

Statisztikai Szemle

Közzététel: 2024. június 27.

A tanulmány címe:

A hólabdafolyamatról – még egyszer

Szerző:

HUNYADI LÁSZLÓ

a Statisztikai Szemle nyugalmazott főszerkesztője

E-mail: hunyadi44@gmail.com

DOI: <https://doi.org/10.20311/stat2024.06.hu0509>

Az alábbi feltételek érvényesek minden, a Központi Statisztikai Hivatal (a továbbiakban: KSH) *Statisztikai Szemle* c. folyóiratában (a továbbiakban: Folyóirat) megjelenő tanulmányra. Felhasználó a tanulmány vagy annak részei felhasználásával egyidejűleg tudomásul veszi a jelen dokumentumban foglalt felhasználási feltételeket, és azokat magára nézve kötelezőnek fogadja el. Tudomásul veszi, hogy a jelen feltételek megszegéséből eredő valamennyi kárért felelősséggel tartozik.

1. A jogszabályi tartalom kivételével a tanulmányok a szerzői jogról szóló 1999. évi LXXVI. törvény (Szt.) szerint szerzői műnek minősülnek. A szerzői jog jogosultja a KSH.
2. A KSH földrajzi és időbeli korlátozás nélküli, nem kizárólagos, nem átadható, térítésmentes felhasználási jogot biztosít a Felhasználó részére a tanulmány vonatkozásában.
3. A felhasználási jog keretében a Felhasználó jogosult a tanulmány:
 - a) oktatási és kutatási célú felhasználására (nyilvánosságra hozatalára és továbbítására a 4. pontban foglalt kivétellel) a Folyóirat és a szerző(k) feltüntetésével;
 - b) tartalmáról összefoglaló készítésére az írott és az elektronikus médiában a Folyóirat és a szerző(k) feltüntetésével;
 - c) részletének idézésére – az átvevő mű jellege és célja által indokolt terjedelemben és az eredetihez híven – a forrás, valamint az ott megjelölt szerző(k) megnevezésével.
4. A Felhasználó nem jogosult a tanulmány továbbértékesítésére, haszonszerzési célú felhasználására. Ez a korlátozás nem érinti a tanulmány felhasználásával előállított, de az Szt. szerint önálló szerzői műnek minősülő mű ilyen célú felhasználását.
5. A tanulmány átdolgozása, újra publikálása tilos.
6. A 3. a)–c) pontban foglaltak alapján a Folyóiratot és a szerző(ke)t az alábbiak szerint kell feltüntetni:
„*Forrás: Statisztikai Szemle* c. folyóirat 102. évfolyam 6. számában megjelent, **Hunyadi László** által írt, **A hólabdafolyamatról – még egyszer** című tanulmány (link csatolása)”
7. A Folyóiratban megjelenő tanulmányok kutatói véleményeket tükröznek, amelyek nem feltétlenül esnek egybe a KSH vagy a szerzők által képviselt intézmények hivatalos álláspontjával.

Hunyadi László

A hólabdafolyamatról – még egyszer On the Snowball process – once again

Hunyadi László, a Statisztikai Szemle nyugalmazott főszerkesztője
E-mail: hunyadi44@gmail.com

A „hólabda” egy terjedési típusú, nemlineáris autoregresszív folyamat. A hólabdafolyamat időmegoldása (ami bizonyos feltételek mellett explicit formában is előállítható) egy szaturációs függvény, a logisztikus vagy a Gompertz-görbéhez hasonlóan felső korlátos növekedési folyamatot ír le. A hólabdafolyamatot az 1970-es években írtam le először, ám akkor lényegileg szakmai visszhang nélkül maradt. Az azóta eltelt időben a számítástechnika gyors fejlődése lehetőséget teremtett analitikusan nehezen kezelhető feladatok, egyebek között a hólabdafolyamat néhány tulajdonságának vizsgálatára. A cikk ezekből mutat be egyet, de lényeges mondanivalója a hólabdafolyamat alkalmazhatóságának bemutatása három, az elmúlt években nagy figyelmet kapott társadalmi-gazdasági jelenség (a mobiltelefonok elterjedése, a Covid19-járvány lefutása és az utóbbi évek háborús inflációja) példáján.

Kulcsszavak: idősorlemzés, autoregresszív folyamatok, növekedési függvények, Covid, infláció

The Snowball is a diffusion-type, nonlinear autoregressive process. Its solution for the time variable (which is available only in certain cases) yields a growth function with saturation level, similar to the logistic, or the Gompertz functions. First time, the snowball process was described by me, in the 70-s, but this trial remained without any echoes of the model-builders. In the last 40 years, the rapid development of informatics made us possible to analyse problems that are difficult to handle analytically. So, the possibility to analyse some problems of the snowball process has been realized. The article gives an example of these, but its focus is on the application of the snowball model. Three different social-economic events (the diffusion of the mobile phones, the COVID, and the war-induced inflation of the last years) are examined here.

Keywords: time series analysis, autoregressive processes, growth functions, COVID, inflation

Valamikor 2020 körül, egy meleg júliusi napon Gáborral ültünk egy sör mellett Révfülöpön. Éppen hazafelé tartottunk a Káli-medencéből, egy félnapos kerékpártúráról. Beszélgettünk erről-arról, amikor egyszer csak megkérdeztem:

– Gábor, nem kellene könyvet írunk az idősorok elemzéséről?

Tudtam, mindkettőnknek van múltunk, vannak eredményeink ezen a területen.

Mire ő így felelt:

– De, írjunk!

Ennyi volt, aztán némán ittuk tovább a sört.

Így kezdődött, de utána nem folytatódott ennyire simán. Végül nehezen, de elkészült a könyv, és internetes változatban hozzáférhető lett minden érdeklődő számára. Sajnos már jórészt egyedül írtam meg el (*Hunyadi, 2022*). Ennek egy alfejezetét adom most közre kicsit átdolgozva, részben kibővítve.

1. A modellek

A hólabdafolyamatot és annak modelljeit már több mint 40 éve elkészítettem, és a *Sigma* című folyóiratban publikáltam (*Hunyadi, 1978*). Azt reméltem, hogy az akkor leírtak ösztönzők lehetnek mások számára, hogy folytassák ezt az utat, de csalódnom kellett. Még a szűkebb szakmában sem figyeltek fel igazán az ebben rejlő lehetőségekre. Később a *Statistikai Szemle* hasábjain tértem vissza – igaz, csak érintőlegesen – a hólabdára (*Hunyadi, 2004*), de a visszhang hasonló volt. Azóta annyit fejlődött a számítástudomány, hogy már könnyen és gyorsan működő R-programot készíthettem a hólabda számításaira, így néhány újabb alkalmazást készítettem (egyebek közt az elmúlt évek Covid19-járványának terjedésére, valamint a 2020 táján kezdődött és ma is tartó inflációra). A hólabdafolyamat egyes, eddig nem publikált (illetve csak az említett internetes megjelenésben közreadott) vonatkozásait, valamint néhány alkalmazást mutatok be ebben a rövid írásban.

A hólabdafolyamat feldolgozása során több, egymáshoz hasonló modellel dolgoztam. Ezek részletes leírása megtalálható az említett *Sigma*-cikkben, ezért itt csak nagyon rövid ismertetésekre térek ki. Az alapmodell abból a feltételezésből indul ki, hogy adott egy kitüntetett tulajdonság (pl. fertőzés), amellyel indulásként

s_0 számú egyed rendelkezik, és az egyedek száma a sokaságban véges. Feltételezem továbbá, hogy ezt a kitüntetett tulajdonságot az egyedek továbbadják, alapértelmezésben időegység alatt átlagosan egy további egyednek. Ha viszont olyan egyed kapja meg a kitüntetett tulajdonságot, aki/amely már rendelkezett vele (már fertőzött volt), akkor természetesen még egyszer nem fertőződik meg (a hír már nem új számára stb.). A feladat, hogy határozzuk meg azt a folyamatot, amely leírja, miként változik időben a kitüntetett tulajdonsággal rendelkezők (fertőzöttek) száma.

Tekintsük a következő jelöléseket: legyen N a sokaság (közeg) elemszáma, legyen n a terjedési sebesség (amelyet állandónak, sőt az alapmodellben 1-nek tekintünk), azaz az, hogy egy elem időegység alatt hány másiknak adja tovább a tulajdonságot (hányat fertőz meg), és jelöljük s_0 -al, hogy induló állapotban hány elem rendelkezett a vizsgált tulajdonsággal. Ekkor az említett mennyiségekre a következő természetes értékek, illetve korlátok adódnak:

$$N \text{ véges, } 0 < s_0 < N, \text{ és } n = 1.$$

Az utóbbi egyenlőséghez magyarázatként annyit teszek hozzá, hogy ez egyelőre csak egyszerűsítő feltételezés, de nem alaptalan, hiszen (egyelőre) úgy választhatjuk meg az időegységet, hogy ez a feltétel teljesüljön.

Az alapmodell kiinduló pontja az

$$s_1 = s_0 + g_0 \tag{1-1}$$

egyenlet, ahol g_0 azoknak az egyedeknek a száma, akik (vagy amelyek) kétszer (vagy többször) kerültek volna be a fertőzöttek közé. Mint látni fogjuk, ez a szám kritikus, hiszen ez adja meg a folyamatnak azt a korlátját, amelytől az eltér az exponenciális növekedésű folyamatoktól. Ekkor tehát kétféle elemet kell megkülönböztetnünk: olyanokat, amelyek rendelkeznek a kiemelt tulajdonsággal, ezek száma s_0 , és olyanokat, amelyek nem. Ez utóbbiak száma induláskor $N - s_0$.

Az első kérdés, hogy mi a valószínűsége annak, hogy a mintába *újonnan* bekerülők száma k , ahol $0 \leq k \leq s_0$. Mivel egyszerű binomiális eloszlásról van szó, g_0 eloszlása a következő:

$$\Pr(g_0 = k) = \binom{s_0}{k} \left(\frac{s_0}{N}\right)^k \left(\frac{N-s_0}{N}\right)^{s_0-k} \quad k = 0, 1, 2, \dots, s_0 \tag{1-2}$$

Ekkor g_0 és s_1 várható értéke:

$$E(g_0) = s_0 \frac{s_0}{N}, \text{ és } E(s_1) = 2s_0 - E(g_0) = 2s_0 - \frac{s_0^2}{N}.$$

Tovább lépve, és most már csak a folyamat változójára felírva:

$$s_{t+1} = 2s_t - \frac{s_t^2}{N},$$

vagy másképp felírva (differenciaegyenletként):

$$\Delta s_t = s_{t+1} - s_t = s_t - \frac{s_t^2}{N} = s_t \left(1 - \frac{s_t}{N}\right). \tag{1-3}$$

Az (1–3) differenciaegyenlet az alapmodell lényege, és mielőtt ennek elemzésébe belemennénk, nézzük meg a továbbfejlesztett modellt!

Ez a modell – mint már utaltam rá – csupán abban különbözik az előzőtől, hogy itt egy elem a vizsgált kitüntetett tulajdonságot nemcsak egy másiktól, hanem egyszerre akár többtől is megkaphatja. Ezáltal ez a modell egy lépésnek tekinthető a folytonos kezelés irányába, de nem vizsgálható annyira tisztán, mint az előző, és jóval több megfontolás, matematika is kell hozzá, amit itt mellőzök. Az esetleges érdeklődők a hivatkozott cikkben (Hunyadi, 1978) nézhetnek utána a részleteknek, itt csak a kiindulópontot és a végeredményt mutatom be.

A modell ezúttal is az

$$s_1 = s_0 + s_0 - \tilde{g}_0 = s_0 + \tilde{h}_0$$

egyenletből és további lépések után az

$$s_{t+1} = s_t + \tilde{h}_t \quad (1-4)$$

egyenletből indul, ahol tehát \tilde{h}_t azon elemek számát jelöli, amelyek a vizsgált tulajdonságot a $(t, t + 1)$ intervallumban megkapják. A cél most is az, hogy keressük \tilde{h}_t eloszlását, várható értékét, és végső soron a folyamat várható időbeli lefutását és annak jellemzőit. Eltekintve most a részletektől, a folyamat lényegét leíró rekurzió a következő lesz:

$$s_{t+1} = s_t + (N - s_t) \left[1 - \left(1 - \frac{1}{N} \right)^{s_t} \right] \quad (1-5)$$

Ez a forma némiképp bonyolultabb, mint az eddigiek, és nem is fogom széles körben használni, ám néhány tulajdonságára a későbbiek során még visszatérek.

Végül, harmadikként bemutatom az ún. *statisztikai modellt*, amely az egységnyi terjedési sebesség korlátozását oldja fel. Az alapmodell esetén eddig azzal az egyszerűsítő feltevessel éltem, hogy a terjedés sebességét (amit n -nel jelöltem) egységnek tekintetem. Egy elméleti modellben ez természetesen megtehető, de amint tényleges adatokat használunk fel, tarthatatlanná válik az a feltevés, hogy bármekkora a terjedési sebesség, ahhoz választjuk meg a modell pontjait, azaz az (egyelőre nem létező) megfigyeléseket. Abban a pillanatban azonban, hogy valós adatokra kívánjuk átvinni az elemzést, fel kell oldanunk ezt a korlátot, és az $n = 1$ helyett a $0 < n \leq 1$ feltevessel kell élnünk.

A részletek mellőzésével láthatjuk, hogy ebben az esetben a folyamat lefutását leíró egyenlet (és a folyamat differenciaegyenlete) a következő lesz:

$$s_{t+1} = (n + 1)s_t - \frac{ns_t^2}{N} \quad \text{és} \quad \Delta s_t = s_{t+1} - s_t = ns_t \left(1 - \frac{s_t}{N} \right) \quad (1-6)$$

Magát a hólabdafolyamatot e három változat köré építem úgy, hogy a modell elméleti tulajdonságait a legkönnyebben kezelhető és explicite is megoldható alapmodellen keresztül mutatom be, míg a gyakorlati számításokra a statisztikai modellt használom. A továbbfejlesztett modellt érdekességként ugyan bemutattam, de a későbbiek során ebben a tanulmányban csak hivatkozom rá, érdemben nem használom.

2. A modellek tulajdonságai

Az $s_{t+1} = 2s_t - \frac{s_t^2}{N}$ alapmodell olyan folyamatot ír le, amelynek legfontosabb tulajdonságai:

- s_t monoton növekvő,
- s_t felső korlátja N ,
- határértéke $t \rightarrow \infty$ esetén N ,
- ábrája egy inflexiós ponttal rendelkező S alakú növekedési görbe.

Ezek a tulajdonságok könnyen beláthatók, de a folyamat explicit megoldása egy kicsit bonyolultabb. A levezetést nem is mutatom meg – ezt az érdeklődők a többször hivatkozott *Sigma*-cikkből megtalálhatják. A végleges forma alkalmas formára rendezve:

$$s_t = N \left[1 - \left(1 - \frac{s_0}{N} \right)^{2^t} \right] \quad (2-1)$$

Ez a Gompertz-függvény egy speciális változata. Speciális azért, mert én ezt a függvényt csak diszkrét értékekre értelmeztem, holott az eredeti formájában folytonosként definiálják. Ezen túlmenően speciális azért is, mert itt az időváltozó 2 kitevőjeként értelmezett, noha általános Gompertz-függvény esetén ez tetszőleges pozitív és egynél nagyobb alapokra értelmezett. A (2-1) alapján a korábban említett tulajdonságok már könnyen beláthatók.

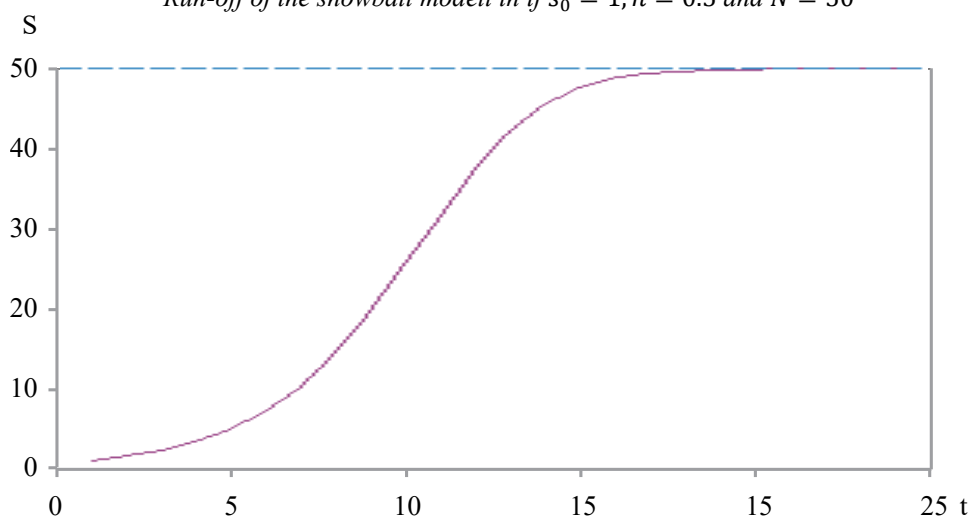
A továbbfejlesztett modellel kevesebbet foglalkoztam, de mivel az alapmodellhez hasonlóan viselkedik, az is hasonló, Gompertz-típusú növekedést mutat (*Kehl-Sipos, 2014*). A (2-1) forma igazi értelme számunkra most az, hogy magyaráz bizonyos idősorelemzésben hasznos függvényeket, és emellett alkalmas további, idősoros elemzésekre is.

A statisztikai modell explicit megoldását ez idáig nem sikerült előállítani, ám a modell számítógépen könnyen kezelhető,¹ és mint az 1. ábra is mutatja, az eddigiekhez hasonló alakú és tulajdonságú (lefutású függvényt) ad.

¹ A számításokat itt és a későbbiekben az R segítségével végeztem el. A számítások részleteit nem közlöm, de ha valaki érdeklődik ezek iránt, a szerző szívesen elküldi a megfelelő kódokat.

1. ábra

A hólabdamodell lefutása $s_0 = 1$, $n = 0.5$ és $N = 50$ esetén
Run-off of the snowball modell in if $s_0 = 1$, $n = 0.5$ and $N = 50$



Még néhány megjegyzés kívánkozik az elmondottakhoz, a hólabdafolyamat eddig bemutatott modelljei kapcsán:

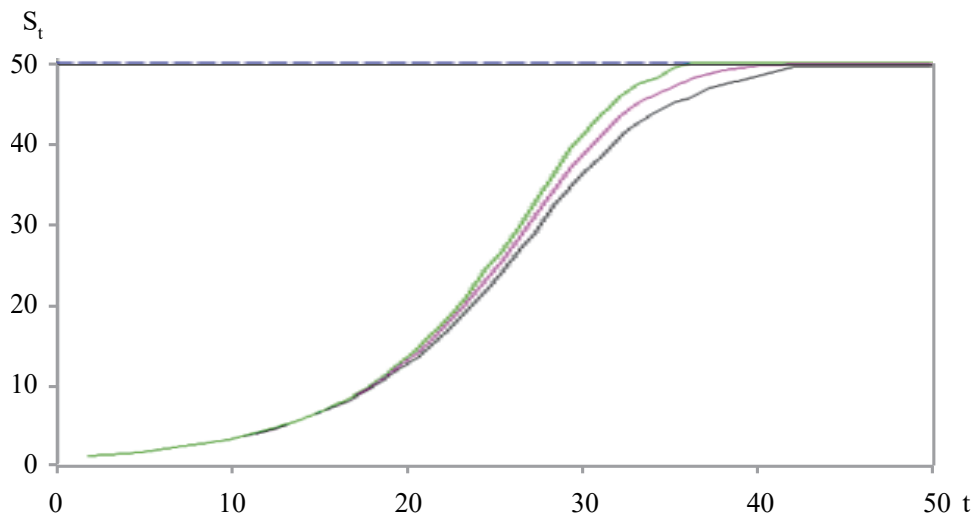
- A hólabdamodellek hasonlítanak ismert trendfüggvényekre. Ezt az állítást korábban részletesen is kifejtettem, lényege, hogy a hólabdamodellek szoros rokonságot tartanak a korlátos növekedési függvényekkel, mindenekelőtt a logisztikus függvénnyel, valamint a Gompertz-függvénnyel.
- A hólabdamodellek felfoghatók nemlineáris AR-modelleknek is, hiszen pl. az (1– 6) formula egyértelműen mutatja az elsőfokú és a másodfokú egykésleltetéses autoregresszív tagokat. Ebben az értelemben ez a modell típus az AR-, vagy később az ARMA-modellecsalád egy új szintű első képviselője lehet.
- Ezek a modellek alkalmasak konfidenciasáv készítésére, hiszen a folyamat minden t esetében egy várható érték, mégpedig egy ismert eloszlás várható értéke. A konfidenciasávokat explicit formában nem sikerült (és alighanem lehetetlen) felírni, de az összes ponthoz tartozó eloszlások ismeretében számítógéppel meglehetősen könnyű őket szimulálni. A következőben bemutatom a statisztikai modell futását, szimulált konfidenciasávokkal. A felhasznált program a megadott paraméterek (itt $s_0 = 1$, $n = 0,15$ és $N = 50$) esetében kiszámítja és kirajzolja a 90%-os konfidenciasávokat, és tetszés szerinti ponthoz ki is írja annak értékeit. Megjegyzem, hogy ez a program a

korábban bemutatott binomiális eloszlásból indul ki, azt sokszor (a példában 1000-szer, de ennek sokszorosa sem jelent problémát) lefuttatva, annak empirikus várható értékét tekinti a folyamat közepének, és ahhoz számítja ki a konfidencia-intervallumokat.

Például a $t = 35$ ponthoz a 90%-os konfidencia-intervallum határai (47,05 várható érték mellett) 45,03 és 49,71. Megjegyzendő, hogy ezt az intervallumot az teszi lehetővé, hogy határozott modellfeltevésekből indultam ki (terjedés leírása). A 2. ábrában ábrázolt lefutás az említett hólabdafutást és 90%-os konfidenciasáv-jait mutatja.

2. ábra

A statisztikai modell 90%-os konfidenciasávja
($s_0 = 1$, $n = 0.15$ és $N = 50$)
90 percent confidence-bands of the snowball model



- Ezek a modellek (még a statisztikai modell is) így csak a vázat adják meg – hasonlóan az idősorelemzés alkalmazott trendfüggvényekhez. Hiába a háttéreloszlás, csak egy (vagy esetleg több) véletlen tag hozzáadása teremt meg a lehetőséget arra, hogy használjuk őket.

Mіндеzen tulajdonság ismeretében most rátérek a hólabdamodellek alkalmazására.

3. Alkalmazások

A hólabdafolyamat és modelljei – bár nemlineáris autoregresszív formát mutatnak – viszonylag könnyen és kényelmesen kezelhetők. A paraméterek becsléséhez az (1–6) formából indulok ki. Ekkor az

$$y_t = s_{t+1}, \quad x_1 = y_{t-1}, \quad \text{és} \quad x_2 = y_{t-1}^2$$

helyettesítéseket elvégezve és a paramétereket

$$\beta_1 = n + 1, \quad \text{és} \quad \beta_2 = -\frac{n}{N} \quad \text{ként} \quad (3-1)$$

értelmezve azonnal adódik a megvalósítható lineáris regresszió

$$y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

alakban. Ebből a paramétereket és a paraméterek „mögötti” jellemzőket egyszerűen lehet legkisebb négyzetek módszerével becsülni.

Az alkalmazásokat három valós feladaton keresztül mutatom be. Az első a magyarországi (vonalas) telefon-előfizetők számának alakulása. Kiindulópontom a telefon-előfizetők száma volt 1962 és 2000 között, a KSH adatai szerint.

Ebben az időszakban a (vonalas) telefon viszonylag gyorsan elterjedt, és úgy tűnt, hogy a telítődésig (ami a magyarországi lakásszámokon kívül a munkahelyi potenciális telefonokkal együtt szakértői becslések alapján valahol 10 millió körül lehetett volna) a folyamat szépen kirajzolja a vizsgált görbét. Megjelent azonban a mobiltelefon, és a trend 2000 körül, illetve utána megszakadt. Ezért a példa a biztatóan induló telefonidősornak csak az elejét tudja megragadni.

A számítások eredményeit az 1. táblázat mutatja, rajzolatát pedig a 3. ábra adja meg.

A becslések alapján a folyamat ábrája bizonyítja a táblázatban megadott jó illeszkedést, és megerősíti a becsült t értékek sugallta helyes, jól illeszkedő modellt. Ugyanakkor azt is jelzi, hogy a folyamat még nagyjából a közepén járhatott (a telítődést illetően), amikor megszakadt, ami, mint említettem, a mobilok megjelenésének és gyors elterjedésének tudható be.

1. táblázat

**A telefon-előfizetők számának alakulására készített
hólabdamodell becslési eredményei**

Estimation results of the snowball model for the numbers of telephone subscribers

	Paraméter	Standard hiba	t-érték
x_1	1,145	0,0287	39,26
x_2	-0,000015	0,0000104	-1,50
$R^2 = 0,9963$	korrigált $R^2 = 0,9961$		
$F = 5146$	szabadságfok = 2,38	$p = 0,0000$	

A becsült regressziós együtthatókból (3–1) alapján kiszámíthatók a hólabda-modell paraméterei

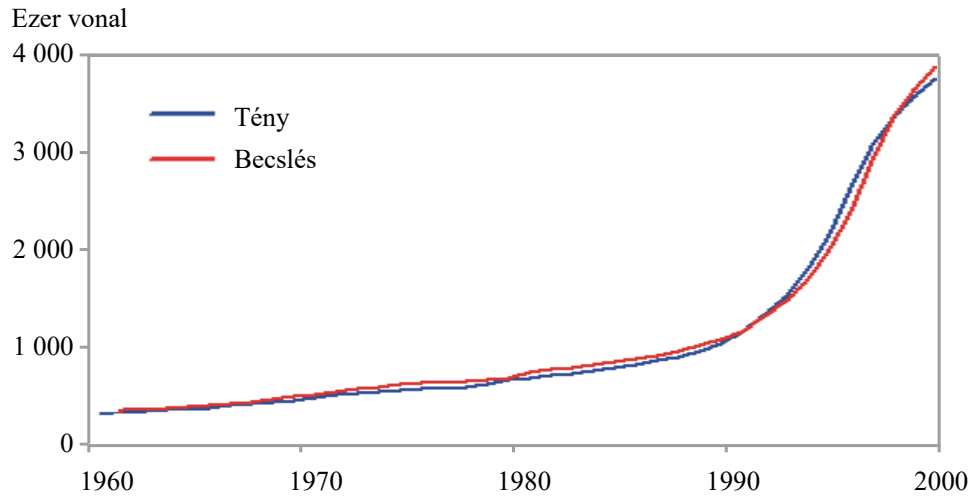
$$n = \beta_1 - 1 = 1,144 - 1 = 0,144 \text{ és } N = \frac{\beta_1 - 1}{-\beta_2} = \frac{0,144}{0,000015} = 9332,47.$$

Ez annyit jelent, hogy a terjedés sebessége 0,144 ennél a folyamatnál, a telítődési szint pedig 9332 ezer vonal, ami nagyjából egybevág az előzetes szakértői becslésekkel.

3. ábra

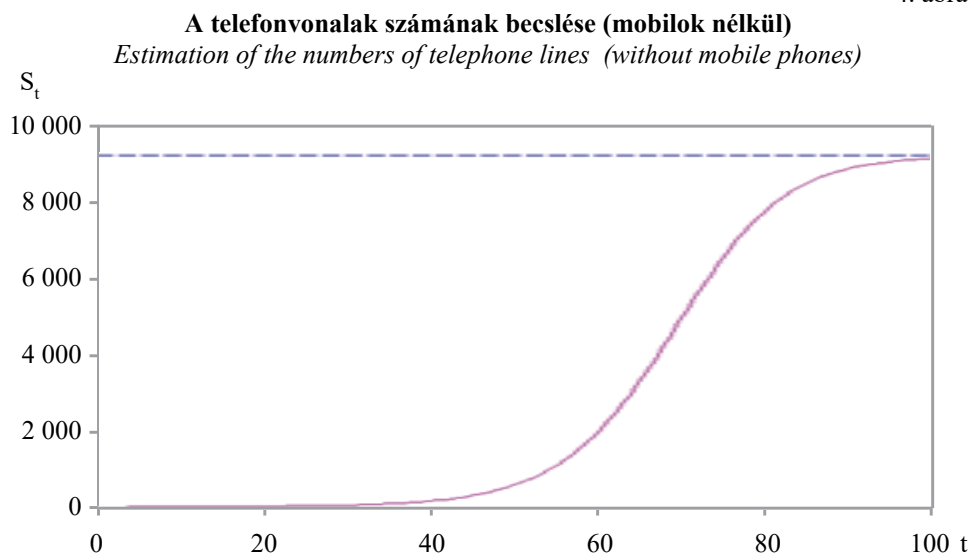
A telefonvonalak száma és azok becslése hólabdafüggvénnyel

The numbers of telephone lines and their estimation by the snowball function



A példa számára nem lényeges ugyan, de érdekességképp a 4. ábrán bemutatom ennek a folyamatnak a teljes (feltételezett) lefutását. Mint látható, ez egy szép korlátos növekedési folyamat, ugyanakkor ez a példa figyelmeztet a hólabdafolyamat modellezésének korlátjaira.

4. ábra



Második példám, a terjedési folyamat még inkább meggyőző példáját a világon (és természetesen Magyarországon is) 2020-ban végigsöprő (és a következő években ismétlődő) Covid19-járvány terjedése adja. A téma különböző vonatkozásaival természetesen részletesen foglalkozott a hazai statisztikai szakirodalom (pl. Hortai–Szóke, 2020; Kemény et al., 2022; Dudás–Szántó, 2022), de érdekes módon a szakma figyelmét az idősor jellege, kezelhetősége, előrejelzése nem igazán ragadta meg. A folyamat természetesen lényegesen összetettebb, bonyolultabb annál, minthogy egy ilyen szűk alfejezetben részletesen elemezni tudnánk, de élelítőképpen egy kis szeletét, a magyarországi terjedés első szakaszát és annak modellezését az alábbiakban bemutatom.

Első szakasznak a járvány hazai kezdetétől, azaz 2020. március 4-től a pandémia első fázisának lecsengéséig, azaz 2020. június 10-ig tartó időszakot tekintem. Az adatokat a John Hopkins Egyetem honlapjáról (coronavirus.jhu.edu) töltöttem le. A modellt, amelyet ezekre az adatokra illesztettem, az (1–6) statisztikai modell, paraméterbecslését a (3–1) módon végeztem el. Az eljárásra külön R-kódot készítettem, amely a becslés, a tényadatok, a becslések és előrejelzések mellett azt is megmutatja, hogy a modell szerint mikor ér véget a járvány, azaz mikor jutunk el oda, hogy az új fertőzöttek száma legfeljebb k legyen. (Ezt az értéket a futtatások során $k = 1$ értékre rögzítettem.)

Az eredmények azt mutatták, hogy a hólabdamodell – bár egy-egy kiugró, a folyamat menetétől lényegesen eltérő inputadat (*outlier*) meg tudja zavarni – alapvetően jól teljesített. Mivel a programot úgy terveztem, hogy az új és a még újabb

adatok beérkeztével mindennap új becslést készítettem, naponta (2020. április óta) egy-egy újabb futási eredményhez jutottam. A hólabdafolyamat modellje esetében 2 napot leszámítva (amikor kiugró számú új fertőzöttet regisztráltak, mivel egy idősoththonba betört a vírus, és a rá különösen fogékony idősök körében gyorsan elterjedt) a naponta becsült paraméterek többnyire egyre kisebb eltérést mutattak, és ha nem is teljes mértékben, de tendenciájukban stabilizálódni látszottak. Ennek igazolására mutatok be egy futást abból az időszakból, amikor a járvány első szakasza éppen túljutott a legmeredekebben növekvő szakaszán (2. táblázat, 5. ábra).

2. táblázat

Az új típusú koronavírus terjedése az első 55 napon
Diffusion of COVID in Hungary in the first 55 days

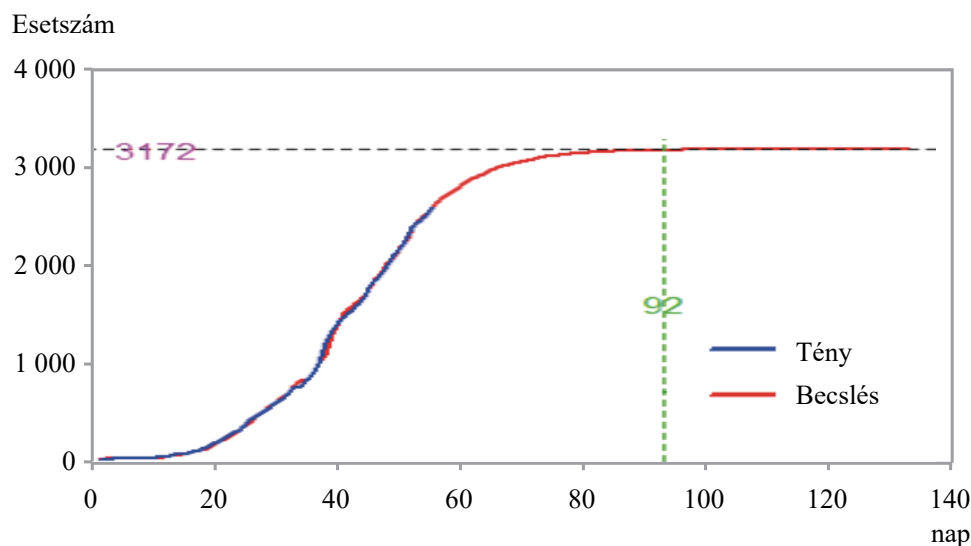
Készítés napja: 55.		Készítés dátuma: 2020. április 27.	
Becsült paraméterek			
$n = 0,1202$	$N = 3172$	$R^2 = 0,9991$	
Előrejelzések			
04.28.	2 686		
04.29.	2 735		
04.30.	2 781		
Utolsó nap: 92.		Utolsó nap dátuma: 2020. június 2.	

Az eredmények eléggé meggyőzőek. Látható, hogy a becsült görbe jól követi a tényadatokat, a várható lefutás megegyezik az elméleti modellel. Kiemelendő, hogy a többszörös determinációs együttható ennek megfelelően kivételesen magas, és mindez úgy valósul meg, hogy egy nagyon egyszerű feltételrendszerből indultam ki, és ennek megfelelően egyszerű modellt konstruáltam.

Természetesen ennek is megvannak a hátulütői. Ezek a későbbi futások (amelyeket itt helyhiány miatt nem mutatok be részletesen) során meg is jelentek. A futások voltaképpen jó eredményekre utalnak, hiszen az illeszkedés ezúttal is jó, sőt a determinációs együttható, ha lehet, még szorosabb illeszkedést jelez. A probléma az, hogy a folyamatot leíró paraméterek valójában nem stabilizálódnak kellő módon. A terjedési sebesség mutatója lassan, de fokozatosan és folyamatosan csökken, de ami talán még ennél is elgondolkodtatóbb, a folyamat felső korlátja nem igazán stabil, minden egyes nappal egy kicsit feljebb tolódik, holott azt várnánk, hogy egy szint körül ingadozik. A különbség persze nem nagy, de mindenesetre figyelmeztet arra, hogy még nem tökéletes a modell, illetve nem jutott el a stabilizálódásig a folyamat.

5. ábra

Az új típusú koronavírus terjedése Magyarországon az első 55 napon
Diffusion of COVID in Hungary in the first 55 days



Ezek után még egy futás eredményét mutatom meg, a folyamat végét, ami a 90. nap, azaz június 1. Ez már a folyamat első szakaszának abban a stádiumában jött létre, amikor közel voltunk ahhoz a szinthez, amelyet az első fázis lezárásának tekinthetünk. A 3. táblázatból és a 6. ábrán jól látható, hogy a korábban jelzett tendencia nem szakadt meg, azaz a terjedési sebesség tovább csökkent, a felső korlát, ha lassan is, de tovább nőtt, tehát a folyamat valójában nem stabilizálódott egészen. Mindazonáltal az első fázis megszűnését – amelyet a modell a 100. napra, azaz június 10-ére becsült – nagyjából jól jelezte előre, hiszen körülbelül ekkor hirdette meg a válságot figyelő és szakmailag kézben tartó testület az első szakasz végét. Igaz, ez még a napi 1 új fertőzést megelőzően volt, de a modell működését, jó tájékoztatóképességét ez nem kérdőjelezi meg.

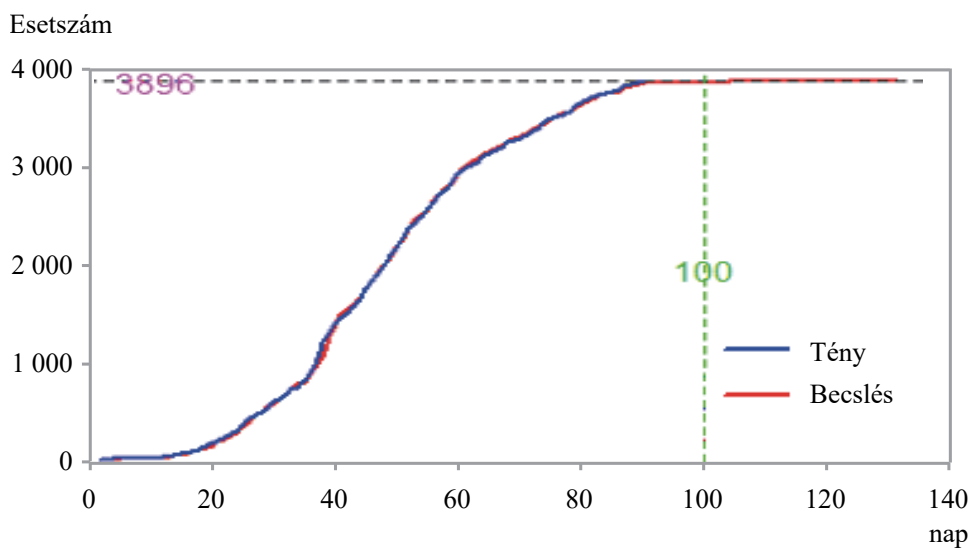
3. táblázat

Az új típusú koronavírus terjedése az első 90 napon
Diffusion of COVID in Hungary in the first 90 days

Készítés napja: 90.		Készítés dátuma: 2020. június 1.	
Becsült paraméterek			
$n = 0,0913$	$N = 3\,896$	$R^2 = 0,9998$	
Előrejelzések			
06.02.	3 876		
06.03.	3 878		
06.04.	3 879		
Utolsó nap: 100.		Utolsó nap dátuma: 2020. június 10.	

6. ábra

Az új típusú koronavírus terjedése Magyarországon az első 90 napon
Diffusion of COVID in Hungary in the first 90 days



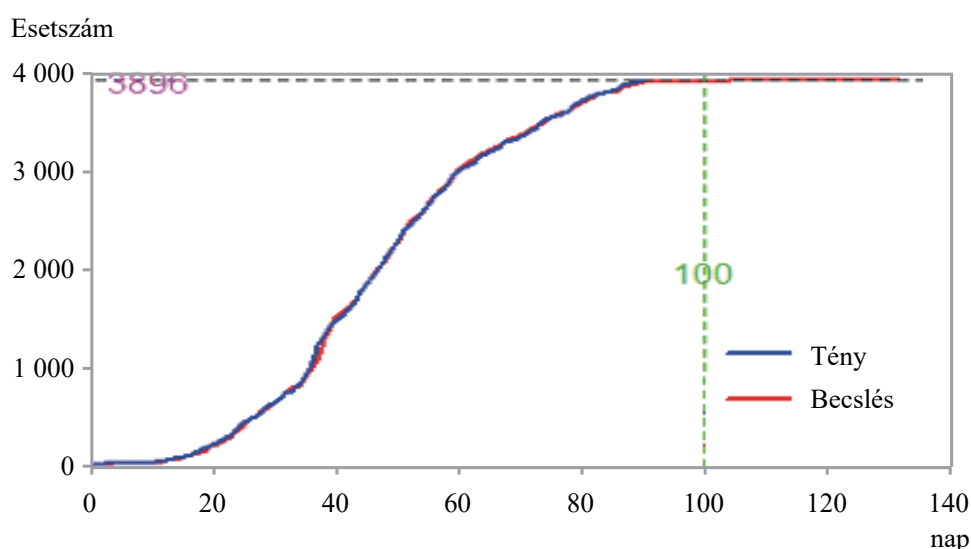
Csak érdekességképp megmutatom, hogy ezekre az adatokra a logisztikus modellel készített utolsó futás a 7. ábrán látható eredményt adta. Bár a végső eredmények némiképp hasonlóak az eddigiekhez, ez az ábra jól mutatja, hogy a folyamat lefutása lényegesen eltér attól, amit a logisztikus közelítés adott volna. Azt is jól jelzi, hogy bár a telítődési szinttel rendelkező folyamatok sok szempontból rokonságot mutatnak egymással, a gyakorlati esetekben olykor nagyon nehéz különbséget tenni köztük, kiválasztani a mindenkor legalkalmasabbat, az alkalmazások so-

rán érdemes több oldalról megfontolni, megvizsgálni, kipróbálni, hogy melyik alkalmazható leginkább, mert akkor elkerülhetők azok a buktatók, amelyek a 7. ábrán jól láthatók.

7. ábra

**Az új típusú koronavírus terjedése Magyarországon az első 90 napon
(közelítés logisztikus függvénnyel)**

*Diffusion of COVID in Hungary in the first 90 days
(approximation by a logistic function)*



Harmadik példám a 2020-as évek elején lezajló magyarországi inflációs folyamat modellezése. Természetesen az infláció – és kiváltképp a mostani infláció – felkeltette a szakma figyelmét (pl. *Bareith–Fertő, 2023*), de idősoros közelítéssel, leírásával, előrejelzésével a hazai szakirodalomban nem találkoztam. Jelen írás szellemével megegyezően ezúttal sem a jelenség mélyebb összefüggéseire, közgazdasági tartalmára, okaira és következményeire fordítom a fő figyelmet, „csupán” azt vizsgálom – feltételezve a terjedési folyamatot, hogy az infláció bizonyos értelemben hasonlóan terjed, halad előre, mint ahogy azt az eddig vizsgált egyéb jelenségek tették –, hogyan lehet az infláció lefutását hólabdával modellezni.

Az adatokat (2020. október = 1,00 bázisú bázisindexeket a fogyasztói árak változására) a KSH honlapjáról vettem. A regressziós becslés eredményei a 4. táblázatban láthatók.

4. táblázat

Az inflációs folyamat paramétereinek becslése
Estimation of the parameters of the inflation

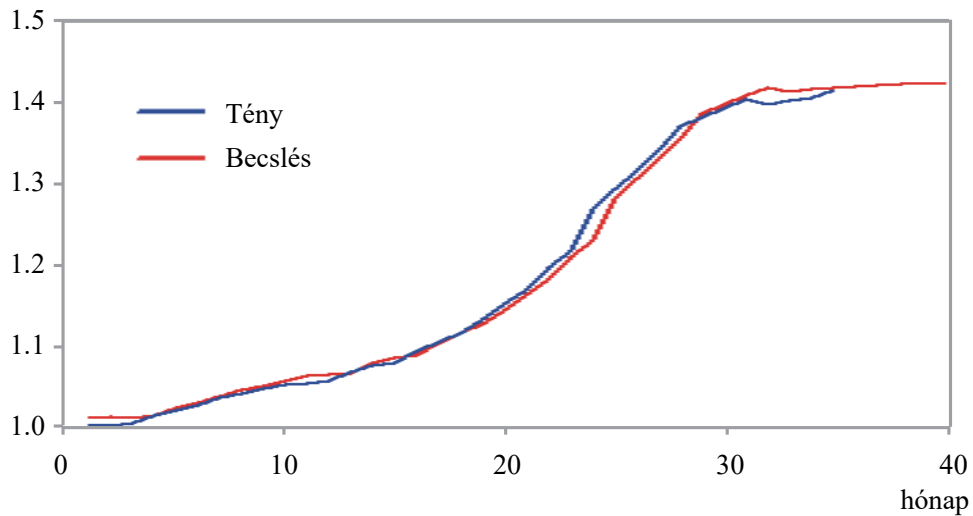
	Paraméter	Standard hiba	t érték
x_1	1,006040	0,012268	82,05
x_2	0,003618	0,010129	0,357
$R^2 = 0,9999$	korrigált $R^2 = 0,9999$		
$F = 213700$	szabadság fok = 2,33	$p = 0,0000$	

Innen a (3–1) alapján azonnal kaphatók a hólabdaparaméterek: $n = 0,006$ és $N = 1,6694$.

8. ábra

Az infláció lefutása és a hólabda közelítése
Run-off of the inflation and the snowball approximation

2020. október = 1,0



A becslésekkel kapcsolatban meg kell jegyezni, hogy az eredeti (lineáris) regresszióban az x_2 együttható gyenge t értékkel rendelkezik, ezért becslése bizonytalan. Mivel a felső korlát N paramétere ennek függvénye, természetesen ez is felettébb bizonytalan, ezért ezt ki se rajzoltam. Ugyanakkor a 8. ábra – összhangban a magas korrelációval – azt jelzi, hogy az illeszkedés jó, a modell jól leírja a folyamatot.

Az előrejelzések a következőket mutatják (megjegyzendő, hogy 2023 októberében készültek).

5. tábla

Az infláció előrejelzése
Forecasting the inflation

Év, hónap	Fogyasztóiár-index (előző év azonos időszaka = 1,00)
2023. október	1,155
2023. november	1,110
2023. december	1,089
2024. január	1,071
2024. február	1,051

Az előrejelzések meglehetősen reálisnak tűnnek, és egybevágóak azokkal a központi előrejelzésekkel, amelyek az év végére egy számjegyű (10%-nál alacsonyabb) inflációt várnak. Ez megerősíti azt, hogy ez a meglehetősen egyszerű, és igen kevés külső információt igénybe vevő modell ebben az esetben is jól teljesít.²

Ennek kapcsán egy gondolatot szeretnék még megemlíteni. Az összes eddig lezajlott vizsgálat (azok is, amelyeket ezzel a modellel eddig, más feladatok kapcsán végeztem, de itt nem mutatom be őket) azt jelezte, hogy valamilyen szinten értékelhető, használható a hólabdafolyamat modellje. Természetesen ezek a jelenségek (és itt elég a telefonellátásra, a Covidra vagy éppen az inflációra utalni) jóval összetettebbek annál, mintsem két változóval és egy ügyesen választott formulával alaposan ki lehessen elemezni őket. Ráadásul maga a terjedési folyamat, amely a modell mögött áll, sohasem érvényesül tisztán, azt kívülről mindig megpróbálják befolyásolni (pl. a járványok esetén védőoltásokkal, vagy az infláció esetén központi árintézkedésekkel stb.). Ezt nyilvánvalón mindenki tudja. Ugyanakkor az itt bemutatott eredmények azt sugallják, hogy első eszközként, egyetlen gyorsan és kényelmesen alkalmazható módszerként az ilyen jellegű folyamatok esetén célszerű a hólabdafolyamat modelljét használni. Más szóval, a „hólabda” nemcsak le fogja írni ezeket az egymástól meglehetősen távol álló, mégis bizonyos rokonságot mutató folyamatokat, de segíthet megoldani is az ilyen problémákat. A gondos elemzés természetesen ezekben az esetekben is több oldalú megközelítést, több módszer összevetését, konfrontáltatását jelent. Ezek egyike lehet, sőt kívánatos lehet a „hólabda”.

² A cikk megírása óta megjött a KSH legújabb, novemberi jelentése az inflációról, miszerint a lassulás valamivel gyorsabb a tervezettnél és az itt számítotttnál. Ez persze sok mindennel magyarázható, sőt az új adat birtokában új számítások is készíthetők, de erre a rendelkezésre álló idő már nem elegendő.

4. Összegzés

Befejezésül annyit kell még a témához hozzátenni, hogy az az egyszerű gondolat, amely mögötte van – nevezetesen a telítődési folyamatok lényegének, terjedésének, diffúziójának megragadása –, roppant hasznos lehet az idősorok elemzésének más területein is, bár azt is le kell szögezni, hogy ez a fajta gondolkodás ezen a területen meglehetősen ritka. Ezért – bár az említett tanulmány, amelyre ezt az előző alfejezetet alapoztam – már több mint 40 éves, folytatásról, hasonló szellemű modellekről sajnos nem tudok, pedig az idősorok elemzésénél nagy szükség lenne rájuk.

Irodalom

- Bareith T. – Fertő I. (2023): Stabilizálhatja-e a monetáris politika az élelmiszer-inflációt? *Statisztikai Szemle*, 101(4), 354–380.
- Box, G. E. P. – Jenkins, G. M. (1976): *Time Series Analysis, Forecasting and Control*. Holden Day, San Francisco.
- Dudás L. – Szántó R. (2022): Mi befolyásolja a járványintézkedések támogatottságát? *Statisztikai Szemle*, 100(5), 49–513.
- Éltető Ö. – Hunyadi L. (1972): *On the Estimation of the Parameters of the Logistic Function*. Paper Presented at the European Meeting of the Econometric Society, Budapest.
- Harvey, A. C. (1981): *Time Series Models*. Philip Allan P. L, Oxford.
- Hortay O. – Szőke T. (2020): A kijárási korlátozás hatása a villamosenergia-rendszer terhelési és árgörbéire Magyarországon. *Statisztikai Szemle*, 98(10), 1131–1150.
- Hunyadi L. (1978): Egy terjedési folyamat elemzése (a hólabdamodell). *Sigma*, XI. 191–209.
- Hunyadi L. (2004): A logisztikus függvény és a logisztikus eloszlás. *Statisztikai Szemle*, 82(10–11), 991–1011.
- Hunyadi L. (2022): *Gazdasági idősorok elemzése* (kézirat).
- Kehl D. – Sipos B. (2014): A telítődési, a logisztikus és az életgörbe alakú trendfüggvények becslése Excel parancsfájl segítségével. *Statisztikai Szemle*, 87(4), 381–411.
- Kemény I. – Kulhavi N. M. – Kun Zs. (2022): A távorvoslás igénybevételét befolyásoló tényezők a Covid19-járvány miatti félelem tükrében. *Statisztikai Szemle*, 100(1), 7–43.