőrendű autoregresszív modellnek*, *AR(1)-nek* nevezik [9-10, 25].

p-edrendű autoregresszív modell esetében az idősor t időponti értéke az idősor saját korábbi $t-1, t-2, \dots, t-p$ időbeli értékeinek lineáris kifejezéséből és a véletlent reprezentáló változótól függ, azaz

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + \varepsilon_t,$$

ahol ε_t a véletlen ingadozást fejezi ki, p az autoregresszivitás rendjét jelöli. A modell szokásos jelölése AR(p).

Ha z_t kifejezhető a jelen és egy előző véletlen lineáris kifejezésével, azaz

$z_t = \varepsilon_t - \theta \varepsilon_{t-1}$, ahol θ paraméter, ez *elsőrendű mozgóátlag MA(1) modellt* jelent.

q-adrendű mozgóátlag modell esetében az idősor t időbeli értéke a $t, t-1, t-2, \dots, t-q$ időbeli véletlen változók lineáris kifejezésével adható meg, azaz

$$z_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q},$$

ahol θ a mozgóátlag folyamat rendjét jelöli.

Az *autoregresszív és mozgóátlag (ARMA) modellek* az AR és a MA modellekből álló (p, q)-adrendű vegyes modellek.

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}.$$

Az ε_t véletlen változók sorozatáról feltételezzük, hogy fehér zaj folyamat. Ez esetben a várható értékük nulla, a szórásnégyzetük konstans és autokorrelálatlanok, azaz $E(\varepsilon_t) = 0$, $\text{var}(\varepsilon_t) = \sigma^2$, $\rho_k = 0$ ($k = 1, 2, \dots$). Az előző kifejezés szerint egy z_t idősort lehet modellezni az előző értékeiből és/vagy előző hibáiból.

A fentiekben megadott modellek speciális esetei az *integrált p-edrendű autoregresszív és q-adrendű mozgóátlag modelleknek*, melynek jelölése ARIMA(p, d, q). d a differenciaképzést jelenti. Jelöljük ∇ -al az idősor adataiból a differenciaképzést.

Az elsőrendű differenciák képzése: $\nabla z_t = z_t - z_{t-1}$.
A másodrendű differenciák képzése: $\nabla^2 z_t = \nabla(z_t - z_{t-1}) = \nabla z_t - \nabla z_{t-1} = z_t - 2z_{t-1} + z_{t-2}$.

Box és Jenkins által javasolt elnevezés, mint ARIMA modellezési módszer vált ismertté. Az AR, MA, és az ARMA modellek az ún. *stacionárius idősorok* elemzésére alkalmazhatók. Ilyen idősorok jellemzői: a várható érték, a szórásnégyzet, az autokorrelációs együtthatók időben állandóak, azaz függetlenek a t változótól. Vannak olyan idősorok, melyek stacionárius idősorrá transzformálhatóak d -szeres differenciaképzéssel. (rendszerint $d = 0, 1, 2$). Vagyis $w_t = \nabla^d z_t$, ahol w_t stacionárius idősor.

A *stacionárius idősorok* nem tartalmaznak trendhatást, az idősor értékei egy állandó átlagérték körül állandó szórással ingadoznak. A stacionárius idősorokra jellemző az idősorok autokorrelációs együtthatóinak időbeli állandósága is, mely az idősor adatai közötti belső összefüggéseket jelenti. Az autokorrelációs együtthatók értékei a változók egymás közötti távolságától, k -tól függenek, a t időtől nem. $k = 1, 2, \dots, K$. k -t késletetésnek nevezik. Az autokorrelációs együtthatókat a késletetés függvényében az autokorrelációs függvény adja meg, melyet *korrelogramnak* is szokás nevezni.

A stacionárius idősorok jellemzői a tapasztalati idősből becsülhetők. A megfigyelt időszori adatok átlaga és szórásnégyzete becslést ad a stacionárius folyamat várható értékre és a szórásnégyzetre.

$$\bar{z} = \frac{1}{n} \sum_{t=1}^n z_t, \quad S_z^2 = \frac{\sum_{t=1}^n (z_t - \bar{z})^2}{n}.$$

Az elméleti autokorrelációs függvény az empirikus autokorrelációs függvényvel becsülhető. Ha $k = 0, 1, 2, \dots, c_k$ és r_k az egymástól k időegységnyi távolságra lévő időszori értékek empirikus autokovariánciáit, illetve autokorrelációs együtthatóit jelöli [2, 3].

$$c_k = \frac{1}{n} \sum_{t=1}^{n-k} (z_t - \bar{z})(z_{t+k} - \bar{z}), \quad r_k = \frac{c_k}{c_0}.$$

Ezek alapján az autokorrelációs mátrix:

$$\mathfrak{R}_k = \begin{pmatrix} 1 & r_1 & \dots & r_j & \dots & r_{k-1} \\ r_1 & 1 & \dots & r_{j-1} & \dots & r_{k-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_j & r_{j-1} & \dots & 1 & \dots & r_{k-j+1} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{k-1} & r_{k-2} & \dots & r_{k-j+1} & \dots & 1 \end{pmatrix}.$$

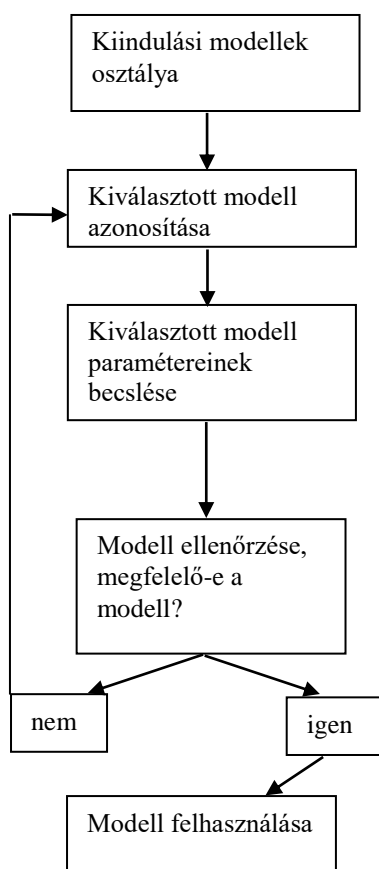
A parciális autokorrelációk (jelöljük ρ_k -val) az \mathfrak{R}_k mátrixból képezhetők, ha $k=0, 1, 2, \dots$

$$\rho_k = \frac{\det \mathfrak{R}_k^*}{\det \mathfrak{R}_k}.$$

Az \mathfrak{R}_k mátrix utolsó sorába behelyettesítve $r_1, r_2, r_3, \dots, r_k$ értékét kapjuk \mathfrak{R}_k^* -t.

Ha az idősor autokorrelálatlan, akkor a becült autokorrelációs együtthatók szignifikánsan nem különböznek nullától. Ezenkívül ha az idősor adatai nulla körül szóródnak állandó szórással, akkor az idősor véletlen, azaz fehér zaj folyamatot képez.

A SZTOCHASZTIKUS IDŐSORI MODELL-KÉSZÍTÉS FOLYAMATA



Két idősor adatai közötti kapcsolatot a keresztkorrelációs függvény fejezi ki. A z_t és y_t idősorokra a keresztkorrelációs függvény: $\rho_{zy}(k) = \text{korreláció}(z_t, y_{t+k})$, $k = 0, \pm 1, \pm 2, \dots$ k: az időkésltetés. Az empirikus keresztkorrelációs függvény interpretálása nehéz, különösen ha mindkét sorozat önmagában autokorrelált. Nagyon nehéz következtetni két idősor közötti megfelelő sztochasztikus kapcsolatra ezen függvény felhasználásával. Két idősor közötti kapcsolat elemzésére két lépésből álló módszer ismert [19-20]. Először mindkét idősorra meg kell adni a megfelelő egyváltozós modellt, majd a stacionárius idősorra transzformálás után nyert fehér zaj folyamatokra, a reziduumok sorozatára meghatározható a keresztkorrelációs függvény. Az autokorrelált sorozatból fehér zaj előállítását előszűrésnek nevezik. Az előszűrés után az idősorokból készített keresztkorrelációs függvény azt mutatja, hogy milyen k időkésltetés esetén van a két idősor kapcsolatban.

A nem stacionárius idősorok nagy része megfelelő számú differenciaképzéssel stacionáriussá tehető. Lineáris trendhatás esetében már az első differenciák idősora konstans szint körül ingadozik. A stacionáriussá tett idősorból készülnek a további modellezések.

A modell azonosításához az autokorrelációs- és a parciális autokorrelációs függvények felhasználhatók. Ha a megfigyelési adatok autoregresszív folyamatból származnak, akkor az autokorrelációs függvény értékei exponenciálisan vagy exponenciálisan és csillapodó szinusz görbéhez hasonlóan csökkennek. Ha az elsőrendű autokorrelációs együttható pozitív, azaz $\phi_1 > 0$, akkor az autokorrelációs függvény értékei mind pozitívak, ha $\phi_1 < 0$, akkor az autokorrelációk értékei váltakozó előjelűek.

p -edrendű autoregresszív folyamat esetében a parciális autokorrelációs függvény első p darab együtthatója szignifikánsan különbözik nullától, a többi együtthatóra ez nem teljesül. A parciális autokorrelációk tehát az AR folyamat rendjének becsléséhez alkalmazhatók [10, 12-13].

A MA folyamatok esetében, ha a stacionárius vagy azzá transzformált idősor első q darab autokorrelációs együtthatója szignifikánsan különbözik nullától, q -adrendű mozgóátlag modell illesztésével célszerű próbálkozni. Ekkor a parciális autokorrelációs függvény exponenciálisan vagy csillapodó szinusz görbéhez hasonlóan csökken.

Vegyes jellegű folyamat, azaz autoregresszív mozgóátlag modell (ARMA(p, q)) esetében az autokorrelációs- és a parciális autokorrelációs függvény értékei néhány jellemző érték után exponenciálisan vagy csillapodó szinusz görbe szerint csökkennek.

Az előzők szerint a becült autokorrelációs- és parciális autokorrelációs függvények alapján eldönthető, milyen típusú a modell és p és q értékeit is meg lehet határozni.

A választott modell ellenőrzéséhez meg kell vizsgálni a becült paraméterek standard hibáját és

szignifikanciáját, valamint a tapasztalati reziduumok véletlen jellegét.

A becült paraméterek megbízhatóságát konfidencia intervallumokkal is becsülhetjük. Többféle módon történhet a konfidencia intervallumok megadása.

KONFIDENCIA INTERVALLUM BECSLÉSEK AZ ELSŐRENDŰ AUTOREGRESSZÍV FOLYAMATOK PARAMÉTEREIRE

Az elsőrendű autoregresszív folyamatok ϕ paraméterének becslését konfidencia intervallumokkal az irodalomban leírt módszerek szerint végeztük [2, 6, 8, 12].

Az elsőrendű autoregresszív folyamat ($z_t = \phi z_{t-1} + \varepsilon_t$) ϕ paraméterének becslését jelölje $\bar{\phi}$, mely az empirikus idősor alapján meghatározható. A paraméter becslésére konfidencia intervallum a normális eloszlás alkalmazásával az alábbi összefüggés alapján nyerhető, ahol x_p a standard normális eloszlás p -kvantilise, n az idősor elemeinek száma [2]:

$$\bar{\phi} - x_p \sqrt{\frac{1-\phi^2}{n-1}} < \phi < \bar{\phi} + x_p \sqrt{\frac{1-\phi^2}{n-1}}.$$

Az AR(1) paraméterre konfidencia intervallum a White félé becsléssel kapható a következő képlet szerint, ahol t_p az $n+1$ szabadságfokú Student eloszlás p -kvantilise [2]:

$$\bar{\phi} - t_p \sqrt{\frac{1-\phi^2}{n+1}} < \phi < \bar{\phi} + t_p \sqrt{\frac{1-\phi^2}{n+1}}.$$

Ritkán alkalmazott eljárás az ún. folytonos idejű modellből adódó közelítés [2, 4, 6]. A folytonos idejű modellből adódó becsléssel az idősor adataiból az AR(1) folyamatra $\phi = e^{-\lambda T}$ összefüggés alapján becsülhető a ϕ paraméter [2]. Ha az idősor adatai egyenlő időközönként adóttak a $[0, T]$ intervallumban, akkor $\Delta = 1/T$. Behelyettesítve az előző kifejezésbe $\phi = \exp(-\lambda/T)$. Az egyenlet mindkét oldalának logaritmusát képezve és átrendezés után adódik $\lambda = -T \ln \phi$. Amennyiben $T = n$, ahol n az idősor elemeinek száma, akkor $\lambda = -\ln \phi$. Vezessük be a $\kappa = \lambda T$ jelölést. A [2] irodalomban megadott táblázat – „A csillapodási tényező becslésének az eloszlása” – tartalmazza a κ paraméter becült értékeit, melyek felhasználhatók az AR(1) folyamatok paramétereinek megbízhatóságát jellemző konfidencia intervallumok becslésére [2].

Ismert, hogy $r_1 \approx \phi$, azaz az idősor adataiból a becsülhető elsőrendű autokorrelációs együtthatóval közelíthető a ϕ paraméter értéke. A választott p értékhez tartozó csillapodási tényezőt jelöljük λ_1 -el. Hasonlóan meghatározható λ_2 . Képezhető a λ_1/T és a λ_2/T arány, majd a konfidencia intervallum ϕ_1 és ϕ_2 végpontjai az alábbi kifejezésekkel kaphatók meg:

$$\phi_1 = e^{-\frac{\lambda_1}{T}} \quad \text{és} \quad \phi_2 = e^{-\frac{\lambda_2}{T}}.$$

A fentiekben ismertett módszerekkel történt a két kiválasztott betegség okozta halálozási adatokból az AR(1) együtthatók konfidencia intervallumainak meghatározása.

Ha a konfidencia intervallum becsléseket p különböző értékeire elvégezzük, azt kapjuk, hogy a p értékének csökkenésével a konfidencia intervallumok hossza növekednek.

ARIMA MODELL KÉSZÍTÉSÉHEZ ÉS A MODELLEK PARAMÉTEREINEK BECSLÉSÉHEZ FELHASZNÁLT ADATOK

Az ARIMA modellezési eljárás alkalmazására példaként bemutatásra kerülő adatok egy WHO adatbázisból származnak. Az adatbázisban 1970-től 1997-ig a főbb halálozási okok standardizált adatai szerepeltek két korcsoportban (0-64 évig és 65 év feletiek) és összesítve, mindkét korcsoportban nőkre és férfiakra is. Az elemzések eredményeit két jellemző példán mutatjuk be.

KRÓNIKUS MÁJBETEGSÉG ÉS MÁJZUGOR OKOZTA HALÁLOZÁSI ADATOK ELEMZÉSE

A férfiak és a nők halálozási arányai az összes korcsoportban 1994-ig növekedtek, a férfiaknál fokozottabban, mint a nőknél. Később mindkét nem esetében kismértékű csökkenést láthatunk az 1. ábrán.

1. ábra: Krónikus májbetegség és májzsugor okozta halálozási arányok

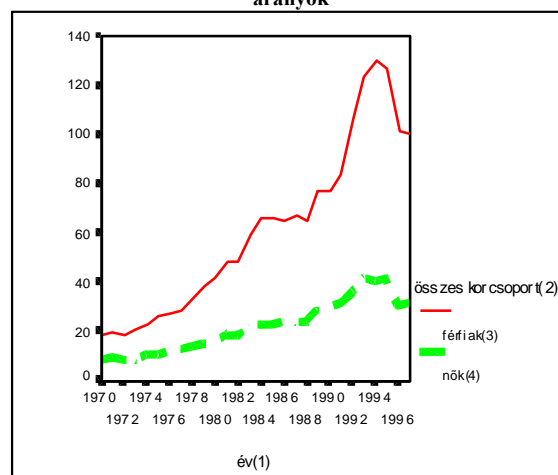


Figure 1: Mortality rates for chronic liver disease and cirrhosis year(1), all age groups(2), men(3), women(4), deaths per 100000 population(5)

A férfiak és a nők halálozási adataira az előzőekben röviden ismertett Box-Jenkins félé modellezési eljárást alkalmaztuk [10, 26]. Ehhez a modell azonosítási szakaszában meg kell határozni az idősorok autokorrelációs és parciális autokorrelációs

függvényeit. Az autokorrelációs függvény értékek pozitív értéktől fokozatosan csökkennek (2. ábra).

2. ábra: Krónikus májbetegség és májsugor okozta halálozási arányok autokorrelációs függvénye a nők adataiból

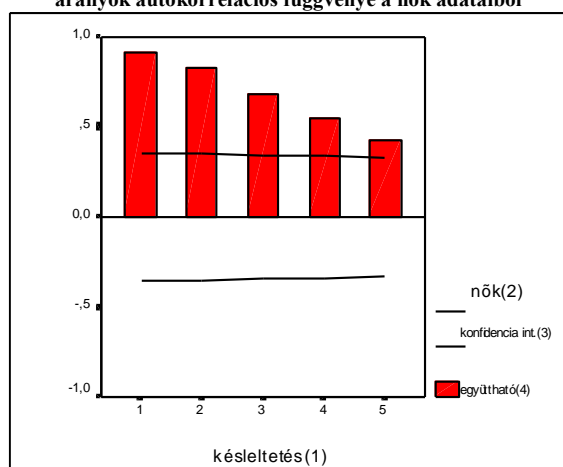


Figure 2: Autocorrelation function of mortality rates for chronic liver disease and cirrhosis from data of women lag(1), women(2), confidence interval(3), coefficient(4), autocorrelation(5)

A parciális autokorrelációs függvénynek $k=1$ késletetésnél szignifikáns értéke van, további késletetések esetén a függvényértékek a konfidencia intervallumon belül maradtak (3. ábra). Az ábrákon a konfidencia intervallumok kétszeres standard hiba távolságra vannak nullától, azaz $\pm 2\sigma$.

3. ábra: Krónikus májbetegség és májsugor okozta halálozási arányok parciális autokorrelációs függvénye a nők adataiból

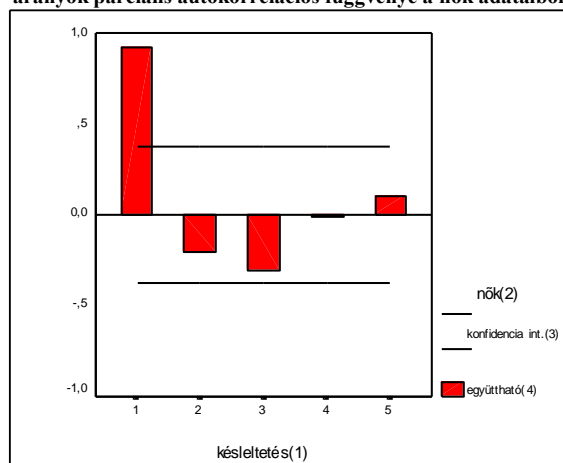


Figure 3: Partial autocorrelation function of mortality rates for chronic liver disease and cirrhosis from data of women lag(1), women(2), confidence interval(3), coefficient(4), partial autocorrelation(5)

Az autokorrelációs és parciális autokorrelációs függvények ilyen tulajdonságai alapján ismert [10, 12, 26], hogy elsőrendű autoregresszív modell (AR(1)) illesztését célszerű választani, melyet a $z_t = \phi_1 z_{t-1} + \varepsilon_t$ kifejezés jellemez. A ϕ_1 paraméter becült értéke az idősor adataiból számított r_1 empirikus autokorrelációs együttható, melyre a nők adataiból 0.917 és a férfiakéból 0.919 adódott a számítások

alapján. Így a nők halálozási arányát leíró sztochasztikus egyenlet $z_t = 0,917z_{t-1} + \varepsilon_t$; $\sigma_N = \pm 0,189$; férfiakra $z_t = 0,919z_{t-1} + \varepsilon_t$; $\sigma_F = \pm 0,189$. σ a standard hibát jelöli.

Az autoregressziós folyamatok paramétereinek megbízhatósága az előző fejezetben ismertett módon konstruált konfidencia intervallumokkal elemezhető. A normális eloszlás alkalmazásához és a White féle közelítéshez a szükséges $x_{0,05}$ és $t_{0,05}$ értékeket a standard normális eloszlás illetve a Student eloszlás ismert táblázataiból kikereshetők. A ϕ paraméter közelítő értéke az elsőrendű autokorrelációs együttható, azaz r_1 , melyet az idősor elemeiből az ismert képletek adtak meg. A konfidencia intervallum számítások eredményei a nők adataiból: Normális eloszlás alkalmazásával (0,709; 1,125), White féle becsléssel (0,712; 1,012), folytonos idejű közelítéssel (0,786; 0,999). A konfidencia intervallum számítások eredményei a férfiak adataiból: Normális eloszlás alkalmazásával (0,713; 1,124), White féle becsléssel (0,717; 1,121), folytonos idejű közelítéssel (0,789; 0,994).

A helyes modell választás igazolásához, az AR(1) paraméter vizsgálatán kívül, a reziduumok tulajdonságait is ellenőrizni kell. Ha a reziduumok sorozata nem autokorrelált és normális eloszlású nulla várható értékkel, $1/n$ varianciával, akkor megfelelő a modell [10, 12]. A χ^2 próbát alkalmazhatjuk ennek igazolására [19-20].

Az autokorrelációkból számolt $\chi^2_F = 8,971$; $\chi^2_N = 2,990$ értékek és $\chi^2_{0,05} = 9,488$ táblázatbeli érték alapján a választott modell helyesnek bizonyult [12, 19-20].

A férfiak és a nők halálozási arányait jellemző idősorok közötti kapcsolat elemzéséhez a keresztkorrelációs függvényt használtuk. Az alapadatokból képzett keresztkorrelációs függvényt mutatja a 4. ábra.

4. ábra: Krónikus májbetegség és májsugor okozta halálozási arányok keresztkorrelációs függvénye férfiak és nők adataiból

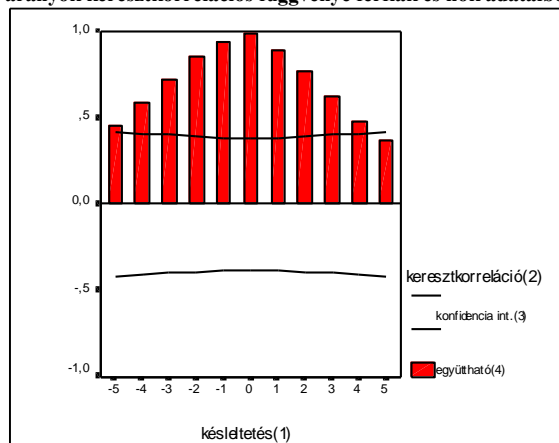


Figure 4: Cross-correlation function of mortality rates for chronic liver diseases and cirrhosis from data of men and women lag(1), cross-correlation(2), confidence interval(3), coefficient(4), cross-correlation(5)

Az első differenciák idősorából képzett keresztkorrelációs függvénynek $k=0$ késletetésnél

szignifikáns értéke van, mivel a keresztkorrelációs függvény értéke nagyobb a kétszeres standard hibájánál. A többi együtttható nullától szignifikánsan nem különbözik, értékük nem nagyobb a keresztkorrelációs függvényük kétszeres standard hibájánál (5. ábra). Két idősor közötti keresztkorrelációs függvény olyan időkésleltetésnél ad szignifikáns értéket, mikor az idősorok között korreláció van [10, 19-21].

5. ábra: Krónikus májbetegség és májzsugor okozta halálzási arányok keresztkorrelációs függvénye férfiak és nők adataiból az első differenciaképzés után

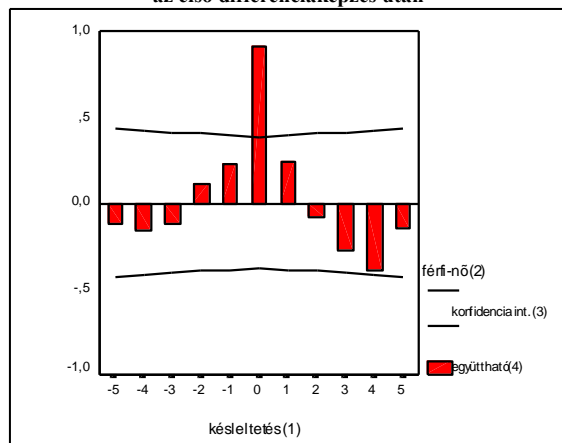


Figure 5: Cross-correlation function of mortality rates for chronic liver disease and cirrhosis from data of men and women after prewhitening
lag(1), men-women(2), confidence interval(3), coefficient(4), cross-correlation(5)

Az idősorok elemeinek hasonló viselkedése, azaz az idősorok „szinkronizációja”, a két időorból előállított első differenciák sorozatából az alábbi ábrán jól látható (6. ábra). Azokban az években, mikor növekedett a férfiak halálzási aránya, akkor a nőké is növekedett, ugyanez igaz a csökkenésekre is.

6. ábra: Krónikus májbetegség és májzsugor okozta halálzási arányok idősorai az első differenciaképzés után

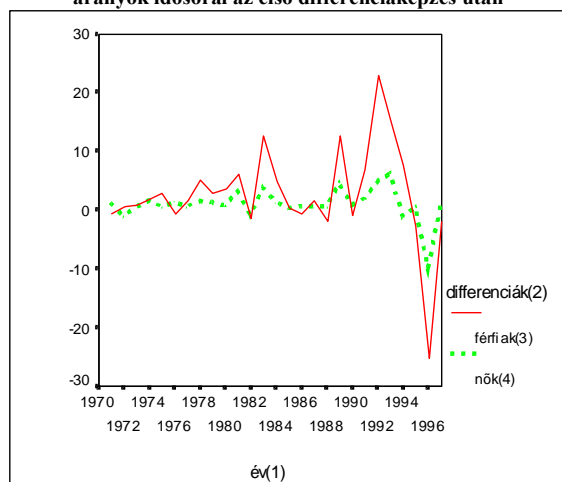


Figure 6: Time series of mortality rates for chronic liver disease and cirrhosis after prewhitening
year(1), differences(2), men(3), women(4), deaths per 100000 population(5)

LÉGCSŐ-, HÖRGŐ-, TÜDŐRÁK OKOZTA HALÁLZÁSI ARÁNYOK ÖSSZEHASONLÍTÁSA A 0-64 ÉVES KORCSOPORTOKBAN FÉRFIAKRA ÉS NŐKRE

A megfigyelt időszakban a nők és a férfiak halálzási arányai is növekedtek, a férfiaké lényegesen nagyobb mértékben (7. ábra).

7. ábra: A légső-, hörgő-, tüdőrák okozta halálzási arányok a 0-64 éves korcsoportokban férfiak és nők adataiból

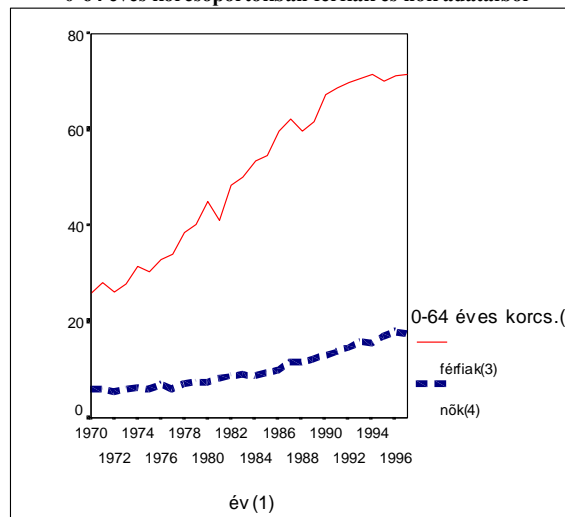


Figure 7: Mortality rates for tracheal, bronchial and lung cancers in the age classes of 0 to 64 years from data of men and women
year(1), 0-64 age-groups(2), men(3), women(4), deaths per 100000 population(5)

A Box-Jenkins féle modellezési eljárást az előzőekhez hasonlóan alkalmaztuk. Az autokorrelációs és parciális autokorrelációs függvények hasonlóak, mint a korábban elemzett betegség okozta halálzási adatokra. Az elsőrendű autokorrelációra a nők adataiból 0,843, a férfiak adataiból pedig 0,864 adódott. Így a megfelelő AR(1) modellek, $z_t = 0,864z_{t-1} + \varepsilon_t$; $\sigma_f = \pm 0,189$; nőkre $z_t = 0,843z_{t-1} + \varepsilon_t$; $\sigma_n = \pm 0,189$ [10, 13, 26]. σ a standard hibát jelöli.

A ϕ paraméter közelítő értéke az elsőrendű autokorrelációs együtttható, azaz r_1 , melyet az idősor elemeiből az ismert képletek adtak meg. A konfidencia intervallum számítások eredményei a nők adataiból: normális eloszlással (0,563; 1,123), White féle becsléssel (0,567; 1,118), folytonos idejű közelítéssel (0,675; 0,999). A konfidencia intervallum számítások eredményei a férfiak adataiból: Normális eloszlás alkalmazásával (0,601; 1,167), White féle becsléssel (0,615; 1,122), folytonos idejű közelítéssel (0,704; 0,999).

A helyes modell választás igazolásakor hasonlóan jártunk el, mint az előző elemzésnél. Az autokorrelációkból számolt $\chi^2_F = 6,147$; $\chi^2_N = 8,045$ értékek és $\chi^2_{0,05} = 9,488$ táblázatbeli érték alapján a választott modell helyesnek bizonyult [12, 19-20].

A férfiak és a nők halálzási arányait jellemző idősorok közötti kapcsolat elemzéséhez a

keresztkorrelációs függvényt használtuk. Az alapadatokból képzett keresztkorrelációs függvényt mutatja a 8. ábra.

8. ábra: A légső-, hörgő-, tüdőrák okozta halálási adatokból a 0-64 éves korcsoport adataiból a keresztkorrelációs függvény

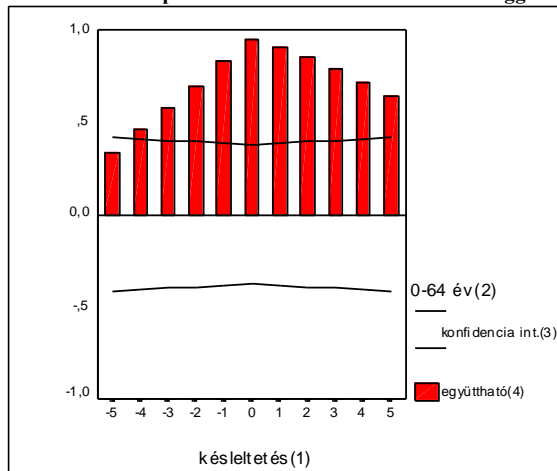


Figure 8: Cross-correlation function of mortality rates for tracheal, bronchial and lung cancers in the age classes of 0 to 64 years from data of men and women lag(1), 0-64 age-groups(2), confidence interval(3), coefficient(4), cross-correlation(5)

Az alapadatokból képzett keresztkorrelációs függvény értékei viszonylag tág késleltetési tartományba estek (8. ábra).

Az első differenciák idősoraiból készített keresztkorrelációs függvény nem mutatott szignifikáns értéket (9. ábra).

9. ábra: A légső-, hörgő-, tüdőrák okozta halálási adatokból a 0-64 éves korcsoport adataiból a keresztkorrelációs függvény az első differenciaképzés után

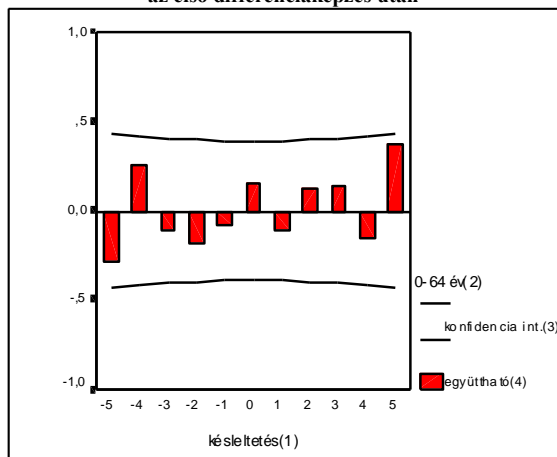


Figure 9: Cross-correlation function of mortality rates for tracheal, bronchial and lung cancers in the age classes of 0 to 64 years from data of men and women after prewhitening lag(1), 0-64 age-groups(2), confidence interval(3), coefficient(4), cross-correlation(5)

A 10. ábra azt mutatja, hogy a reziduumok idősorainak viselkedése eltérő minden időkésltetésnél, így azt mondhatjuk, hogy nincs

„szinkronizáció”. Ebből arra lehet következtetni, hogy azonos külső környezeti tényező hatása nem mutatható ki.

10. ábra: A légső-, hörgő-, tüdőrák okozta halálási adatokból a reziduumok idősorai

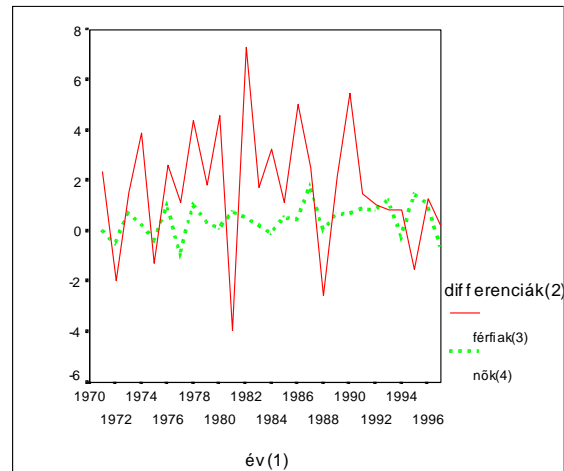


Figure 10: Series of the residuals from data of mortality rates for tracheal, bronchial and lung cancers year(1), differences(2), men(3), women(4), deaths per 100000 population(5)

MEGBESZÉLÉS

A krónikus májbetegség és májzsugor okozta halálási adatok idősorai hasonlóan, vagyis „szinkronizáltak” viselkedtek. A két nem adataiból készített keresztkorrelációs függvénynek a $k=0$ értéknél szignifikáns értéke volt. Ebből arra lehet következtetni, hogy minden időpillanatban hasonló a két idősor, azonos külső hatás hat rájuk, összhangban a [10, 19-21] irodalmakkal.

A légső-, hörgő-, tüdőrák okozta halálási adatok idősorai a két összehasonlított csoport adatai alapján nem mutattak „szinkronizáltságot”. Olyan következtetés vonható le, hogy nem lehet igazolni külső környezeti tényező hatását. A környezet hatásának értékelésekor azonban tekintetbe kell venni a módszer és a betegségek sajátosságait is.

A sztochasztikus idősorok elemzésére alkalmazható ARIMA modelleket és a modell paramétereinek becslését összehasonlításként több halálást előidéző betegségre is elkészítettem [14]. Az elemzési eredményeket e közleményben ismertett két példa jól jellemzi.

Magyarországi adatokból levont fenti következtetések megegyeznek az irodalomban, e tárgykörben végzett korábbi elemzésekből nyert következtetésekkel [21].

Az AR(1) folyamat ϕ paraméterére vonatkozó konfidencia intervallum becsléseiből látható, hogy a normális eloszlás felhasználásával és a White féle módszerrel körülbelül azonos nagyságú konfidencia intervallumokat lehetett nyerni, bár a White féle becslés kismértékben jobbnak bizonyult, mint a normális eloszlás alkalmazása, viszont folytonos becsléssel a legszűkebb konfidencia intervallumok adódtak.

Ha $\phi \approx 1$, akkor a normális eloszlással való közelítés nem minden esetben használható, mert a konfidencia intervallum felső határa 1-nél nagyobb is lehet. Ekkor stacionárius esetben használhatatlan az eredmény. Az első két eljárással a konfidencia intervallumok felső hataraként 1-nél nagyobb értékek is adódtak, melyek használhatatlanok [4]. A folytonos idejű modellből adódó becsléssel kapott konfidencia intervallumok felső határai minden esetben egy alatt maradtak. A fenti példák jól mutatják a folytonos idejű becslés előnyeit a másik két módszerhez képest.

Az elemzések arra is adatokat szolgáltatottak, hogy az AR(1) együtthatók konfidencia intervallumokkal történő becslésére alkalmazott módszerek közül a hazai kutatók által kidolgozott folytonos becslés

előnyösebben alkalmazható az irodalomban korábban leírtakhoz képest [3, 5, 35, 43].

A közleményben ismertetett elemzési módszerek és az autogeresszív paraméterek becslési eljárásai természetesen más empirikus idősorok elemzésére is alkalmazhatók. Az agrártudományok területén keletkező adatok vizsgálatára jól felhasználható lehet. Az agrárökonometriai számítások jól kiegészíthetők lehetnek a fentiekben ismertetett módszerekkel. A növénytermesztésben az egyes időjárási tényezőknek a termésátlagok alakulására gyakorolt hatása is elemezhető idősor analízissel. Az egészségügy területén alkalmazott elemzési módszereket a későbbiekben az agrártudományok egyes területeinek tudományos vizsgálatához tervezem felhasználni.

IRODALOM

- Allard, R.(1998): Use of time series analysis in infectious disease surveillance. *Bull World Health Organ* 76. 327-33.
- Arató M.-Benczúr A. (1986): Gauss-Markov folyamatok maximum-likelihood becslésének egzakt eloszlása. in: (Idősorok analízise, Ed. Tusnády G.-Ziermann M.) Műszaki Kiadó, Budapest, 85-117.
- Arató M.-Knuth E.(1970): Sztochasztikus folyamatok elemei. Tankönyvkiadó, Budapest
- Arato, M. (1982): Linear stochastic systems with constant coefficients: A statistical approach, Springer, Berlin
- Arató M. (1972): Komplex stacionárius Gauss Markov folyamat „csillapodási” paraméterének becslése és konfidencia intervallumainak megszerkesztése. MTA SZTAKI Közlemények, 122-161.
- Arató M.-Benczúr A. (1972): Szimulációs eredmények az elemi Gauss folyamat paraméterei becsléseinek eloszlására. MTA SZTAKI Közlemények, 8. Budapest, 3-34.
- Armstrong, J. S. (1978): Forecasting with Econometric Methods: Folklore versus Fact with Discussion. *Journal of Business* 51. 549-600.
- Benczúr A. (1971): Stacionárius Gauss Markov folyamat csillapodási paraméterének konfidencia határai. MTA Számítástechnikai Központja Közlemények, 6. Budapest, 1-14.
- Box, G. E. P.-Tiao, G. C. (1975): Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70. 70-79.
- Box, G. E. P.-Jenkins, G. M. (1970): Time Series Analysis Forecasting and Control Holden-Day. San Francisco
- Crabtree, B. F.-Ray, S. C.-Schmidt, P. M.-O'Connor, P. J.-Schmidt, D. D. (1990): The individual over time: time series applications in health care research. *J Clin Epidemiol*, 43. 241-60.
- Csáki P. (1986): ARMA folyamatok. in: (Idősorok analízise, Ed. Tusnády G.-Ziermann M.) Műszaki Kiadó, Budapest, 50-84.
- Éltető Ö.-Meszéna Gy.-Ziermann, M. (1982): Sztochasztikus módszerek és modellek. Közgazdasági és Jogi Könyvkiadó, Budapest
- Fazekasné Kis M. (2000): Idősor analízis alkalmazása halálózási okok adatainak elemzésére. NJSZT Neumann Kollokvium
- Gál Z.-Iglói E.-Terdik Gy. (1999): Nagysebességű informatikai hálózat adatforgalmának matematikai statisztikai jellemzése. *Alkalmazott Matematikai Lapok*, 19. 29-38.
- Garádi J.-Krámlí A.-Ratkó I.-Ruda M. (1975): Statisztikai és számítástechnikai módszerek alkalmazása kórházi morbiditás vizsgálatokban. MTA SZTAKI Tanulmányok, 35. 3-31.
- Gerencsér, L.-Michaletzky, Gy.-Vágó, Zs. (1998): Risk-sensitive identification of ARMA processes. *Proceedings of the 37th IEEE Conference on Decision and Control*, 1. 215-216.
- Hajtmán B. (1971): Bevezetés a matematikai statisztikába. Akadémiai Kiadó, Budapest
- Haug, L. D.-Box, G. E. P. (1977): Identification of Dynamic Regression (Distributed Lag) Models Connecting Two Time Series. *J Am Stat Assoc*, 72. 121-130.
- Haug, L. D. (1976): Checking the independence of two covariance stationary time series: a univariate residual cross-correlation approach. *J Am Stat Assoc*, 71. 378-385.
- Helfenstein, U. (1990): Detecting hidden relations between time series of mortality rates. *Methods Inf Med*, 29. 57-60.
- Helfenstein, U.-Wanner, H. U. (1991): The environmental accident at 'Schweizerhalle' and respiratory diseases in children: a time series analysis. *Statistics in Medicine*, 10. 1481-1492.
- Helfenstein, U. (1991): The use of transfer function models, intervention analysis and related time series methods in epidemiology. *International Journal of Epidemiology*, 20. 808-815.
- Huhn E. (1984): ARMA folyamatok egzakt sűrűségfüggvénye. *Alkalmazott Matematikai Lapok*, 10.
- Huhn E. (1990-91): Rekurzív algoritmus ARMA folyamatok likelihood függvényének számolására. *Alkalmazott Matematikai Lapok*, 15. 303-307.
- Hulyák K. (1976): Idősorok sztochasztikus modelljei. KSH Ökonometriai Laboratórium, Budapest
- Izsák J. (1997): A halálózási struktúra változásainak vizsgálata disszimilitási indexekkel. *Statisztikai Szemle*, 75. 130-140.
- Jenkins, D. M.-Watts, D. G. (1968): Spectral analysis and its applications Holden-Day. San Francisco
- Józan P. (1994): Epidemiológiai válság Magyarországon a kilencvenes években (I.). *Statisztikai Szemle*, 72. 5-20.
- Józan P. (1994): Epidemiológiai válság Magyarországon a kilencvenes években (II.). *Statisztikai Szemle*, 72. 101-113.
- Juvancz I.-Lipták T. (1952): A matematikai statisztika orvosi-biológiai alkalmazásának néhány problémájáról. *Alkalmazott*

- Matematikai Lapok, 175-205.
32. Knuth E. (1975): Központi egység kihasználtságának stimulációs vizsgálata másodrendű autoregresszív folyamatok alkalmazásával. MTA SZTAKI Közlemények, 15. 141-145.
 33. Kovács K. (1984): Idősorok alapján történő előrejelzés főbb módszertani kérdései: különös tekintettel a Box-Jenkins modellekre. Disszertáció. MKKE, Budapest
 34. Kormos J. (1989): Hipotézis vizsgálat közel nem stacionárius AR(1) esetén. Alkalmazott Matematikai Lapok, 14. 27-35.
 35. Krámlí A.-Ratkó I.-Ruda M.-Soltész J. (1977): A statisztikai adatfeldolgozás matematikai és számítástechnikai problémái. Hospitalizált morbiditási statisztikákkal kapcsolatos megfontolások (Esettanulmány). MTA SZTAKI Tanulmányok, 70. 5-67.
 36. Krámlí, A.-Pergel, J. (1974): The connection between Gaussian Markov processes and autoregressive-moving average processes. MTA SZTAKI Közlemények, 13. 53-57.
 37. Krámlí A.-Pergel J. (1975): Auttoregressziós típusú folyamatok által generált mértékek Radom-Nikodym deriváltjáról. Alkalmazott Matematikai Lapok, 1. 73-79.
 38. Liéli R. (1999): Egyváltozós időszori modelleken alapuló inflációs előrejelzések. Statisztikai Szemle, 77. 522-549.
 39. Martínez-Schnell, B.-Zaidi, A. (1989): Time series analysis of injures. Statistics in Medicine, 8. 1497-1508.
 40. Nelson, C. R.-Plosser, C. I. (1982): Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications. Journal of Monetary Economics, 10. 139-162.
 41. Odriozola, A. J.C.-Jimenez, D. J.-Rubio, M. J. C.-Perez, M. I. J.-Ortiz, P. M. S.-Rodrigues, R. P. (1998): Air pollution and mortality in Madrid. Spain: a time series analysis, Arch Occup Environ Health, 71. 543-549.
 42. Pierce, D. A. (1977): Relationship – and the lack thereof – between economic time series, with special reference to money and interest rates. Journal of the American Statistical Association, 72. 11-26.
 43. Ratkó, I.-Ruda, M. (1974): On an estimate for the parameter of a multidimensional stationary Gaussian Markov process, and an application. MTA SZTAKI Közlemények, 13. 21-30.
 44. Ratkó, I. (1984): Microcomputer in epidemiology. in: (Lecture Notes in Medical Informatics, Ed. Lindberg, D. A. B.-Reichert, P. L.) Springer-Verlag Berlin, 477.
 45. Ruda, M. (1974): Parameter estimation in the first order autoregressive process. Biometrika, 61. 632-633.
 46. Sugár A. (1991): Determinisztikus vagy sztochasztikus modell? Az időszori modellezés problémái. Disszertáció. BKTE, Budapest
 47. Terdik Gy. (1977): Az autoregressziós mező mozgó átlag előállítására és együtthatóinak maximum likelihood becslése. Alkalmazott Matematikai Lapok, 3. 379-388.
 48. Upshur, R. E.-Knight, K.-Goel, V. (1999): Time-series analysis of the relation between influenza virus and hospital admissions of the elderly in Ontario, Canada, for pneumonia, chronic lung disease, and congestive heart failure. Am J Epidemiol, 149. 85-92.
-