

EÖTVÖS LORÁND TUDOMÁNYEGYETEM

BUDAPESTI CORVINUS EGYETEM

Kelemen Viktória

A SAINT halandóság-előrejelző modell életbiztosítási alkalmazása

Szakdolgozat

Témavezető:

Dr. Vékás Péter

Biztosítási és pénzügyi matematika mesterszak

Aktuáriusi specializáció



Budapest, 2024.

Köszönetnyilvánítás

Szeretném kifejezni mély hálámat Vékás Péternek, aki vállalta a konzulensem szerepét, és értékes szakmai tanácsokkal segítette a dolgozatom létrejöttét. Külön köszönet illeti családomat, barátaimat és szaktársaimat, akik végig támogattak és inspiráló környezetet biztosítottak számomra az egyetemi évek alatt, így téve ezt az időszakot igazán emlékezetessé.

Tartalomjegyzék

1. Bevezetés	8
Szakdolgozatom felépítése.....	9
2. A mortalitás modellezés módszertani alapjai	10
Folytonos modellezés.....	10
Diszkrét modellezés.....	11
Becslés diszkrét modellből.....	13
3. Szakirodalmi áttekintés	15
Lee-Carter modell és hiányosságai.....	15
Referencia populáció.....	19
SAINT modell.....	20
4. A felhasznált halandóság előrejelző modellek	22
Lee-Carter modell.....	22
A modell egyenlete.....	22
Paraméterek értelmezése.....	23
Paraméterek becslése.....	24
Előrejelzés.....	25
SAINT modell.....	26
5. Numerikus elemzés	28
Adatbemutatás.....	28
Egypopulációs modellezés.....	30
Referencia populáció Lee-Carter modelljei.....	30
Járadékosi populáció Lee-Carter modelljei.....	35
SAINT modellezés és eredményei.....	37
USA: férfi adatok eredménye.....	39
USA: női adatok eredménye.....	40
Svájc: férfi adatok eredménye.....	41
Svájc: női adatok eredménye.....	42
Modellek alkalmazása díjszámításban.....	43

6.	<i>Összegzés</i>	46
7.	<i>Hivatkozások</i>	48
8.	<i>Melléklet</i>	51

Táblázatjegyzék

1. Táblázat MSE értékek a valós és Lee-Carter modellel becsült értékek között a referencia populációra	33
2. táblázat MSE értékek a valós és Lee-Carter modellel becsült értékek között a biztosított populációra	35
3. Táblázat Alkalmazott ARMA folyamatok	38
4. táblázat A két modell MSE értékei az amerikai férfiakra	39
5. táblázat A két modell MSE értékei az amerikai nőkre	40
6. táblázat A két modell MSE értékei a svájci férfiakra.....	41
7. táblázat A két modell MSE értékei a svájci nőkre	42
8. táblázat Járadék díjak periódus táblából.....	44
9. táblázat Járadék díjak kohorsz táblából.....	44
.....	

Ábrajegyzék

1. ábra A 2020-as magyar nők halandósági táblája 23 éves korig. (Forrás: Human Mortality Database).....	12
2. ábra Az átlagos logaritmikus halandóság ábrázolása az 1980-2010-es adatokkal Magyarországon (saját szerkesztés).....	23
3. ábra A mortalitási index alakulása az 1980-2010-es magyarországi adatok felhasználásával (saját szerkesztés).....	24
4. ábra Az életkorfüggő érzékenység alakulása Magyarországon az 1980-210-es adatok szerint (saját szerkesztés).....	24
5. ábra Uniszex log-halandósági ráták (bal) és uniszex halálozási számok (jobb) alakulása 1959-1991 között az Amerikai Egyesült Államokban évente (saját szerkesztés).....	29
6. ábra Uniszex log-halandósági ráták (bal) és uniszex halálozási számok (jobb) alakulása 1945-1995 között Svájcban évente (saját szerkesztés).....	30
7. ábra Egyesült Államok férfi (bal) és Svájc férfi (jobb) Lee-Carter modell reziduálisainak hőtésképe.....	31
8. ábra 1959-1991 között a férfiak Lee-Carter modelljének paraméterei 60-100 éves kor között USA-ban.....	31
9. ábra 1959-1991 között a nők Lee-Carter modelljének paraméterei 60-100 éves kor között USA-ban.....	31
10. ábra 1945-1995 között a férfiak Lee-Carter modelljének paraméterei 60-100 éves kor között Svájcban.....	32
11. ábra 1945-1995 között a nők Lee-Carter modelljének paraméterei 60-100 éves kor között Svájcban.....	32
12. ábra Valós és LC modell becsült értékeinek eltérése a referencia USA női adatokon (saját szerkesztés).....	33
13. ábra Valós és LC modell becsült értékeinek eltérése a referencia USA férfi adatokon (saját szerkesztés).....	34
14. ábra Valós és LC modell becsült értékeinek eltérése a referencia Svájc férfi adatokon (saját szerkesztés).....	34
15. ábra Valós és LC modell becsült értékeinek eltérése a referencia Svájc női adatokon (saját szerkesztés).....	34
16. ábra Egyesült Államok férfi (bal) és Svájci férfiak(jobb) biztosított adatainak Lee-Carter modell reziduálisainak hőtésképe (saját szerkesztés).....	35
17. ábra Valós és LC modell becsült értékeinek eltérése a biztosított USA férfi adatokon (saját szerkesztés).....	36

18. ábra Valós és LC modell becsült értékeinek eltérése a biztosított USA női adatokon (saját szerkesztés).....	36
19. ábra Valós és LC modell becsült értékeinek eltérése a biztosított Svájc női adatokon (saját szerkesztés).....	36
20. ábra Valós és LC modell becsült értékeinek eltérése a biztosított Svájc férfi adatokon (saját szerkesztés).....	37
21. ábra Optimalizált paraméterek az amerikai férfiak (bal) és nők (jobb) adataihoz 60 éven felettiekre (saját szerkesztés).....	37
22. ábra Optimalizált paraméterek a svájci férfiak (bal) és nők (jobb) adataihoz 60 éven felettiekre (saját szerkesztés).....	38
23. ábra Amerikai férfi adatokra modellenként az átlagos eltérés 60-100 éves korban (bal) és 60-90 éves korban (jobb) (saját szerkesztés).....	39
24. ábra Amerikai női adatokra modellenként az átlagos eltérés 60-100 éves korban (saját szerkesztés).....	40
25. ábra Svájci férfi adatokra modellenként az átlagos eltérés 60-100 éves korban (bal) és 60-95 éves korban (jobb)(saját szerkesztés).....	41
26. ábra Svájci női adatokra modellenként az átlagos eltérés 60-100 éves korban (bal) és 75-100 éves korban (jobb) (saját szerkesztés).....	42
27. ábra Amerikai férfi adatokra modellenként az átlagos eltérés 10-100 éves korban (saját szerkesztés).....	46
28. ábra Svájci férfi adatokra modellenként az átlagos eltérés 20-100 éves korban (saját szerkesztés).....	47
29. ábra Svájci férfi adatokra modellenként az átlagos eltérés 45-90 éves korban (saját szerkesztés).....	47

1. Bevezetés

Az emberiség történelmében mindig is jelen volt a vágy, hogy kiélvezzük a jelen pillanatait, akárcsak ahogy a latin szállóigévé vált mondás szerint: *Carpe diem!* Vagyis: Élj a mának! Ragadd meg a napot! Az egyének életüket a mának élve igyekeznek élményekkel gazdagítani, kapcsolatokkal és boldogsággal. Ha nem is szeretünk róla beszélni, rágondolni, akkor is el kell fogadnunk, hogy mindannyiunk élete véges, így keressük a minél kielégítőbb válaszokat arra, hogyan és meddig is élhetünk boldog, autentikus életet, akárcsak Tolsztoj hőse:

„Kereste régi, megszokott halálfélelmét, és nem lelte sehol.

Hová lett? Miféle halál? Nem volt sehol a félelem, mert a halál sem volt sehol.”

Tolsztoj: Ivan Iljics halála (1886)

A mára koncentráció személyes életfilozófiája ellenére számos olyan tényező és döntés jellemzi életünket, amelyek a jövőre irányulnak, és befolyásolják, hogy meddig élünk, és milyen körülmények között. Hosszú távú tervek, hitelek, elköteleződések, mint a házasság és a több éves tanulmányok mind olyan elemek, amelyek a *carpe diem* elveivel látszólag szembemennek. Az egyén szintjén ezek inkább kósza gondolatoknak tűnhetnek, miközben a mindennapi életben próbáljuk élvezni az elröppenő pillanatokot. Ugyanakkor társadalmi szinten ezek a döntések rendkívül fontosak, mivel nagy hatással vannak akár adott országok működésére, nyugdíjrendszerére, vagy a biztosítók működésére. A gazdaságok fejlődése, az egészségügy és az életmódunk, étkezéseink minőségének emelkedése mind hozzájárulnak ahhoz, hogy egyre magasabb várható élettartammal kell számolhatunk. Ez a pozitív változás azonban kockázatokkal is jár, amit a szakirodalomban élettartam kockázat (angolul *longevity risk*) néven emlegetnek. Egyre fontosabb meggondolnunk, hogyan gondoskodhatunk saját magunkról (vagy az államról) idősebb korunkban. Ennek egyik módja lehet a járadékbiztosítások kötése, így a biztosítók társaságok számára rendkívül fontos a halandósági ráták minél pontosabb előrejelzése a jövőbeli kihívásokra való felkészülés érdekében.

Szakdolgozatom felépítése

A 2. fejezetben bemutatom a halandósági modellépítésekhez elengedhetetlen módszertani alapokat folytonos és diszkrét modellezésre is, illetve leírom a megfigyelésekből adható becslések módját. Majd ezt követően a 3. fejezetben az úttörő Lee-Carter modellről értekezem, melyet az 1992-ben publikáltak. Ezen túl ismertetem néhány korlátját a Lee-Carter modellnek és megoldásokat mutatok azok kezelésére más szakirodalmak által. A 4. fejezetben részletesen ismertetem a szakdolgozatomban használt két halandósági modell matematikai hátterét, mint a Lee-Carter modell összetevőit, paramétereit, becslését, majd az előrejelzéssel kapcsolatos tudnivalókat, illetve az általam tesztelni kívánt SAINT modell részleteit, melynek célja, hogy leküzdje a kis méretű populációk halandósági adataival kapcsolatos problémákat. A cél az, hogy egy nagyobb, referencia populáció segítségével információt szerezzünk a halandóság változásának dinamikájával összefüggésében. Numerikus eredményeimet az 5. fejezetben prezentálom az Egyesült Államok és Svájc férfi és női biztosított adatain elvégzett számítások alapján. Összehasonlítom a SAINT és a Lee-Carter modellek segítségével kapott eredmények pontosságát a 4 populáció adatain. A halandóság-előrejelzési módszerek alkalmazását szemléltetve járadék díjakat számolok periódus táblák és az előrejelző modellek segítségével alkotott kohorsz táblák felhasználásával. Végül összegzem konklúziómat a SAINT modell hasznosságáról. és kitekintést adok a témával kapcsolatban.

2. A mortalitás modellezés módszertani alapjai

Ebben a részben leírom a mortalitáshoz kapcsolódó legfontosabb alapfogalmakat, összefüggéseket, amelyek során a szakirodalomban széleskörűen elterjedt jelöléseket alkalmazom. Az itt bemutatott módszertani alapok Vékás (2016) tanulmánya alapján készültek.

Folytonos modellezés

A folytonos halandósági modellezés az élettartam változón alapszik, aminek jele L . Ez egy folytonos valószínűségi változó, ami továbbá nemnegatív tulajdonságú is. Fontos alapfogalom a *túlélési függvény*, amely az élettartam változó eloszlását bemutató függvény. Ezt jelöli $G: \mathbb{R}^+ \rightarrow [0, 1]$ értelmezésű függvény, amire

$$G(y) = \mathbb{P}(L \geq y), \text{ ahol } (y \geq 0).$$

A túlélési függvényre természetesen teljesülnek az alábbiak:

$$G(0) = 1 ; G(y) = 1 - F(y), \text{ ahol } (y \geq 0).$$

Itt az F függvény jelöli az L élettartam eloszlásfüggvényét. Ismeretes még a reziduális túlélés függvény, amely a hátralévő élettartam eloszlását próbálja megragadni. Tehát ez esetben az egyén már elért az életében adott $x \geq 0$ kort, így az $L - x$ hátralévő élettartamnak eloszlását azzal a feltétellel vesszük figyelembe, hogy $L \geq x$. A reziduális túlélés függvény összefüggései az alábbiak:

$$G_x(y) = \mathbb{P}(L - x \geq y | L \geq x) = \frac{\mathbb{P}(L \geq x + y)}{\mathbb{P}(L \geq x)} = \frac{G(x + y)}{G(x)}, \text{ ahol } (x, y \geq 0).$$

Ezentúl még különösen fontos mutatószám a várható hátralévő élettartam, ami megmutatja, hogy egy x éves egyén várhatóan még mennyi évet fog élni. Számítása a reziduális túlélés függvény integráltjaként történik:

$$e_x = E(L - x | L \geq x) = \int_0^{\infty} G_x(y) dy, \text{ ahol } (x \geq 0),$$

valamint speciális esete $x = 0$ esetén, tehát a születéskor várható élettartam:

$$e_0 = E(L - x) = \int_0^{\infty} G(y) dy.$$

Ha folytonos eloszlású modelleket alkalmazunk, nem szabad elfelejtenünk, hogy $\mathbb{P}(L = y) = 0$, tehát egy adott időpontban 0 a halálozás valószínűsége. Azonban határtértéket alkalmazva a folytonos elemzés során is vizsgálhatunk halálozási valószínűségnek tekinthető értéket. Ezt nevezzük halálozási intenzitásnak, vagy más néven hazárdrátának:

$$\mu(y) = \lim_{\varepsilon \rightarrow 0^+} \frac{\mathbb{P}(L < y + \varepsilon | L \geq y)}{\varepsilon}, \text{ ahol } (y \geq 0).$$

Gyakorlatban a könnyebb számolás érdekében azonban a következő képletet használják, ahol az f az L élettartam sűrűségfüggvényét jelöli, G továbbra is a túlélési függvény és F az élettartam eloszlásfüggvénye:

$$\mu(y) = \lim_{\varepsilon \rightarrow 0^+} \frac{F(y + \varepsilon) - F(y)}{\varepsilon G(y)} = \frac{f(y)}{G(y)}, \text{ ahol } (y \geq 0; G(y) > 0).$$

Diszkrét modellezés

Gyakorlatban nem a folytonos modellezés eszközeit használják általában az aktuáriusi területeken, mivel az adatok egész évben/korban állnak rendelkezésre, valamint könnyebb nem integrálokkal számítani, így a következőkben bemutatom a diszkrét elemzések alapfogalmait. Elterjedt a halandósági tábla alkalmazása, mely éves szinten tartalmazza a koréves halálozási és a koréves túlélési valószínűségeket. Ezek a mérőszámok annak a valószínűségét adják meg, hogy adott korba belépő egyén megéri vagy elhalálozik következő születésnapja előtt.

A valószínűségek:

$$p_x = \mathbb{P}(L \geq x + 1 | L \geq x), \text{ ahol } (x \in \mathbb{N}),$$

illetve

$$q_x = \mathbb{P}(L < x + 1 | L \geq x), \text{ ahol } (x \in \mathbb{N}).$$

Természetesen adódik a q_x és a p_x definíciójából, hogy összegük 1-et eredményez. Továbbá meghatározhatjuk a koréves túlélési valószínűségeket a túlélési függvény segítségével a következően:

$$p_x = \frac{G(x+1)}{G(x)}, \text{ ahol } (x \in \mathbb{N}, G(x) > 0).$$

Az alábbi ábrán szemléletesen látható egy részlet a 2020-as magyar női halandósági táblából.

x	q_x	l_x	d_x	e_x
0	0,00410	100 000	410	72,21
1	0,00015	99 590	15	71,50
2	0,00012	99 576	12	70,51
3	0,00006	99 563	6	69,52
4	0,00010	99 557	10	68,53
5	0,00013	99 547	12	67,53
6	0,00011	99 535	12	66,54
7	0,00008	99 523	12	65,55
8	0,00020	99 511	10	64,56
9	0,00006	99 501	9	63,56
10	0,00004	99 492	8	62,57
11	0,00006	99 484	7	61,57
12	0,00010	99 477	8	60,58
13	0,00012	99 469	11	59,58
14	0,00016	99 459	16	58,59
15	0,00018	99 442	23	57,60
16	0,00029	99 420	28	56,61
17	0,00043	99 391	35	55,63
18	0,00042	99 356	43	54,65
19	0,00050	99 313	49	53,67

9. ábra A 2020-as magyar nők halandósági táblája 23 éves korig. (Forrás: Human Mortality Database)

A halandósági táblák rendszerint tartalmaznak egy maximális kort, jelen tábla esetén ez 100 év, amit ω -val jelölünk. Mivel ilyen szép kort még mostanában sem sokan élnek meg, így nem túlzó feltételezés, hogy q_{100} -at 1 valószínűségűnek tekintjük, így elhanyagolhatónak tekintve a magasabb korokat megélteket. A tábla az ismerteteken túl tartalmazza még a továbbélési rend, a kihalási rend és a várható hátralévő élettartam mutatóit. Ezeket a következőképpen értelmezzük:

$$l_x = l_0 G(x) \quad (x = 0, 1, 2 \dots, \omega)$$

$$d_x = l_{x+1} - l_x \quad (x = 0, 1, 2 \dots, \omega - 1)$$

$$e_x = \frac{1}{l_x} \sum_{i=x+1}^{\omega} l_i + \frac{1}{2} \quad (x = 0, 1, 2 \dots, \omega - 1).$$

Az l_x továbbélési rend megmutatja, hogy egy hipotetikus 100.000 főből álló populációból hányan élnek meg az x éves kort. A számítás igen súlyos feltételezéssel él, miszerint a koréves halálozási valószínűségeket konstansnak tekinti. A d_x kihalási rend a továbbélési rendből származtatható, tehát annak a számát jelöli, hogy az előzőekben felvázolt hipotetikus populációból mennyien halnak meg az adott évben. Végül az e_x jellemzi a várható hátralévő életéveket.

Becslés diszkrét modellből

Minden halandóságot becselő modell a mortalitási rátákon (*mortality rate*) alapszik, mely leírja, hogy mennyi az egy főre jutó halálozások száma adott korban, adott időszakban. Tehát a megfigyelt halálesetek és a populáció létszámának hányadosát kell venni, amit a következőképpen definiálunk:

$$m_{x,t} = \frac{D_{x,t}}{E_{x,t}}.$$

A képletben a $D_{x,t}$ jelöli a t . évben x évesen meghaltak számát, $E_{x,t}$ pedig a kitettséget, azaz a t . évben x éves emberek megélt életéveinek összegét. Vizsgálhatjuk az időszak elején élő x évesek számát, akkor kezdeti kitettségről beszélünk, jele E_x^0 . A másik használatos forma, ha az x évesek átlagos létszámát tekintjük az időszak során. Ezt központi kitettségnak nevezzük, jele E_x^c és az x évesen megélt évek számából határozzuk meg. Ezek alapján alkalmazhatunk kezdeti $m_{x,t}^0$, illetve központi $m_{x,t}^c$ x éves kori halálozási rátát. Mivel a mortalitási ráták a kortól függően nagyban eltérnek egymástól, mint a fiatal és az idős korosztály esetén, így modellezés során ezeknek a logaritmált, olykor logit verzióját érdemes használni.

Továbbá a modellépítéshez elengedhetetlen összefüggés még a mortalitási ráta és a halálozási valószínűség kapcsolatát leírni. Ezek között bonyolult összefüggések állnak, de egyszerűsítő feltételezésekkel alkalmas összefüggést lehet találni. Májer és Kovács (2011) munkája alapján, ha valószínűsítjük, hogy a kitettség a kor lineáris függvénye, akkor az alábbi alakban írható fel a kapcsolat:

$$q_{x,t} = \frac{m_{x,t}}{1 + \frac{1}{2} * m_{x,t}},$$

$$m_{x,t} = \frac{q_{x,t}}{1 - \frac{1}{2} * q_{x,t}}.$$

Másik esetben azt feltételezzük, hogy a halálozási intenzitás egész évek között konstans, és a központi mortalitási rátával megegyezik:

$$q_{x,t} = m_{x,t}^0 = 1 - e^{-m_{x,t}^c}.$$

3. Szakirodalmi áttekintés

Lee-Carter modell és hiányosságai

Mortalitások vizsgálatában a legelterjedtebb módszer Lee és Carter nevéhez fűződik, melyet 1992-es cikkükben mutattak be az Amerikai Egyesült Államok 1900-1989 évek közötti halandósági adatain. A modell szerint a logaritmált központi mortalitási rátákra érdemes egy log-bilineáris modellt illeszteni, amely időtől és korcsoporttól függő paraméterekkel van ellátva. A naptári évtől függő paramétereket úgy javasolják újra becsülni, hogy a tényleges megfigyelt halálesetek számai megegyeznek a modell által várt számokkal. Az előrejelzéshez ARIMA (autoregresszív integrált mozgóátlagolású) folyamatnak tekintik az újrabeült idősort a mortalitási indexnek. Legjobb modellspecifikációnak az eltolásos véletlen bolyongást találta Lee és Carter. A szerzők a paraméterek becslését nem végezheték el regressziós módszerekkel, hiszen a modellt leíró egyenlet jobb oldalán valójában nincsenek tényleges magyarázóváltozók. Így szingulárisérték-felbontással (SVD) kapták meg a legkisebb négyzetes becsléseket. Várakozások szerint az újrabeült mortalitási indexek lineáris csökkenést mutatnak és a halandósági ráták pedig exponenciális csökkenést.

A modell megjelenése óta számos országban végeztek számításokat ezzel a módszerrel. Baseline, Camarda és Booth szerzők 2023-ban publikált cikkükben összefoglalták a módszert a 30 éves évfordulója alkalmából. Megvizsgálták az ez idő alatt megszületett kiterjesztéseket, amik orvosolni kívánták a felmerülő hiányosságokat és összehasonlították ezek hatékonyságát.

A következőkben bemutatom a Lee-Carter modell néhány hiányosságát és hogy azokat milyen módszerekkel próbálták meg mások javítani:

1. *Korrigált k_t*

A módszer nem súlyoz a korcsoportok létszámaival, így a fiatalabb korok, ahol egyáltalán nem jellemző a halálozás, ugyanolyan súllyal számítanak, mint az idősebbek, ahol a halálesetek zöme történik. Lee és Carter a probléma kezelésére egy második lépéses korrekciót alkalmazott a k_t paraméterre, azonban ezt a módszert Booth és szerzőtársai (2002) nem ajánlják bonyolultsága és a kritériumoknak való nem megfelelése miatt. Alternatív módszerek ennek a hibának korrigálására például Lee és Miller (2001) ötlete, ahol a kiigazított

k_t alapján számított születéskor várható élettartamok minden évben meg kell egyezzenek a megfigyelt születéskor várható élettartamokkal, periódus táblákat feltételezve. Másik lehetőség Brouhns és szerzőtársainak (2005) megoldása, akik a hibatag normális eloszlása helyett a korszecifikus halálozási számok Poisson eloszlását feltételezték. Ötletük a Poisson Lee-Carter modell néven vált ismertté és a tanulmányok alapján sok szempontból jobban teljesít az eredeti modellhez képest. Harmadikként megemlítendő még Rabbi és Mazzuco (2021) munkája, akik a becsült mortalitási index kiigazítását az átlag helyett az életkor szerinti halálozás eloszlásának változékonyságához igazították.

2. *Kiindulási hiba*

Lee és Carter eredeti modellje szerint a ráták jövőbeli pályáit a legutolsó időszaki becsült értékéből hosszabbítja meg. Lee és Miller (2001) szerint azonban így nem a megfelelő értékéből indulunk ki az előrejelzéshez. Pontosabb eredményeket lehet kapni szerintük, ha nem a becsült, hanem a legutolsó ténylegesen megfigyelt értékkel hosszabbítjuk meg a mortalitási ráták pályáját. Ezt elemezte ki Stoeldraijer (2018) munkájában nyolc európai ország adatain. Számításai alapján nagyobb pontosságot tudott elérni az utolsó megfigyelt érték használatával, és nagyobb robusztusságot tapasztalt, amikor a közelmúltban megfigyelt évek átlagával dolgozott.

3. *Statisztikai keretrendszer*

Az LC modell talán legjelentősebb korlátja a statisztikai keretrendszeréből fakad, mivel feltételezi a hibatagok normális eloszlását és homoszkedasztikus voltát. Ezek azonban nem feltétlen reális feltételezések emberi halandóságok vizsgálatokor, hiszen idősebb korban változékonnyabbak a halálozási ráták, így a hibatagok varianciái nem lesznek azonosak. Wilmoth (1993) a probléma alternatív kezelésére SVD helyett a súlyozott legkisebb négyzetek módszerét alkalmazta a paraméterek becslésére, ahol a súlyok a halálesetek száma. Másik javaslata a Poisson Lee-Carter modell, amelyben a paraméterbecslések maximum likelihood (MLE) módszerrel történnek. Végül pedig az utóbbi években egyre jobban elterjedő gépi tanulási módszerekkel lehet javítani az eredmények pontosságán. Számos kutatás található már, ahol ezen eszközökkel vizsgálják meg aktuáriusi kérdéseket, a halandóságok előrejelzését, mint például Vékás és Szentkereszti (2022) tanulmánya a magyar adatokon

visszacsatolt neurális hálózatokkal.

4. Rögzített b_x

Lee és Carter modelljének központi feltételezése a b_x paraméter időben rögzített értéke. Lee és Miller 2001-es munkájukban viszont megjegyezték, hogy a paraméter időben változhat. Mindezt arra alapozva, hogy a 20. század első felében fiatal korban gyorsabb, míg idősebb korban lassabb halandóság javulást észleltek, azonban ez később fordítottá vált (Kannisto és társai 1994, Rau és társai 2008). Booth és társai (2002) az illesztési időszak korlátozásával javították a teljes illeszkedést, így kevésbé torzít a rögzített b_x feltételezése. Camarda és Basellini (2021) háromkomponensű kiterjesztett LC modelljükben megoldják a problémát három független additív komponensre (gyermekkor, felnőttkor és időskor) történő felosztással. Ezek mindegyike Poisson Lee-Carterrel vannak felírva és habár az egyes korszpecifikus b_x -ek rögzítettek, azok kombinációja lehetővé teszi a rugalmasságot. Továbbá érdemes megemlíteni még a problémával kapcsolatban Li, Lee és Gerland (2013) tanulmányát. A szerzők Japán és az Egyesült Államok adatai alkalmazásával terjesztették ki a Lee-Carter modellt, melyet rotált Lee-Carter modellnek neveznek. Ebben a b_x paraméterek az idő függvényében változnak, így kezelve a limitációt. Eredményeik alapján azonban nehézkesnek találták a finom rotációt kezelni a minden korcsoportra kiterjedő halandósági modellekben, emiatt leegyszerűsítették a problémát, azaz a halálozási arányok relatív nagyságára összpontosítottak adott korokban, miközben feltételezéseket alkottak a ráták csökkenéseiről más életkorokban.

5. Simítás

Az életkorfüggő halandósági valószínűségek simítása, annak előrejelzésének egységesítése tovább javítja a modell pontosságát, mivel azok sorozata szabálytalan, véletlen ingadozásúak, így a biztosítási számításokhoz érdemes azok simítása. Ezt alkalmazza például a *p-spline* interpolációs összefüggésekkel bővített kutatása Currie-nek (2004). Kapcsolódó meglátásukat fejtették ki Camada és Basellini (2021) is.

6. *Előrejelzési kofidenciantervallum*

Számos kritika érte az eredeti modell előrejelzési kofidenciantervallumát, mivel az csak a mortalitási index idősorának hibatagjaitól függ. Ismertnek tételezi fel a paramétereket, amiket valójában becslünk, így a becslésből fakadó hibát nem veszi figyelembe és kisebbnek becsli az előrejelzési hiba mértékét. Brouhns és szerzőtársai (2005) a Poisson bootstrap megközelítést javasolja a paraméterbizonytalanság megragadására. Szélesebb kofidenciantervallumokat eredményezett Li és társainak tanulmánya (2009), ahol bootstrap eljárást nem Poisson, hanem negatív binomiális eloszlással illesztett halálozási számokkal nézték meg.

7. *Koherens halálozási előrejelzések*

A hasonló populációkra vonatkozó egymástól független mortalitási előrejelzések gyakran hosszú távú inkonzisztenciát mutatnak az életkorspecifikus halálozás és a várható élettartam viszonylatában. Számos erre irányuló kutatás zajlott az elmúlt időben, kezdve Lee 2000-ben kiadott munkájában, ahol a női és a férfi mortalitásokat vizsgálta, majd Li és Lee (2005) 15 országból álló elemzése (LL módszer néven ismerhetjük). Eredményül azt kapták, hogy a 15 országból 11 esetében a populációnkénti előrejelzések megegyeznek a koherens előrejelzésekkel. Ebből arra következtettek, hogy a hasonló országok (demográfiai adottságok tekintetében) adatainak együttes felhasználásával még jobb előrejelzéseket adhatunk meg. Meghatározták, hogy a két populáció hosszú távon nem divergáló előrejelzéséhez szükséges, hogy azonosak legyenek a rájuk vonatkozó paraméterek: eltolásos véletlen bolyongás paraméterei és a korszpecifikus érzékenységi paraméterek. Az LL módszert kiterjesztették két, illetve több dimenzióra is. Van Baal és munkatársai (2016) Hollandia népességére vonatkozóan vizsgálták nem- és iskolai végzettség-specifikus halálozási valószínűségeket. Lyu és szerzőtársai (2021) egy kétlépcsős béta-konvergencia teszt segítségével gondolták tovább az LL eljárást, hogy magába foglalják a megfigyelt két ország és a halálokok közötti koherenciát. Számos más témában is vizsgálták már az eljárással, mint dohányzással, alkohollal vagy elhízással összefüggő mortalitásokra (Janssen és társai (2013) és (2021)). A koherens előrejelzések során kiemelt szerepe van a nemek közötti koherenciának, melyet De Jong és szerzőtársai (2016) egy komplex számokon alapuló keretrendszer segítségével bővítették az LC modellt. Előnyeként könnyen összehasonlítható független modellel a „komplex LC” modell, azonban csak két populáció esetében. Booth 2020-ban kimutatta, hogy

a férfi halálozási adatok esetében jobb az előrejelzés teljesítménye a nemek közötti koherencia feltételezése esetén.

8. Kohorszhatás

Gyakran éri még kritika a Lee-Carter modellt abból az okból, mert nem veszi figyelembe a lehetséges kohorszhatást. Kohorszhatás alatt azt értjük, hogy a naptári évtől és az életkoroktól függő hatásokon túl felléphet az a jelenség is, hogy az azonos időszakban született egyének halandósága függ a születés időpontjától is. Ilyen jellegzetesség állhat fenn például háborúk idején, vagy egész társadalmakat befolyásoló látványos életminőség javulás esetében. A hiányosságok legismertebb megoldását Renshaw és Haberman dolgozta ki 2006-ban, illetve gondolták tovább azt 2011-ben. Ennek alapja, hogy $\eta_{xt} = \ln(m_{xt})$ szisztematikus komponenst tekintették az $\eta_{xt} = a_x + k_t - c_{t-x}$ egyenletben. Itt c_{t-x} a kohorszhatás paraméter, mely a $t-x$ időszakban születettek halandóságának a tipikus mortalitási pályához képesti eltérését reprezentálja.

Az általam említett módszereken túl számos kiterjesztés született még Lee és Carter modellje óta, azonban ezeknek részletezése már nem képezi tanulmányom tárgyát.

Referencia populáció

Aktuáriusi tanulmányaim során megtanultam mennyire fontos a felhasznált adatok minősége, és egyáltalán az adatok léte milyen értékes is. A gyakorlati életet tekintve azonban a biztosítóknak legtöbbször hosszú évekre és nagy kiterjedtséggel rendelkező állományra van szükségük ahhoz, hogy megbízható számolásokat, előrejelzéseket tudjanak kalkulálni. Azonban ez induló biztosítók, kisebb piacokon, kisebb állomány méretek és egyéb okok folyamán nem egyszerűen megoldható. Például Wang, Yue és Chong (2018) kis populációkra specifikus kutatása alapján, ha a megfigyelt népesség mérete kisebb 200 000 főnél, akkor már különösen érzékelhetők a halandósági modellek torzításai, és ezek becslési módszertől függetlenek, mindegyik esetében észre vehetőek. Ezen probléma köré irányított ahhoz, hogy a SAINT modellel (Jarnes és Kryger, 2008) kezdjek foglalkozni részletesebben, mely egy referencia populáción keresztül kíván pontosabb becslést adni.

Az aktuáriusi tudományokban nem újdonság a referencia adatokra való hagyatkozás. Erre alapszanak például a *credibility* modellek is, mint az eredeti Bühlmann-modell. Ezt az alábbiak szerint definiálhatjuk Arató (2001) könyve alapján.

A következő jelöléseket használhatjuk azon feltételezés mellett, hogy X_1, X_2, \dots, X_t kockázatok rögzített θ (rizikóparaméter) mellett feltételesen függetlenek és azonos eloszlásúak:

$$\begin{aligned}\mu(q) &= \mathbb{E}(X_i | \theta = q) \\ m &= \mathbb{E}(X_i) = \mathbb{E}(\mu(\theta)) \\ a &= \mathbb{D}^2[\mathbb{E}(X_i | \theta)] \\ \sigma^2(q) &= \mathbb{D}^2[X_i | \theta = q] \\ s^2 &= \mathbb{E}[\mathbb{D}^2(X_i | \theta)] = \mathbb{E}(\sigma^2(\theta)).\end{aligned}$$

Ekkor a lineáris $g(x_1, \dots, x_t) = c_0 + c_1 x_1 + \dots + c_t x_t$ alakú függvényeknél a

$$g(x_1, \dots, x_t) = z \frac{x_1 + \dots + x_t}{t} + (1 - z)m, \quad z = \frac{at}{s^2 + at}$$

függvény minimalizálja az $\mathbb{E}(\mu(\theta) - g(X_1, \dots, X_t))^2$ várható értéket. Azaz, a biztosító a várható érték becslésénél z -ed résznyire veszi figyelembe a saját tapasztalatot és $1 - z$ -ednyire m -et, a referencia értékét. (Arató, 2001)

SAINT modell

Szakedolgozatomban a Jarner és Kryger által 2008-ban kidolgozott többpopulációs halandósági modellt, a SAINT (*Spread Adjusted International Trend*) modell hatékonyságát vizsgálom meg, összehasonlítva a Lee-Carter metódus jóságával. Jarner és Kryger dán adatokon számolt munkája arra alapszik, hogy egy kisebb populáció adatait egy nagyobb referencia populáció adatainak segítségével határozzák meg, így a kisebb népességű megfigyelésekre is megbízhatóbb eredmények kaphatók. Felvetéseik szerint a kisebb populáció hosszú távon hasonlóan fog viselkedni a referenciához képest, viszont rövid- és középtávú eltérések megfigyelhetők lehetnek. A kisebb populációnak lehet választani például életbiztosítási /nyugdíjbiztosítási cég halandósági adatait, referencia populációnak pedig az adott nemzet/régió halandósági adatát. A kis populáció mortalitása a modell alapján meg fog egyezni a referencia populáció mortalitás és az attól való eltérés összegeként (ezen logika mentén

született a SAINT modell név: *Spread Adjusted INternational Trend*). Ezzel a felbontással elválaszthatjuk a biztosnak tekinthető hosszútávú trendet a rövidebb távú eltérések bizonytalanságától. (A SAINT modell keretrendszerét a következő fejezetben mutatom be.)

A SAINT modellhez kapcsolódóan megemlíteném még itt Wan és Bertschi (2015) módszerét, akik szintén a kis populációk mortalitását kutatták. Jarner és Kryger ötletén alapuló koherens sztochasztikus halandóság-modellezést fejlesztettek, melyben ők is két lépésben dolgoztak: először modellezték a referencia populáció halandóságát, majd ezt követően megbecsülték a mortalitási különbséget a két populáció között. Lee-Carter és Plat modelleket használtak a referencia csoporthoz, mely nem historikus, geográfiai vagy gazdasági okok alapján választódott meg, hanem egyszerűen 15 alacsony halálozással rendelkező ország halmazát alkotta. Svájc népességi adatain tesztelték ötletüket és azt tapasztalták, hogy az alapmodellek mind a nők és férfiak esetében is alul becsülték az élettartam kockázatot.

Menziatti, Morabito és Stranges (2019) tanulmányában Málta népességére alkalmazzák a SAINT modell koncepcióját. Két részből álló modellezésük egy trend és egy eltérés komponenst tartalmaz a referencia populáció alapján. Írásukat kifejezetten az idősebb népességre koncentrálják, hiszen annak van a legnagyobb érdekeltsége az aktuáriusi alkalmazásokban. Tanulmányuk annyiban tér el az említett kis populációs módszerektől, hogy maga a referencia csoport kiválasztása is egy elemzéssel történik, valamint többféle modell kombinációt is leteszteltek. A legjobb választásként az Egyesült Királyság adataival elemezték a máltai halandóságot, és minden modell esetén drasztikusan tudták csökkenteni a mortalitási előrejelzések bizonytalanságát a sztenderd modellekhez képest.

4. A felhasznált halandóság előrejelző modellek

Lee-Carter modell

A halandóság modellezésének legelterjedtebb módszerét, alapkövét Ronald D. Lee és Lawrence Carter alkották meg 1992-ben. (Lee és Carter (1992)) Egy egyszerű modellt készítettek, melyben log-bilineáris összefüggés illesztését javasolták a vizsgált időszaktól és az életkoroktól függő paraméterekkel. Ezzel elemezték az 1900-89 évek alatti egyesült államokbeli életkorfüggő mortalitási ráták változását.

A modell egyenlete

A modell tehát abból indul ki, hogy ismeretesek minden $t \in \{1,2,3, \dots T\}$ időszakban az $x \in X$ korokhoz az adott populációban jellemző központi halandósági ráták, ahol X a korcsoportok halmaza. Ennek összefüggését egy viszonylag egyszerű egyenlet segítségével írták le a szerzők, az alábbiak képeiben:

$$\ln m_{xt} = a_x + b_x * k_t + \varepsilon_{xt}, \text{ ahol } x \in X \text{ és } t = 1,2,3, \dots T.$$

Az egyenlet paraméterei közül a_x és b_x értékei testesítik meg a kortól való függést, míg k_t alakulása a halandóságot kívánja jellemezni az idő alakulásában. Az $\varepsilon_{x,t}$ pedig a modell véletlen hibahatásait foglalja magába. Az ilyen hibategokról feltételezzük, hogy 0 várható értékűek, varianciájuk nagyobb 0-nál és egymástól független, normális eloszlású valószínűségi változók: $\varepsilon_{x,t} \sim \mathcal{N}(0, \sigma^2)$. Ezek alapján a logaritmus mortalitási ráták egymástól függetlenek és szintén normális eloszlást követnek a következő paraméterek szerint:

$$\ln m_{x,t} \sim \mathcal{N}(a_x + b_x * k_t, \sigma^2), \quad \forall x, t.$$

Továbbá a modell helyes alkalmazásához a szerzők kiegészítése szerint két paramétermegkötést kell még figyelemmel kísernünk, miszerint:

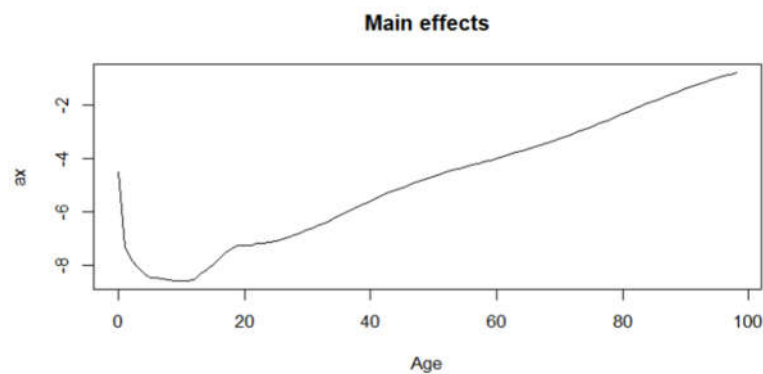
$$\sum_{t=1}^T k_t = 0 \text{ és } \sum_{x \in X} b_x = 1.$$

Habár a normalitás feltevése nem szerepelt az eredeti Lee-Carter cikkben, de hozzá lehet adni a modellhez, mely könnyebb számítást eredményez. Ekkor a maximum likelihood becslés épp azonos lesz a szingulárisérték-felbontással kapott legkisebb négyzetes becsléssel.

Paraméterek értelmezése

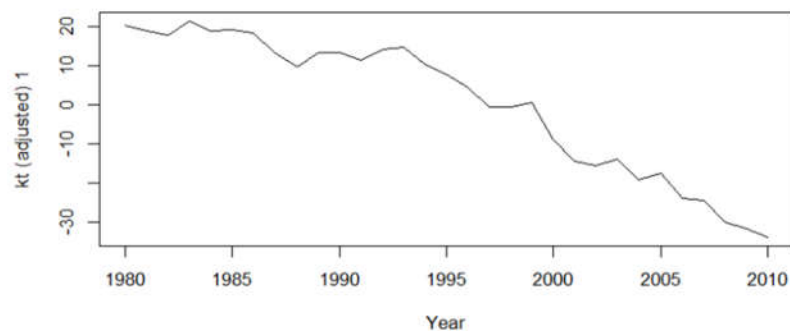
Az előbbieken bemutatott paraméterekhez a következő értelmezéseket lehet kapcsolni:

- * Az a_x , mint *átlagos logaritmusos halandóság* értékeket mutatja meg az egyes korokban, tehát hogy adott életkorokban általában mekkora halandóság jellemzi az embereket. Jellemző alakja, hogy a születéskor igen magas, majd gyors ütemben eléri minimumát, ezután 20 éves kor körül kipúposodik, majd lineárisan növekszik az életkor előrehaladtával. Ezt szemlélteti a következő ábra, mely a magyar populáció adatain készült Lee-Carter módszerrel az 1980-2010-es időszakot figyelembe véve.



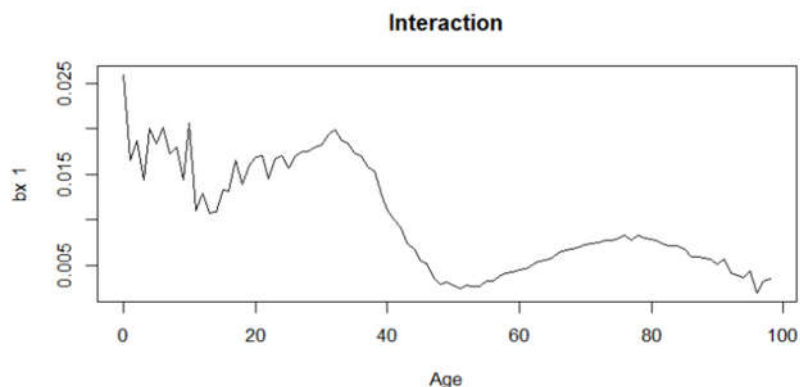
10. ábra Az átlagos logaritmusos halandóság ábrázolása az 1980-2010-es adatokkal Magyarországon (saját szerkesztés)

- * Az egyenlet időtől függő komponense pedig a k_t amit mortalitási indexnek nevezünk. Ez a halandóság mértékének időbeli változását mutatja be a megfigyelt időszakok között. Alakja jellemzően monoton csökkenő, azonban egy-egy rosszabb időszakban megnövekedik, mint háborúk, járványok idején.



11. ábra A mortalitási index alakulása az 1980-2010-es magyarországi adatok felhasználásával (saját szerkesztés)

- * A b_x paraméter értékét úgy értelmezzük, mint *életkorfüggő érzékenység* a halandósági index alakulására nézve. Tehát ha a mortalitási index egy megfigyelt időegység alatt változik egy egységet, akkor várhatóan b_x nagysággal változik meg a logaritmikus halandósági ráta értéke. Az alábbi ábra szemlélteti egy adott populáció esetén:



12. ábra Az életkorfüggő érzékenység alakulása Magyarországon az 1980-210-es adatok szerint (saját szerkesztés)

- * Végül pedig megemlíteném σ^2 paramétert, ami az átlagos négyzetes hiba. Korcsoportonként és időszakonként írja le a logaritmikus halandósági ráták változékonyságát. Amennyiben alacsony értéket vesz fel, akkor a modellünk jól írja le a valóságon alapuló megfigyeléseket.

Paraméterek becslése

A paraméterek becslését leggyakrabban maximum likelihood (ML) módszerrel szokás elvégezni, azonban a szerzők a szinguláris érték felbontást (SVD) javasolták, mely ugyanazt az eredményt fogja visszaadni.

A log-likelihood függvény a log-normális eloszlású mortalitással, amit maximalizálni szeretnénk a paramétermegkötések teljesülése mellett:

$$l = -\frac{1}{2}XT * \ln(2\pi) - XT * \ln \sigma - \frac{1}{2\sigma^2} \sum_{x \in X} \sum_{t=1}^T (\ln m_{x,t} - \alpha_x - \beta_x * \kappa_t)^2.$$

Ebben a becslésben azonos súllyal vesszük figyelembe a megfigyeléseket, azonban ez ellentmond azzal, hogy az LL modellben koronként és évenként az egyes csoportok halálozásának nagysága eltérő mértékű lehet. Emiatt a következő kiigazítással érdemes dolgozni, aminek létezik egyértelmű megoldása, amennyiben minden időszakban megfigyelhető haláleset:

$$\hat{\kappa}^{adj} = \left\{ \kappa \in \mathbb{R}^T : D_t = \sum_{x \in X} D_{xt} = \sum_{x \in X} E_{x,t} \exp(\hat{\alpha}_x + \hat{\beta}_x * \kappa_t) \quad t = 1, 2, \dots, T \right\}.$$

Előrejelzés

Az említettek alapján már készíthetünk előrejelzéseket a paraméterekre, ha elkészítjük még a κ_t (időtől függő) paraméter külön modellezését. Az eredeti módszer alapján empirikus vizsgálatok után ehhez a egy eltolásos véletlen bolyongás a legalkalmasabb (*Random Walk with Drift*, azaz eltolásos ARIMA (0,1,0)). Ekkor a sztochasztikus modell $\hat{\kappa}_1^{adj}$ kezdőérték mellett, s paraméterű eltolással és Φ_t (független, azonos varianciájú normális eloszlású) hibataggal a következő:

$$\hat{\kappa}_t^{adj} = \hat{\kappa}_{t-1}^{adj} + s + \Phi_t \quad (t = 2, 3, \dots, T).$$

Tehát Lee és Carter szerint a mortalitási index egy olyan lineáris sztochasztikus folyamat, amelynek növekményei függetlenek és s a trendjének meredeksége. Az előrejelzésekhez szükséges s és Φ_{RWD} paraméterek megbecslése után a következő formula segítségével adhatunk pontbecslést a jövőbeli mortalitási indexek értékeire:

$$\begin{aligned} \mathbb{E}(\hat{\kappa}_{T+k}^{adj}) &= \mathbb{E} \left(\hat{\kappa}_T^{adj} + \sum_{i=1}^k (\hat{\kappa}_{T+i}^{adj} - \hat{\kappa}_{T+i-1}^{adj}) \right) = \\ &= \hat{\kappa}_T^{adj} + \sum_{i=1}^k \mathbb{E}(\hat{\kappa}_{T+i}^{adj} - \hat{\kappa}_{T+i-1}^{adj}) = \hat{\kappa}_T^{adj} + k\hat{s} \quad (k = 1, 2, \dots). \end{aligned}$$

Ebből a $T + k$. évre vonatkozó előrejelzés a logaritmált mortalitásra:

$$\ln \hat{m}_{x,T+k} = \hat{\alpha}_x + \hat{\beta}_x * (\hat{\kappa}_T^{adj} + k\hat{\delta}) \quad (x \in X, \quad k = 1, 2, \dots).$$

SAINT modell

Jarner és Kryger tanulmánya a halandóság leírására az idő- és korfüggő mortalitást használják $\mu(t, x)$, amire mint mortalitási függvény/intenzitás utalnak (azonban később felteszik, hogy $\mu(t, x) = \mu(x)$). A modell felteszi, hogy a referencia populáció halálozásainak számai Poisson eloszlást követnek az alábbi paraméterekkel:

$$D_{ref}(t, x) \sim Poisson (\bar{\mu}(t, x) E_{ref}(t, x)),$$

ahol $D_{ref}(t, x)$ a referencia populáció halálozásainak száma, $E_{ref}(t, x)$ a referencia populáció központi kitettségeinek száma, $\bar{\mu}(t, x)$ pedig egy átlagos értéke a mortalitási intenzitásnak.

Következésképpen számolandó:

$$\bar{\mu}(t, x) = \frac{1}{4} [\mu(t, x) + \mu(t, x + 1) + \mu(t + 1, x) + \mu(t + 1, x + 1)].$$

A fenti jelölésekkel a log-likelihood függvény így áll elő:

$$l(\theta) = \sum_{t,x} [D_{ref}(t, x) \log (\bar{\mu}(t, x) E_{ref}(t, x)) - \log (D_{ref}(t, x)!) - \bar{\mu}(t, x) E_{ref}(t, x)].$$

A kis populáció modellezése hasonlóan alakul a referenciához képest: modellezve a referenciát egy hozzáadódó eltéréssel. Az alapvető feltevés, hogy hosszútávon hasonlóan fog viselkedni a kis populáció a nagy referenciához képest, azonban lesznek rövid-/középtávú eltérések, amelyek 0 körül fognak fluktuálni.

Ha adott μ referencia trend, akkor feltesszük, hogy részpopulációnk megfigyelései függetlenek, Poisson eloszlásúak a következő paraméterekkel:

$$D_{rész}(t, x) \sim Poisson (\bar{\mu}_{rész}(t, x) E_{rész}(t, x)),$$

ahol

$$\bar{\mu}_{rész}(t, x) = \bar{\mu}(t, x) * \exp(a_t + b_t * r_1(x) + c_t * r_2(x))$$

$$r_1(x) = (x - 60)/40$$

$$r_2(x) = (x^2 - 120x + \frac{9160}{3})/1000.$$

Az a_t paraméter a részpopuláció és a referencia mortalitás közötti különbséget ragadja meg, a b_t és c_t paraméter pedig magasabb rendű korrekciók, amelyek biztosítják a jobb illeszkedést. Ezeket a paramétereket úgy választjuk meg, hogy minimalizálják a részpopulációra (előzőkhez hasonlóan) felírható log-likelihood függvényt. Az $r_1(x)$ és $r_2(x)$ regresszorokat úgy választjuk meg, hogy merőlegesek legyenek és körülbelül 1-re normalizálódjanak 20 és 100 éves korban.

A részpopuláció mortalitásának előrejelzését a modell szerint úgy adhatjuk meg, ha az optimalizált a_t , b_t és c_t paramétereket vektor-autoregresszív (VAR) modell segítségével előrejelezzük a kívánt időszakra.

5. Numerikus elemzés

Adatbemutatás

Adataimat két fő forrásból gyűjtöttem. Az egyik a *Human Mortality Database* (www.mortality.org) honlapja, mely számos ország halandósági adatait szolgáltatja naprakészen, egészen akár az 1800-as évekre visszamenőleg éves gyakorisággal. Innen gyűjtöttem le mindkét nemre az általam vizsgált két ország - Svájc és az Amerikai Egyesült Államok (továbbiakban rövidítve USA-ként hivatkozva) - halálozási számait és a központi kitettségeit éves bontásban.

Másik fő forrásom a *Society of Actuaries* (<https://www.soa.org>) honlapja volt, mely egy 1949-ben az Egyesült Államokban alapított aktuáriusok globális szakmai szervezete által létrehozott honlap, mely nyilvántart számos országra biztosított halandósági táblákat.

A rendelkezésemre álló adatokból Svájc esetén 1941-től 1995-ig gyűjtöttem össze 20-tól 100 éves korig az elemzésemhez szükséges adatokat, melyek 10 év hosszúságú időszakok szerinti bontásban voltak elérhetőek számomra. Mivel időszakosan és nem éves szinten elemeztem első lépésben, így a halálozási számokat és a kitettségeket is ilyen formára összegeztem. Az Egyesült Államok esetén 1957-től 1993-ig sikerült összegyűjtenem a járadékos halandóságokat, melyek 5 éves periódusokban szemléltethetőek. Ezen adatokat 10 éves kortól 110 éves korig használtam fel. 100 éves korban maximalizáltam a figyelembe vett korok intervallumának felső korlátját, azonban fontos tudni, hogy a HMD a 80 év feletti értékeket a populáció kisebb mértéke miatt simítással korrigálja. Ez egy gyakorta használt módszer a magasabb korokra, amit a HMD szervezete a Kannisto eljárással (*Human Mortality Database: Methods Protocol, 2021*) végez el.

A biztosított adatok szűkössége okán kevés megfigyeléssel tudtam csak dolgozni, mely a modellek építésekor súlyos hibákat eredményezett (mint például a szerzők által ajánlott VAR-modell a kevés mintaelemszám miatt értelmezhetetlen), ezáltal az eredmények jósága sem volt elfogadható minőségű. Ennek kiküszöbölése érdekében az 5, illetve 10 éves időszakra megadott adatokat éves gyakoriságúra konvertáltam a lineáris interpoláció képletének alkalmazásával, emiatt jóval több megfigyeléssel tudtam tovább dolgozni. Így Svájc esetében 51, az Egyesült Államok esetében pedig 33 megfigyeléssel rendelkeztem.

A lineáris interpoláció általános képlete, ahol y a megfigyelt változók, x pedig az évek száma:

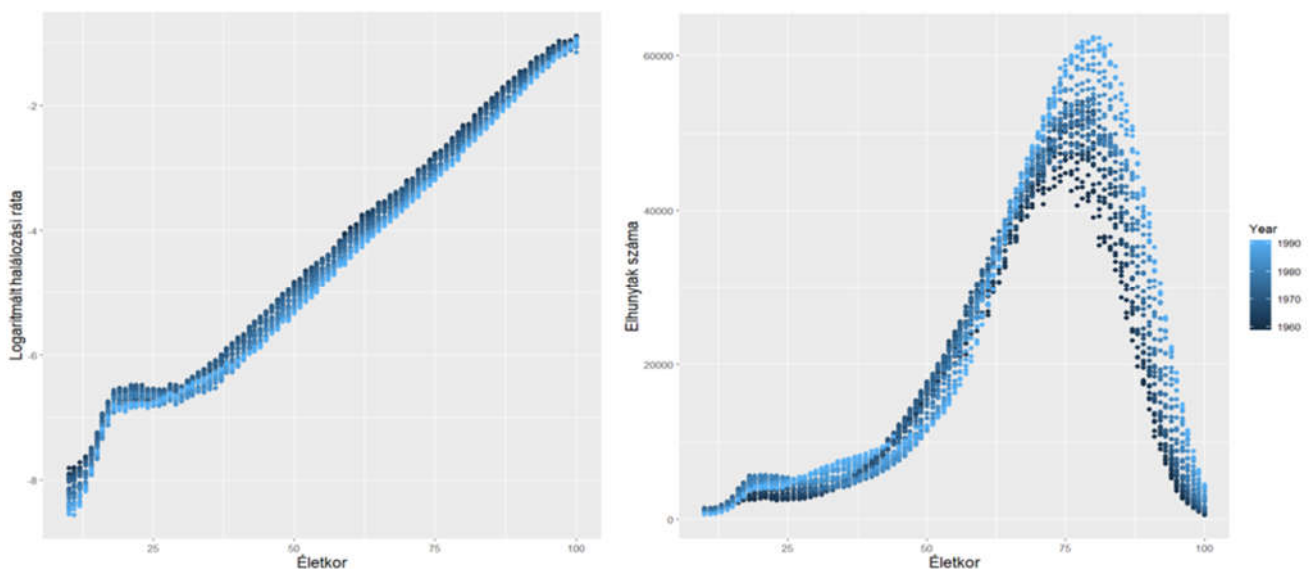
$$y = y_1 + \frac{(y_2 - y_1)}{(x_2 - x_1)} (x - x_1).$$

Dolgozatomban speciálisan:

$$m_{x,t} = m_{x,t_0} + \frac{(m_{x,t_1} - m_{x,t_0})}{(t_1 - t_0)} (t - t_0),$$

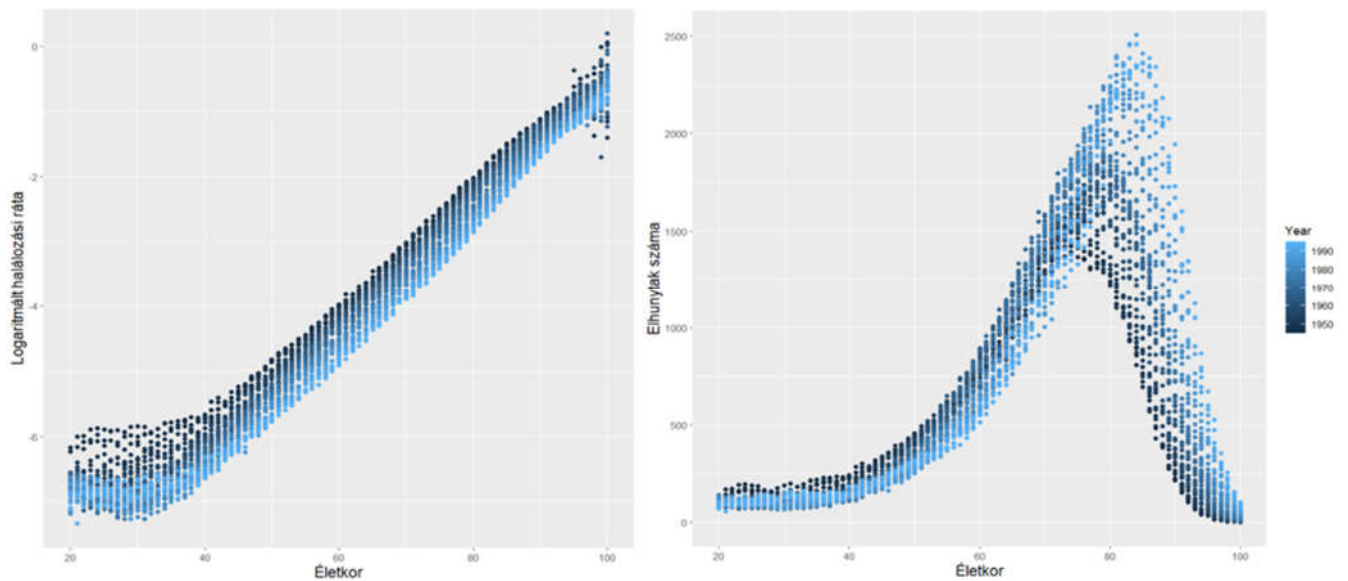
ahol t_0 és t_1 az adott t naptári év előtti és utáni intervallumközéppont.

A vizsgált populációk halandósága az egyes országok biztosítotti adatai, melynek referencia populációja az egész országra vonatkozó megfigyelések. Mivel az én esetemben járadékosi adatokkal dolgoztam, így nem tartom relevánsnak az összes kor figyelembevételét. Az adatbázisom alsó korhatárjának a 60 évet tartom indokolhatónak, felső korhatárjának pedig a 100 évet hagytam, mivel a hosszú élet kockázata egyre nagyobb jelentőségűvé kezd válni. Az alábbiakban néhány grafikon látható, mely a két ország uniszex halandósági rátáinak és halálozási számainak vizualizációját szolgálja.



13. ábra Uniszex log-halandósági ráták (bal) és uniszex halálozási számok (jobb) alakulása 1959-1991 között az Amerikai Egyesült Államokban évente (saját szerkesztés)

14. ábra Uniszex log-halandósági ráták (bal) és uniszex halálozási számok (jobb) alakulása 1945-1995 között Svájcban évente (saját szerkesztés)



Az amerikai log-halálozási ráták esetében látványosan kiemelkedik a fiatal kori púp, amelyet a fiatalabbak kisebb óvatosságának, vakmerőségének szoktak titulálni. A színkódolásból láthatjuk, hogy a halandósági ráták az évek folyamán folyamatosan csökkentek. Érdekes még, hogy az amerikaiak esetében a 75 éves kor körül, míg a svájciaknál 80 év körül jelentkezik a legmagasabb szám az elhunytakban.

Egypopulációs modellezés

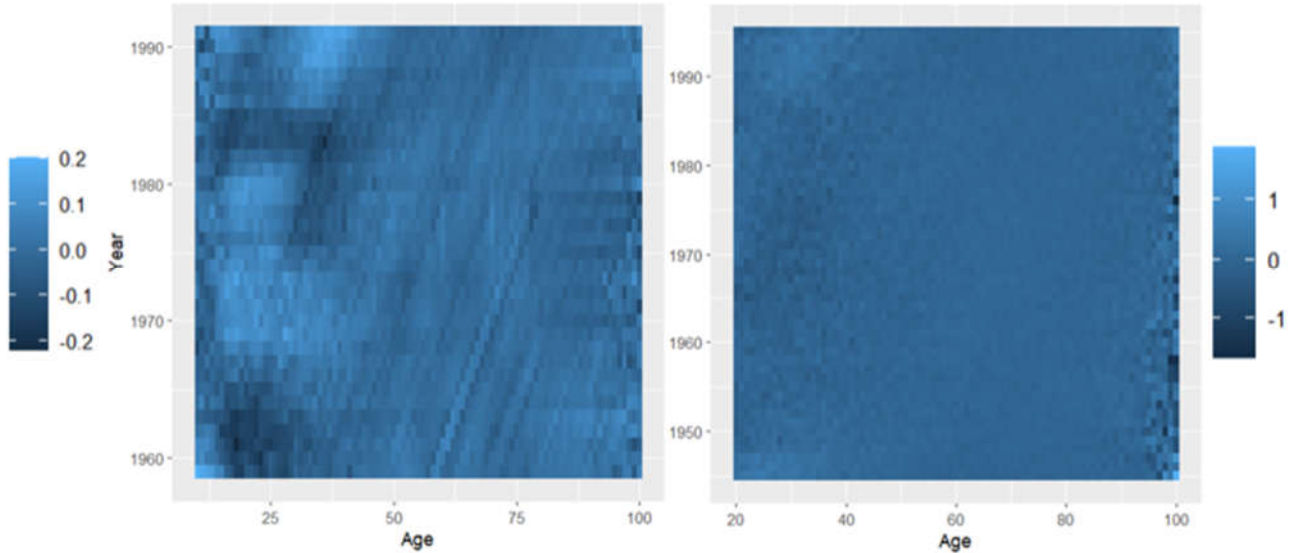
Először a két populációt egymástól függetlenül modelleztem országonként és nemenként Lee-Carter módszerrel, mind a HMD adatokon és mind a biztosított adatokon. A modellek futtatásához az *R* szoftver *StMoMo* csomagját használtam. Kiemelném ezeknek a reziduális hőtérképét, mely vizuálisan szemlélteti további problémáját az LC-modellnek, a kohorszhatás figyelmen kívül hagyását.

Referencia populáció Lee-Carter modelljei

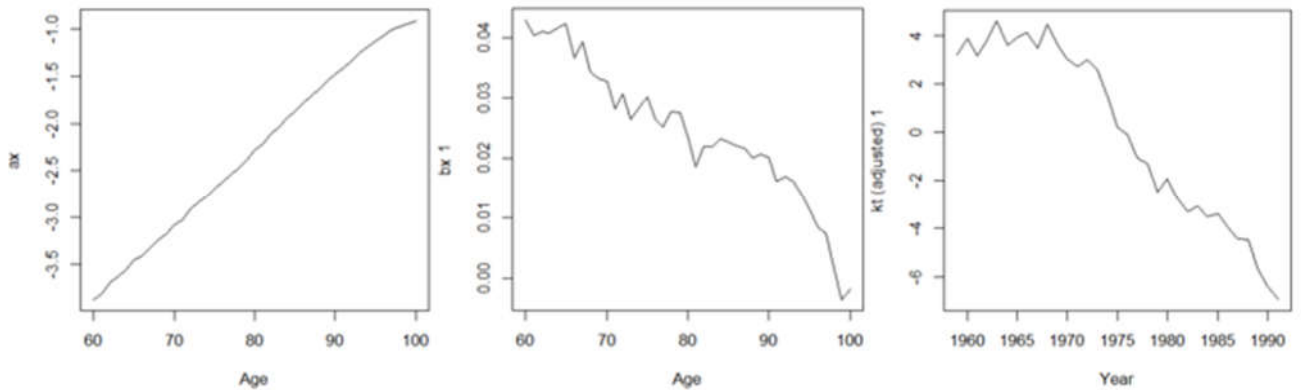
Az azonos évben születettek csoportját nevezzük egy kohorsznak, a 7. ábrán balról jobbra felfelé irányba tartó diagonális vonalak mentén láthatóak. Ez a jelenség az amerikaiak esetében rendkívül szembeötlő, míg a svájciaknál nem igazán vehető észre hasonlóan elkülönülő egyenesek a

hő térképen. A kohorszthatás alapján tehát a mortalitást nem csak az adott évek és korok befolyásolják, hanem az is, hogy melyik a születési évszám.

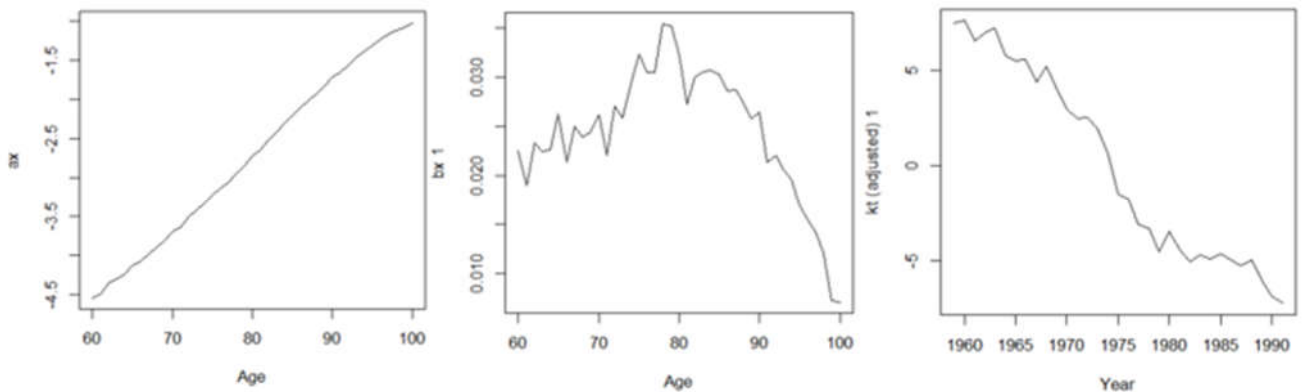
15. ábra Egyesült Államok férfi (bal) és Svájc férfi (jobb) Lee-Carter modell reziduálisainak hő térképe



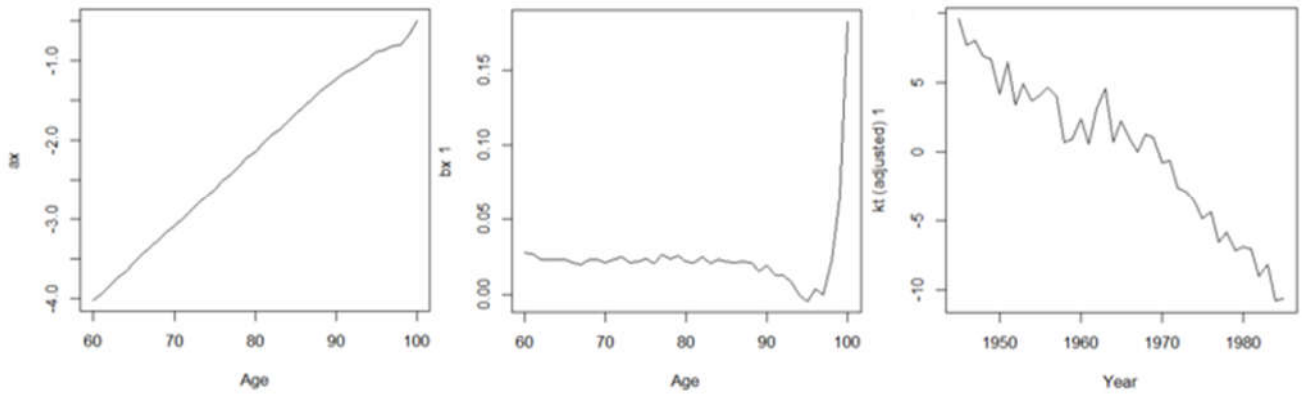
Következőkben a lefutatott négy modellre kapott paraméterek ábráit mutatom be:



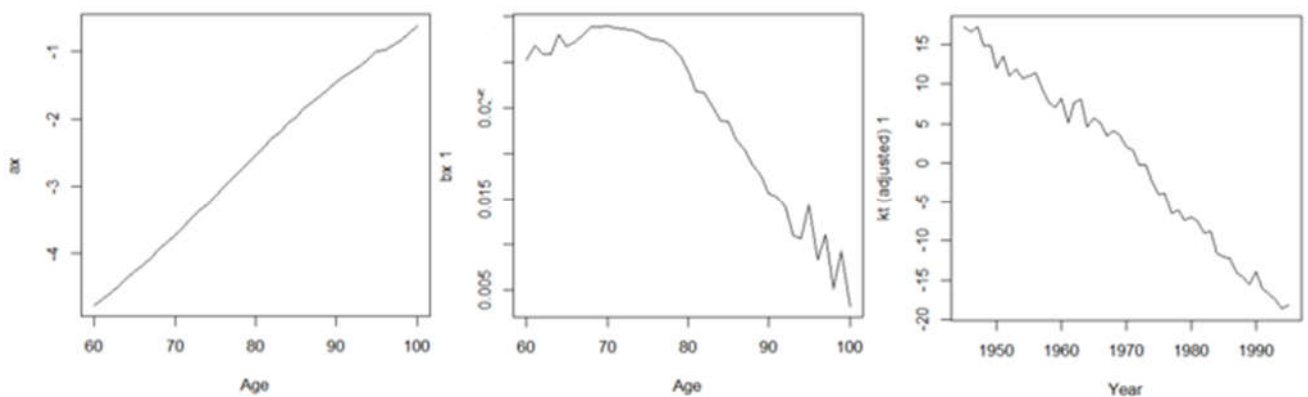
16. ábra 1959-1991 között a férfiak Lee-Carter modelljének paraméterei 60-100 éves kor között USA-ban



9. ábra 1959-1991 között a nők Lee-Carter modelljének paraméterei 60-100 éves kor között USA-ban



10. ábra 1945-1995 között a férfiak Lee-Carter modelljének paraméterei 60-100 éves kor között Svájcban



11. ábra 1945-1995 között a nők Lee-Carter modelljének paraméterei 60-100 éves kor között Svájcban

Az a_x paramétert tekintve mind a két országra és mind a két nemre a várakozások szerint halad, tehát minél magasabb korban vannak az emberek, annál magasabb a halandóság valószínűsége. A b_x minden esetben egy korrallal csökkenő életkor érzékenységet mutat, kivéve a svájci férfiak esetében, ahol 100 éves kor körül hirtelen növekedés látható. Végül pedig a k_t mortalitási indexet leíró időszora látható. USA esetében mindkét nemre látványos javulás érzékelhető a mortalitási index nagyságában, tehát a néphalandóság fokozatosan javult, bár a csökkenés üteme a későbbi években mérséklődött. Svájc esetében szinte monoton csökken a paraméter értéke. Mivel a vizsgált időszakokban nem voltak nagy háborúk, járványok az országokban, így nem mutatkozik a halandóság növekedése.

A modelleket rövidebb megfigyelési időszakra is lefuttattam, ezáltal tesztelni tudtam a Lee-Carter módszer becslését a valós adatokhoz képest. Az amerikai adatokon 33 megfigyelési időszakból 8 évet választottam le tesztelési célzattal, míg a svájci adatokon összesen 51 időszakból 10 évre teszteltem a modell becslőképességét. Egy táblázatban összefoglalva szemléltetném a predikciók teljesítményének értékelését, azaz, hogy a valós értékek mennyire térnek el a modell által becsült

értékektől, melyet átlagos négyzetes hiba (*Mean Squared Error*) mérésével számoltam.

Az alábbi képlet alapján határoztam meg:

$$MSE = \sum_{x \in X} \sum_{t=1}^T \frac{(\ln(\text{előrejelzett } m_{xt}) - \ln(\text{valós } m_{xt}))^2}{n},$$

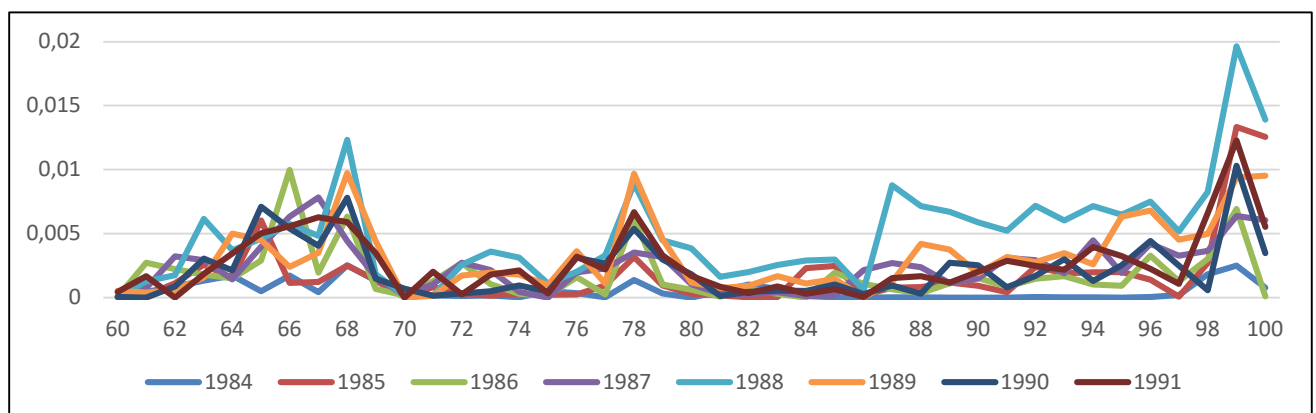
ahol n a megfigyelések száma.

Az alábbi MSE értékeket kaptam a referencia populációra épített modellekre:

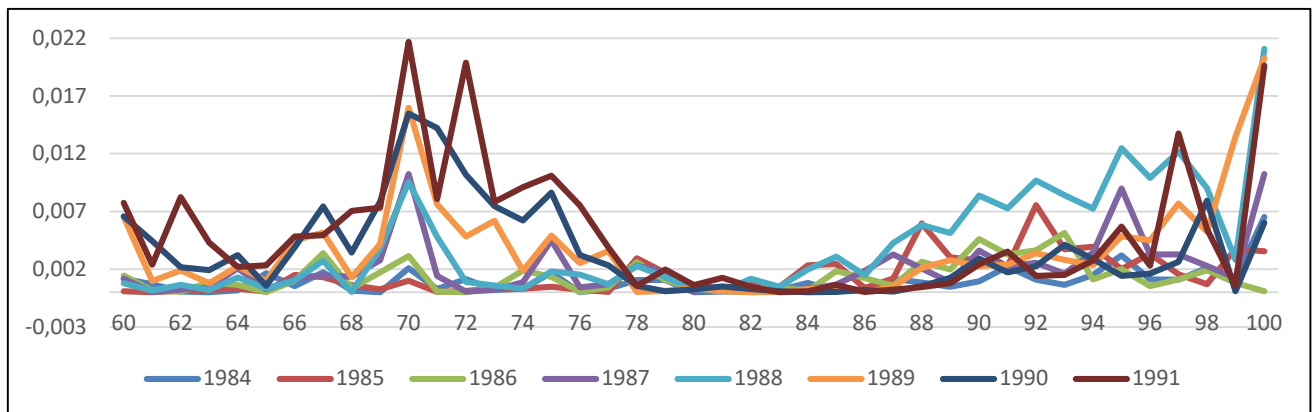
1. Táblázat MSE értékek a valós és Lee-Carter modellel becsült értékek között a referencia populációra

Ország	Nem	MSE
USA	<i>férfi</i>	0,00013
	<i>nő</i>	0,000083
Svájc	<i>férfi</i>	0,0086
	<i>nő</i>	0,00051

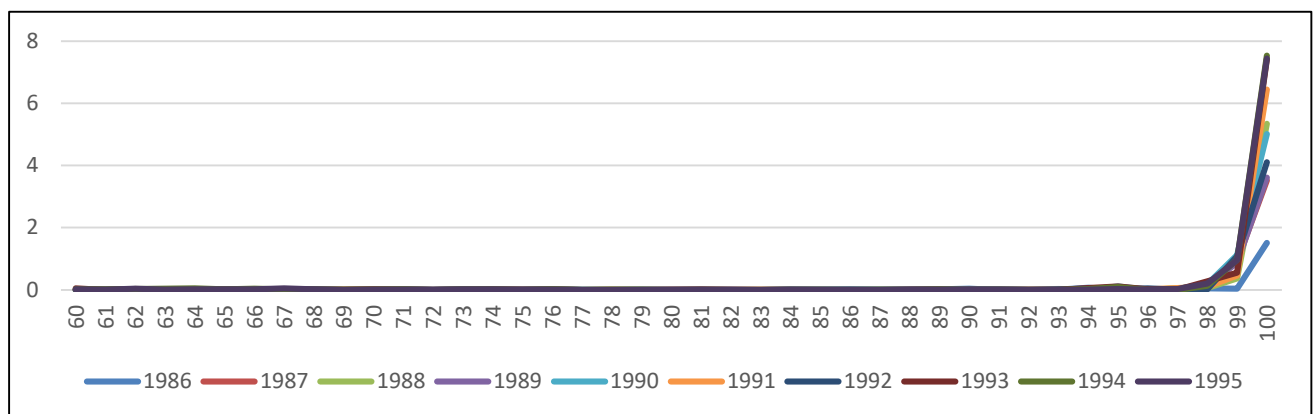
MSE értékelését tekintve az tekinthető pontosabb becslésnek, ahol minél kisebb értéket vesz fel. Tehát a Lee-Carter modell becsülése során a legkisebb értéket az amerikai nők csoportjára tapasztaltam. Habár az MSE mutatók értékei kis számot vesznek fel, kiemelendő, hogy a legkisebb MSE-vel rendelkező amerikai nőkkel ellentétben 100-szor magasabb érték látható a legmagasabb értékkel bíró svájci férfiak populációjára. Továbbá ábrázoltam az egyes korokban a tesztidőszak éveire a valós és becsült mortalitási ráták négyzetes eltéréseinek mértékét. Ezeken megfigyelhető, hogy hol található szisztematikus eltérés.



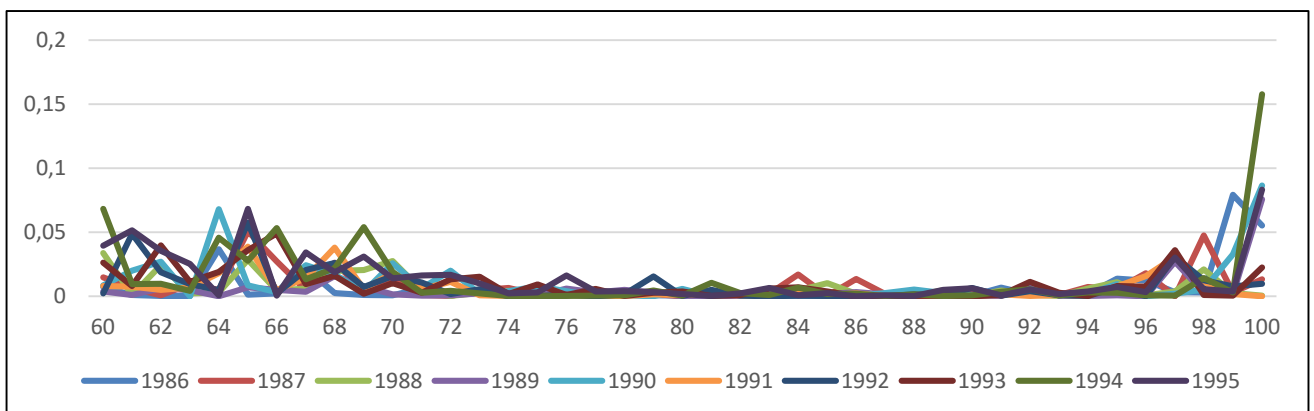
12. ábra Valós és LC modell becsült értékeinek eltérése a referencia USA női adatokon (saját szerkesztés)



13. ábra Valós és LC modell becstült értékeinek eltérése a referencia USA férfi adatokon (saját szerkesztés)



14. ábra Valós és LC modell becstült értékeinek eltérése a referencia Svájc férfi adatokon (saját szerkesztés)

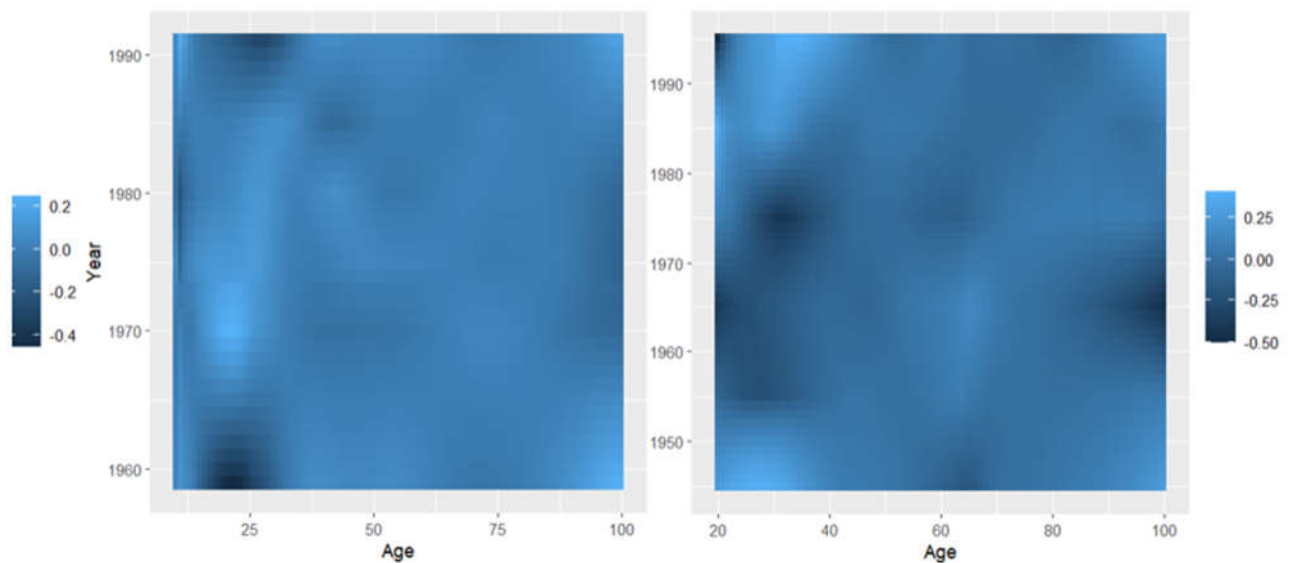


15. ábra Valós és LC modell becstült értékeinek eltérése a referencia Svájc női adatokon (saját szerkesztés)

Látható, hogy az amerikaiak esetében nőknél és férfiaknál is 69-70 éves korban mutatkozik nagy eltérés, illetve a korhatár felső korhatárjánál. Svájci adatoknál rendkívül magas eltérés látható a férfi adatoknál a felső korhatárnál, azonban a nőknél hasonló a tendencia az amerikai diagrammokhoz.

Járadékosi populáció Lee-Carter modelljei

A biztosított adatokat is kielemeztem a Lee-Carter modellel. Megnéztem most is a reziduálisok hőterképét, amelyeken nem figyelhetők meg a kohorszok, mint korábban. Ez várható is volt, hiszen a biztosított populáció sokkal homogénebb mintát jelent, mivel a biztosítók, minél kisebb kockázatú állományokat céloznak meg kialakítani.



16. ábra Egyesült Államok férfi (bal) és Svájci férfiak(jobb) biztosított adatainak Lee-Carter modell reziduálisainak hőterképe (saját szerkesztés)

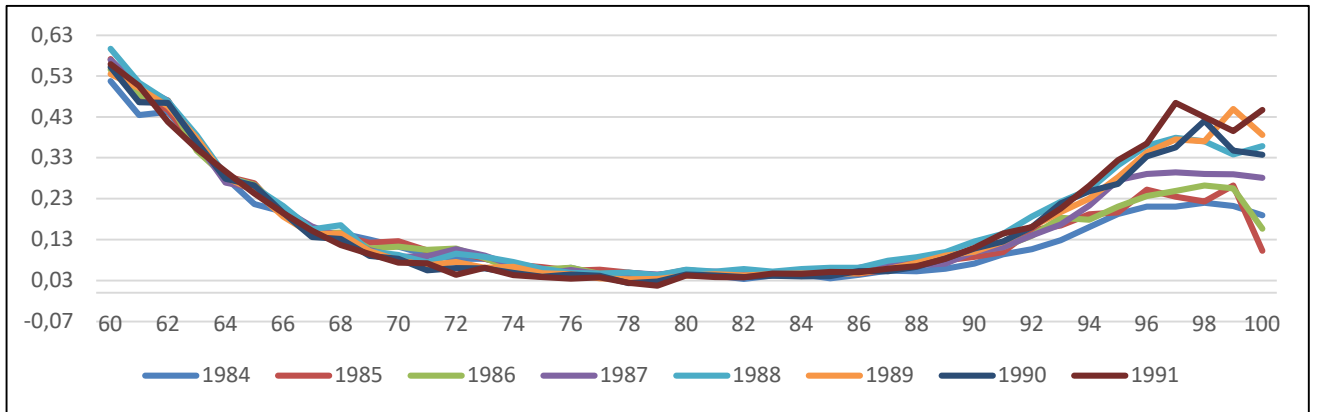
Ezekre az adatokra is néztem tesztidőszakot és ellenőriztem az ellenőrző képességét a modelleknek. Az előzőekhez hasonlóan ugyanannyi időszakot hagytam ki a tesztelésre. Az MSE értékeik a következők:

2. táblázat MSE értékek a valós és Lee-Carter modellel becsült értékek között a biztosított populációra

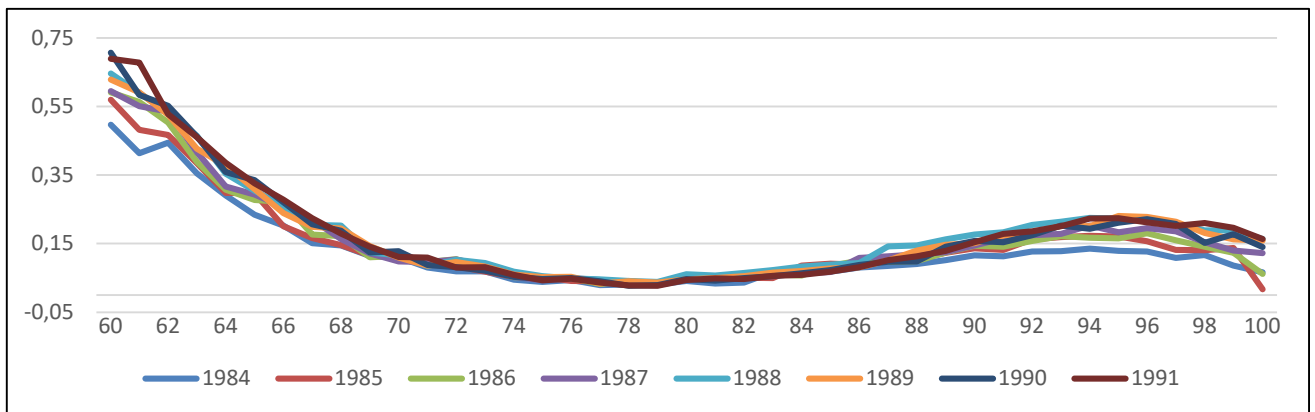
Ország	Nem	MSE
USA	férfi	0,0046
	nő	0,0021
Svájc	férfi	0,0015
	nő	0,0010

Tehát a biztosított adatokra a legkisebb MSE értéket a svájci nők populációja produkálta, míg a

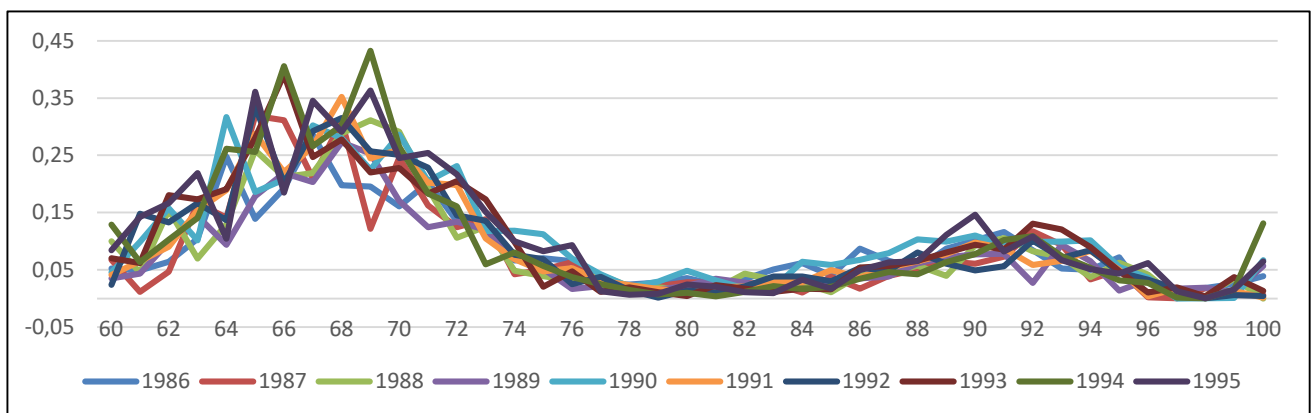
legnagyobbat az amerikai férfiak. A teljes néppességet leíró Lee-Carter MSE eredményeihez képest itt pont az 'ellentett' populációk teljesítettek legjobban/legrosszabbul. A következő diagrammokban megfigyelhetőek a valós és a modell által becsült értékek közötti négyzetes eltérések:



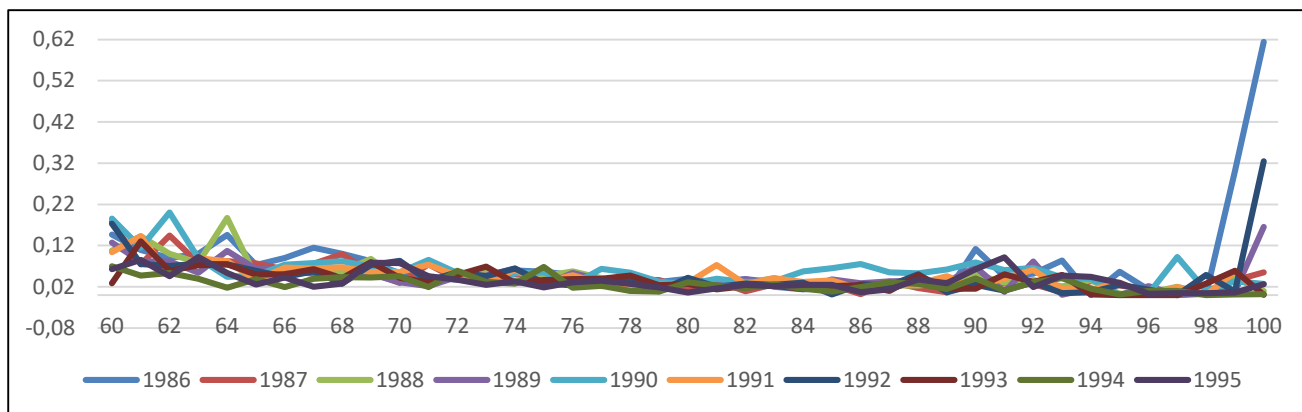
17. ábra Valós és LC modell becsült értékeinek eltérése a biztosított USA férfi adatokon (saját szerkesztés)



18. ábra Valós és LC modell becsült értékeinek eltérése a biztosított USA női adatokon (saját szerkesztés)



19. ábra Valós és LC modell becsült értékeinek eltérése a biztosított Svájc női adatokon (saját szerkesztés)

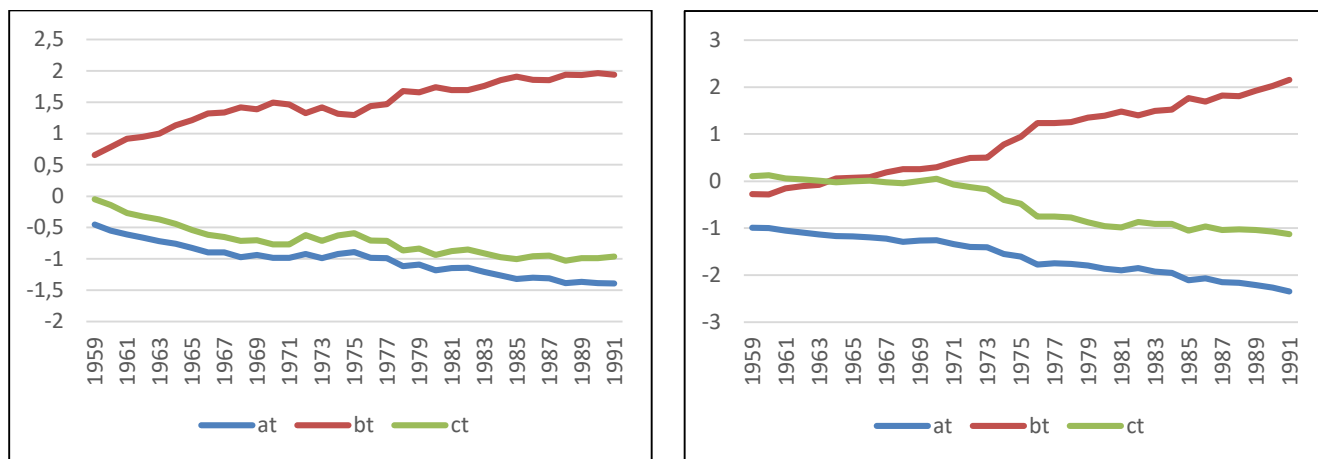


20. ábra Valós és LC modell becsült értékeinek eltérése a biztosított Svájc férfi adatokon (saját szerkesztés)

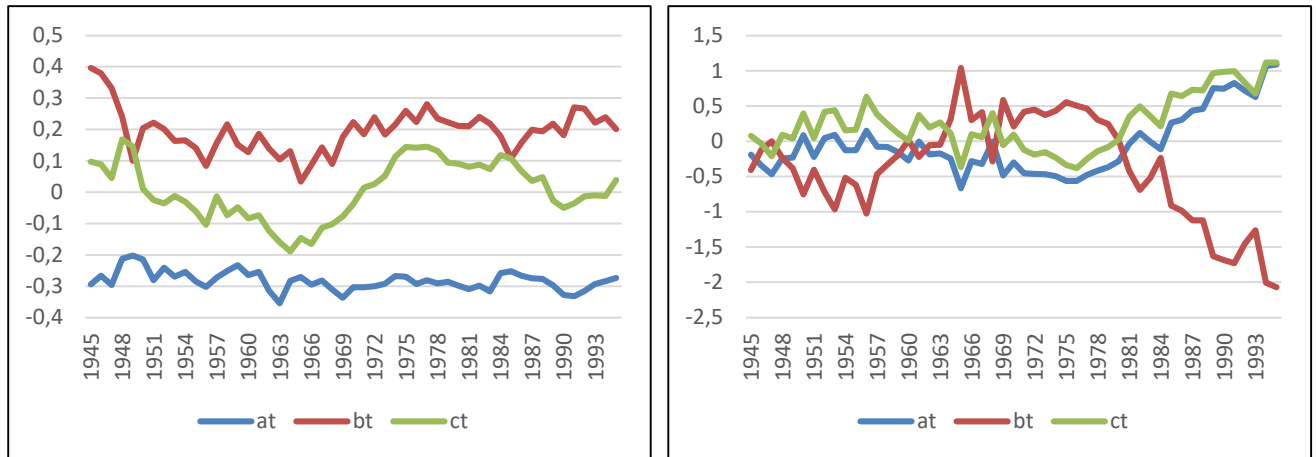
Amerikaiak ábráján mind a nőknél és a férfiaknál is a korintervallumok végein vannak a nagyobb eltérések, és 70-90 éves kor között becsül a legkisebb hibával a modell. Svájc adataira nem egyezik meg a hibák diagramja a két nemre, mg a férfiaknál szinte végig kiegyensúlyozottak az eltérések és csak a 100 éves korhoz közeledve ugrik meg jobban, addig a nőknél 60-75 évesek csoportjára magasabb hibákkal becsül. Nagyságrendileg szinte azonos mértékű eltérések észlelhetők a különböző populációkra, a hibák maximuma a svéd nőkre a legkisebb.

SAINT modellezés és eredményei

Az 4. fejezetben leírtak alapján Excel segítségével meghatároztam a SAINT modellhez szükséges a_t , b_t , és c_t paramétereket, melyek optimalizációs probléma megoldásából adódnak. A teljes megfigyelt időszakra nézve ábrázoltam őket diagrammon.



21. ábra Optimalizált paraméterek az amerikai férfiak (bal) és nők (jobb) adataihoz 60 éven felettiekre (saját szerkesztés)



22. ábra Optimalizált paraméterek a svájci férfiak (bal) és nők (jobb) adataihoz 60 éven felettiekre (saját szerkesztés)

Az ábrák alapján az figyelhető meg, hogy a b_t paraméter időben növekvő ütemet mutat, az a_t paraméter pedig kisebb intenzitásban, de időben csökkenő értékeket vesz fel. A c_t értékeiben viszont kevésbé mutatkozik a férfiak és nők ábráján hasonlóság, a_t -vel körülbelül párhuzamosan mozognak.

A svájci adatokra számolt paraméterek alakulásában kevésbé vettem észre szabályosságokat, főleg a női megfigyelésekre. A férfiak esetében látható egyfajta tendencia, hogy mekkora értékeket vesz fel egy-egy paraméter, illetve, hogy közöttük milyen reláció dominál.

A tanító időszakokon kiszámolt paramétereket a szerzők szerint VAR segítségével javasolt előre jelezni. Sajnos viszonylag kevés évre rendelkezek megfigyeléssel, emiatt a VAR modell az én esetemben nem adott megbízható eredményeket. Ezért egyszerűbb módszerhez, ARIMA folyamatokhoz fordultam. *Augmented Dickey-Fuller* tesztet hajtottam végre a paraméterek idősorán, amely szerint az idősorok nem bizonyultak stacionereknek. Ennek kiküszöbölése érdekében egyszeres differenciázást használtam. Minden modell esetén kipróbáltam többféle ARIMA folyamatot, melyek közül *Akaike* információs kritériuma (AIC) vizsgálatával választottam ki a legjobbat. A következő folyamatokat alkalmaztam az előrejelzésekhez:

3. Táblázat Alkalmazott ARMA folyamatok

USA férfi	ARIMA (1,1,0)
USA nő	ARIMA (1,1,0)
Svájc férfi	ARIMA (1,1,1)
Svájc nő	ARIMA (1,1,0)

Az így megkapott paraméterekkel és a referencia populációra készített LC előrejelzésekkel megkaptam a SAINT módszere szerinti előrejelzést. A korábbiakhoz hasonlóan MSE mutatóval összehasonlítottam a SAINT előrejelzés értékeit a kihagyott valós értékekkel.

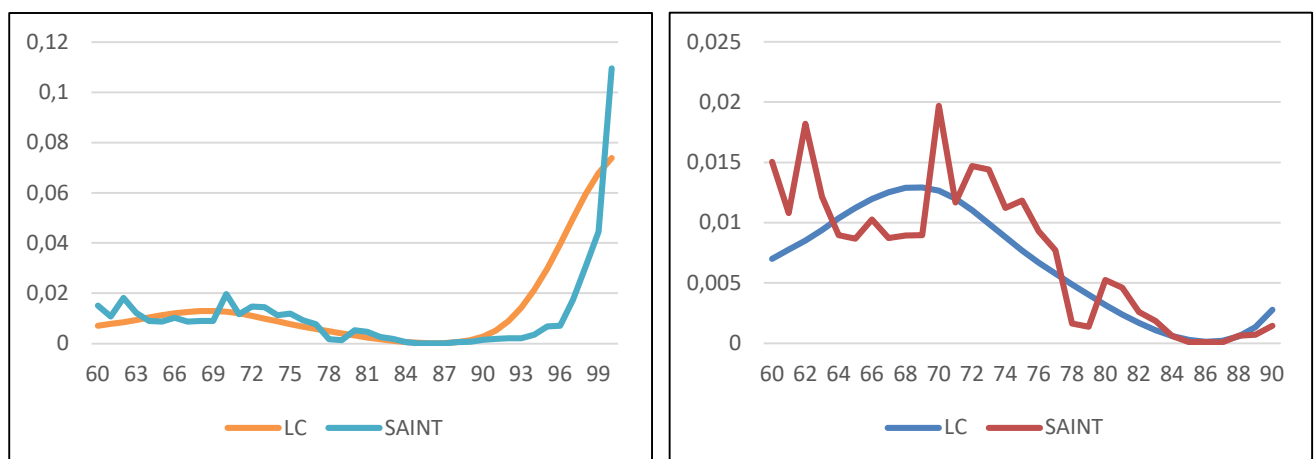
USA: férfi adatok eredménye

Az amerikai adatokból 8 éves tesztidőszakot választottam le az 1959-től 1991-ig tartó megfigyelésekből. A SAINT modell kiértékeléséhez összehasonlítottam a valós értékektől való eltérést a Lee-Carter modell esetén látott eltérésekkel.

4. táblázat A két modell MSE értékei az amerikai férfiakra

	60-100 év	60-90 év
SAINT	0,0111	0,0074
Lee-Carter	0,0139	0,0065

Az átlagos négyzetes hiba, -habár kicsi különbséggel-, de a SAINT féle modellezés alkalmazásánál kisebb értéket vett fel 60 és 100 éves korokra megvizsgálva, így kevesebb tévedéssel adta vissza a valós halandósági valószínűségeket. Szűkítve a vizsgált korintervallumot 60-tól 90 éves korig viszont a Lee-Carter tekinthető jobb modellnek. A következő ábrán kirajzoltam a 8 éves tesztidőszakra koronként az átlagos eltéréseket:



23. ábra Amerikai férfi adatokra modellenként az átlagos eltérés 60-100 éves korban (bal) és 60-90 éves korban (jobb) (saját szerkesztés)

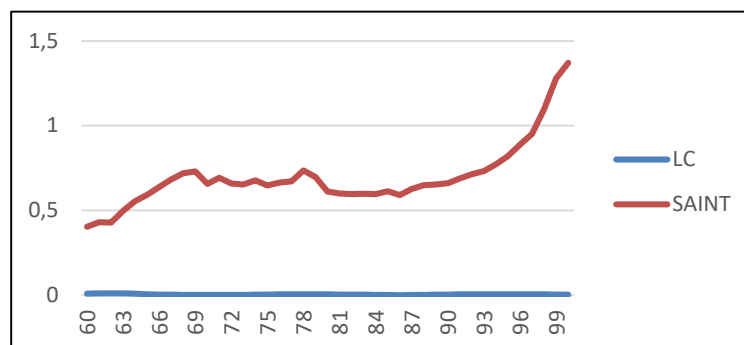
Látható, hogy hullámzik melyik modell teljesít jobban és hogy a felső korhatár felé közeledve mindkét modell esetében nagymértékben megnő az átlagos eltérés. Hogy jobban kirajzolódjanak az eltérések a 30. ábrán 60-tól 90 éves korig láthatjuk. Itt megfigyelhető, hogy valójában több korban is az LC modellnél kisebbek a tévedések. Tehát összességében a SAINT módszer a 64-70, 78-80 és 88 éves kortól teljesít jobban, a többi esetben az LC pontosabb.

USA: női adatok eredménye

Előzőekhez hasonlóan elemeztem ki az amerikai női adatokat is. Itt is láthatjuk az MSE értékeket, azonban a férfiakhoz képest a nőknél határozottan a Lee-Carter modell a pontosabb.

5. táblázat A két modell MSE értékei az amerikai nőkre

Modell	MSE
SAINT	0,695
Lee-Carter	0,003



24. ábra Amerikai női adatokra modellenként az átlagos eltérés 60-100 éves korban (saját szerkesztés)

A korévenkénti átlagos eltérések a SAINT modellenél szinte végig magasabbak 0,5-nél, míg a LC-nél szinte 0 értékeket vesznek fel. Összehasonlítva az amerikai férfiakkal itt sokkal magasabb eltérés értékek lépnek fel. Az amerikai női biztosítotti adatokon tehát egyértelműen a Lee-Carter modellel lehet kevesebb tévedéssel becsülni, azonban indokolt lenne további vizsgálat a többi eredményhez képesti furcsa jelenséget megmagyarázóan.

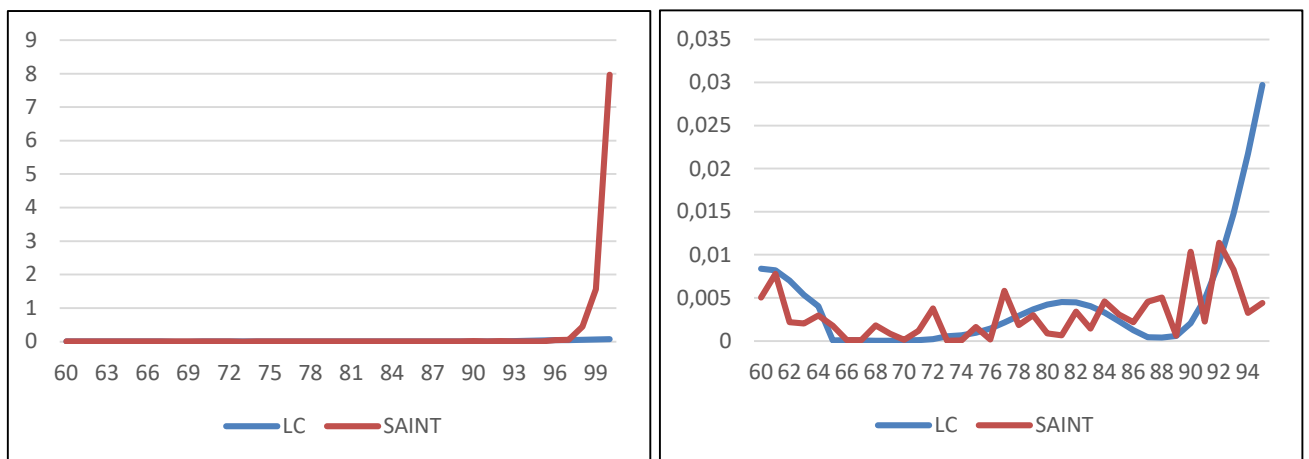
Svájc: férfi adatok eredménye

6. táblázat A két modell MSE értékei a svájci férfiakra

MSE	60-100 év	60-95 év
SAINT	0,248	0,003
Lee-Carter	0,010	0,0042

A svájci adatoknál 10 éves tesztidőszakot választottam le az 1945-től 1995-ig tartó megfigyelésekből. Összehasonlítottam a SAINT és a Lee-Carter módszer becsléseinek eltérését a valós adatoktól.

Férfiakra a SAINT modellnél jóval magasabb MSE értéket kaptam a tesztidőszakra, így sajnos ebben az esetben sem bizonyult jobbnak a Lee-Carternél. A 25. jobboldali ábrán látszik, hogy a mutatót leginkább a felső korhatárnál kiugró eltérés befolyásolta, így megnéztem, hogy rövidebb korintervallumra mi a jellemző. A rövidebb intervallumra is láthatóak az MSE értékek, amelyek alapján viszont már a SAINT modell a pontosabban becslő. A 25. baloldali ábrán megnézhetjük az ehhez tartozó koronkénti átlagos eltéréseket is. Itt megfigyelhetjük, hogy valójában a két modell átlagos eltérései szerint váltakozik mikor melyik a pontosabb. A felső korhatár közeledtével az amerikai férfi adatokhoz hasonlóan itt is megnövekszik a Lee-Carter modell tévedése.



25. ábra Svájci férfi adatokra modellenként az átlagos eltérés 60-100 éves korban (bal) és 60-95 éves korban (jobb) (saját szerkesztés)

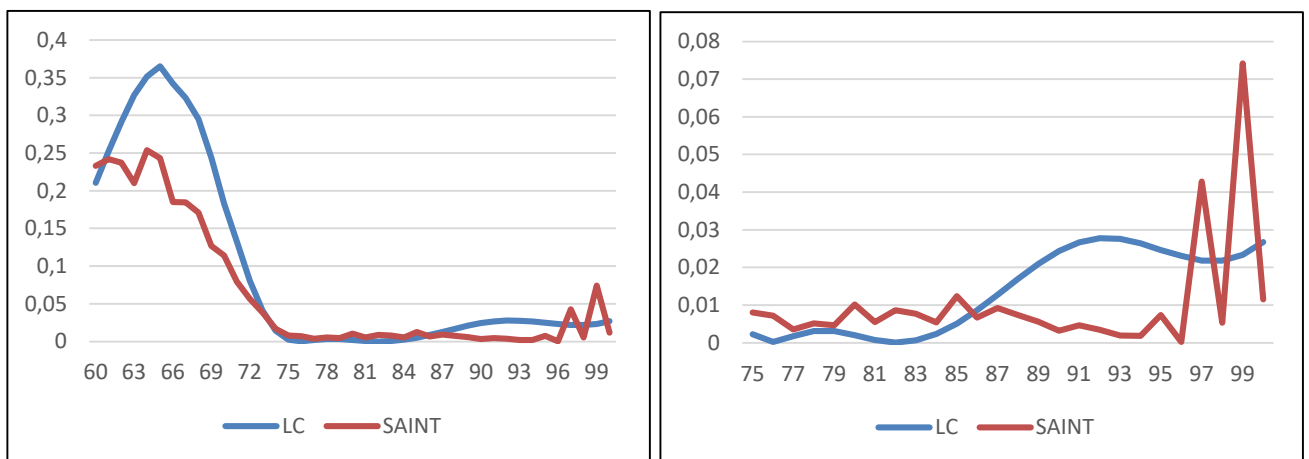
Svájc: női adatok eredménye

A svájci férfi adatokhoz hasonlóan a nők esetén 10 év nagyságú tesztidőszakon néztem meg a modellek becslési képességét.

7. táblázat A két modell MSE értékei a svájci nőkre

MSE	60-100 év	75-100 év
SAINT	0,064	0,01
Lee-Carter	0,092	0,013

A női adatok esetében a SAINT modell rendelkezik jobb becslő erővel a kisebb MSE értéke alapján a 60-tól 100 éves korig terjedő intervallumon. A 75-100 éveseken is elemeztem az átlagos eltérés mutatóját, ami alapján már a korhatár végéhez közeledve kevésbé mondható jobbnak a SAINT az LC modellnél.



26. ábra Svájci női adatokra modellenként az átlagos eltérés 60-100 éves korban (bal) és 75-100 éves korban (jobb) (saját szerkesztés)

Az előző eredményekkel ellentétben érdekes tendenciát figyeltem meg a svájci nők adatainak elemzése során. Míg korábban a felső korcsoportokban láthattunk nagyobb ugrásokat a modellek pontosságában, most az alsó korhatárnál tapasztaltam hasonló jelenségeket. A 75-100 éves korcsoport esetében vettem jobban nagyító alá még az átlagos eltéréseket, és érdekes módon a 75-85 évesek között tapasztaltam a SAINT modell teljesítményének alul maradását. Ez a jelenség tovább fokozódik, ahogy közeledünk a korhatár végéhez.

Modellek alkalmazása díjszámításban

Szeretném röviden bemutatni a mortalitás előrejelzés fontos alkalmazási területét, a járadékárzást. A járadékbiztosítások díjának kiszámításához olyan mortalitási adatokra van szükség, amelyek a feltételezett legmagasabb életkort is tartalmazzák, mely az én esetemben mindkét mintára a 100 éves kor. Kétféleképpen is meghatároztam a díjak mértékét, tehát a dinamikus és a statikus módszerrel is. Statikus halandóság előrejelzéshez periódus halandósági táblát kell alkalmazni, mely egy adott időben élő emberek halandósági valószínűségeit tartalmazza. Dinamikus előrejelzéshez pedig kohorsz tábla használatos, ami már egy adott évben születettek halandósági valószínűségeit tartalmazza, így figyelembe veszi azt is, hogy az idő elteltével a valószínűségek változhatnak. Kohorsz táblát tényadatokkal csak akkor tudnánk elkészíteni, ha az adott kohorsz már kihalt, emiatt fontosak a halandóság-előrejelző technikák, melyekkel megbecsülhetjük a hiányzó értékeket. Mivel ez egy kohorszra mutatja az információkat, így kevésbé torzítottan számolhatunk díjakat, mint periódus táblából. A díjakat az R szoftverbe épített *lifecontingencies* csomag segítségével számoltam. A díjak kiszámításához Banyár (2003) könyve alapján egy évente egy forintot fizető életjáradék egyszeri nettó díjának a képlete a következőként írható fel:

$$\ddot{a}_x = \sum_{i=0}^{100-x} \left(v^i \prod_{k=0}^{i-1} (1 - q_{x+k, T+k}) \right),$$

ahol T az aktuális évszám, v a technikai kamatlábból számolt diszkonttényező és $q_{x,t}$ pedig egy x éves személyre a t . évben érvényes halálozási valószínűség. Technikai kamatlábnak 4 %-ot alkalmaztam, ami a dolgozatom írásának pillanatában az MNB rendelet szerint hatályos hazai maximális mértéke a rendszeres díjas és az 5-évnél hosszabb egyszeri díjas termékek esetében. (MNB, 2015)

A nettó díjszámítást egy 65 éves biztosítottra végeztem el, egy 35 éves tartamú és 100.000 Ft biztosítási összegű szerződést feltételezve. Az általam számolt járadék díját a táblázatban összefoglalva láthatjuk az amerikai és svájci adatokra a periódus tábla szerint, mindkettőnél az utolsó ismert megfigyelési évre:

8. táblázat Járadék díjak periódus táblából

Tábla	Teljes népesség	Férfi népesség	Női népesség	Férfi biztosított	Női biztosított
USA	1 229 407 Ft	1 122 388 Ft	1 316 167 Ft	1 269 673 Ft	1 662 593 Ft
Svájc	1 289 268 Ft	1 175 500 Ft	1 382 076 Ft	1 249 422 Ft	1 385 390 Ft

A periódus tábla alapján kapott értékeknél láthatjuk, hogy a várakozásoknak megfelelően a teljes népességre számított díjak a férfi és a női népességi díjak közé esik. Valamint szintén mutatják a kiszámolt díjak annak hatását, hogy a biztosított populáció kisebb halálozási valószínűségekkel rendelkezik, emiatt magasabb díjat is látunk az ő esetükben, azonban ez az amerikai adatokra nagyobb mértékben feltűnik, svájciak esetében kisebb a differencia. Mind a biztosítói, mind a népességi adatoknál látványos különbséget érzékelhetünk a nemek között, mely a biztosított díjakra még jobban fennáll. Összeségében a két ország díjai szinte azonosak. Ha egy biztosító ezen adatok mentén szabna termékéhez díjat valahol a női és férfi biztosított adatok között lenne, hiszen az EU gender irányelve miatt nem szabhatunk meg nemekre különböző árat.

A járadék kalkulálást elvégeztem kohorsz halandósági tábla alapján is, melynek eredményei láthatók a következő táblázatban szintén mindkét ország adatait felhasználva:

9. táblázat Járadék díjak kohorsz táblából

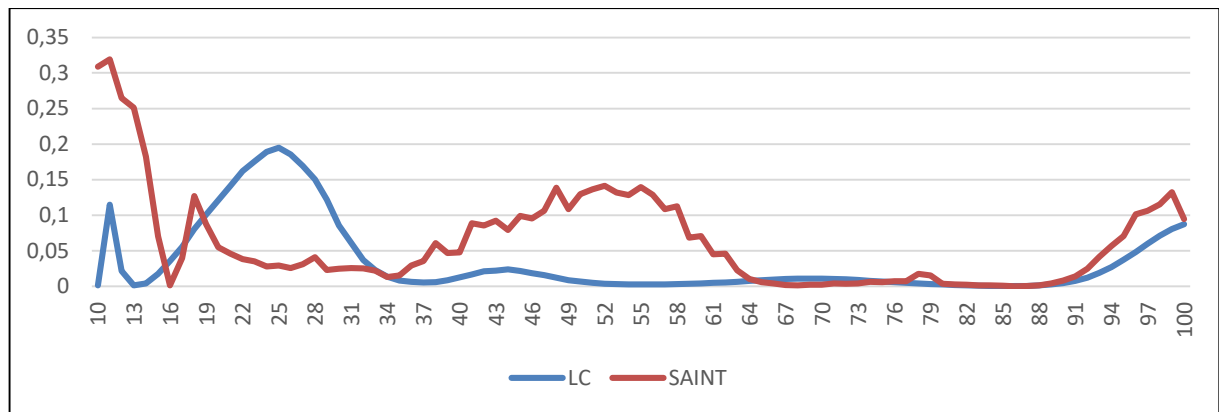
Kohorsz tábla		Teljes népesség	Férfi népesség	Női népesség	Férfi biztosított	Női biztosított
USA	LC	1 274 059 Ft	1 146 546 Ft	1 372 936 Ft	1 332 136 Ft	1 490 328 Ft
USA	SAINT	-----	-----	-----	1 345 783 Ft	1 708 970 Ft
Svájc	LC	1 350 163 Ft	1 213 815 Ft	1 454 398 Ft	1 288 131 Ft	1 432 271 Ft
Svájc	SAINT	-----	-----	-----	1 306 751 Ft	1 422 452 Ft

A kohorsz tábla alapján szinte minden esetben nagyobb díjakat kaptam a periódus táblához képest a várakozásoknak megfelelően, miszerint egy biztosítóra kisebb halandóság jellemző, mint a teljes

népességre. Feltehetően a jobb egészségi állapotok miatt, hiszen a biztosító túl kockázatos egyénekekkel nem fog szerződést kötni. A népességi adatokon itt is a férfínál kisebb, a nőinél pedig nagyobb értéket vesz fel a teljes népességre vonatkozó kalkuláció, mivel a nők magasabb várható élettartama nagyobb kockázatot jelent egy járadékbiztosítási terméknel. Ezeket a Lee-Carter modell segítségével becsült kohorsz tábla alapján számoltam. A biztosított járadékok díját mindkét előrejelző módszerrel meghatároztam. A SAINT modell becsléseiből kapott tábla szerint mindkét nemre magasabb értékek adódtak többségben, mint a Lee-Carterre, azonban visszaemlékezve a korábbi eredményekre, figyelembe kell venni, hogy az amerikaiakra csak a férfiaknál teljesített jobban a SAINT modell. Emiatt a nők oszlopában látható nagy eltérés a torzítás eredménye lehet. A két ország díjai itt is közel azonosak, népességi adatokból következően a svájci népesség túlélési kockázata magasabb.

6. Összegzés

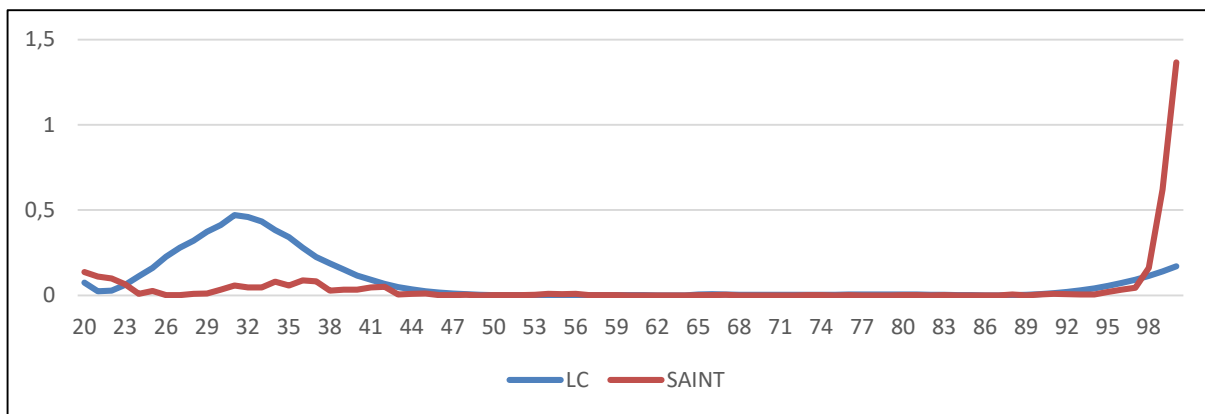
Összeségében láthattuk, hogy teljes mértékben nem teljesített jobban a részpopulációkra a SAINT modell, mint a „benchmark”-nak számító Lee-Carter módszertan. Ha nem is minden modellben, de az amerikai férfiak és a svájci nők esetén kisebb hibával sikerült a részpopulációk referencia populáció segítségével való megbecslése. A vonaldiagrammokon láthattuk, hogy mely különböző korokban teljesített jobban az egyik, illetve a másik modell. Habár elemzésemet csak a 60 éven felüliekre mutattam be részletesen, mivel a járadékbiztosítások számukra releváns kérdés leginkább, azonban megbecsültem a modelleket mindkét országra a rendelkezésemre álló teljes korintervallumon is. A 27. ábrán látható, hogy a SAINT modell a 15-34 éves és a 65-90 éves korokban jobban teljesít, azonban a kettő közötti időszakban a Lee-Carter ad pontosabb becsléseket.



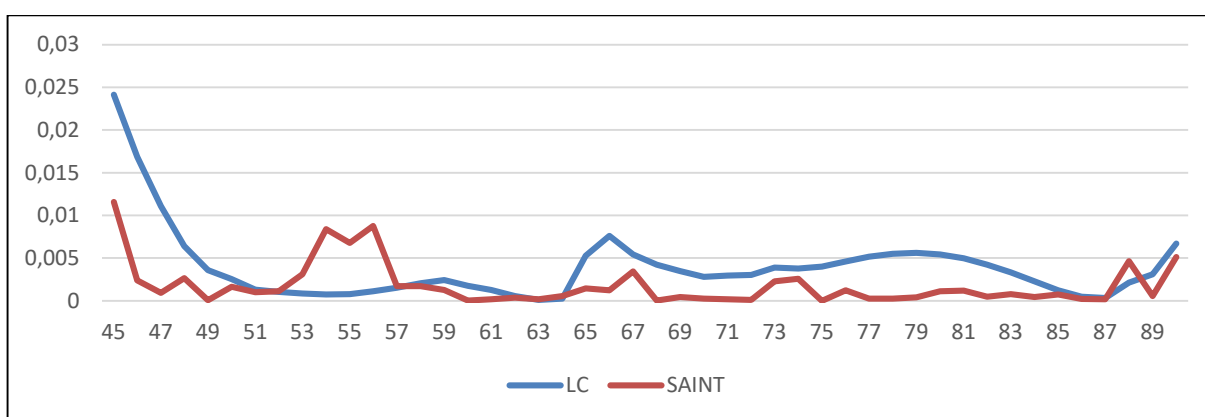
27. ábra Amerikai férfi adatokra modellenként az átlagos eltérés 10-100 éves korban (saját szerkesztés)

Mivel hullámzik, hogy melyik modell is a pontosabb előrejelző, így érdemes lehet akkor alkalmazni az egyes modelleket, ha az adott termék célközönségéhez az bizonyul a pontosabbnak.

A svájci férfiak esetében is ábrázoltam a teljes időszakra vonatkozó átlagos eltéréseket a valós adatokhoz képest. Itt láthatjuk, hogy a fiatalabbakra a Lee-Carter nagyon rosszul tudott becsülni, a korhatár végén pedig a SAINT eredményei szállnak el. Körülbelül 45 és 90 évek közötti időszakban mindkét modell hasonlóan viselkedik ezen ábra alapján. A 29. ábrán kirajzoltam ezt a szűkebb intervallumot is, ahol már jobban fluktuálnak mindkét módszer esetén az eltérések.



28. ábra Svájci férfi adatokra modellenként az átlagos eltérés 20-100 éves korban (saját szerkesztés)



29. ábra Svájci férfi adatokra modellenként az átlagos eltérés 45-90 éves korban (saját szerkesztés)

Végző soron elmondhatjuk, hogy lehet relevanciája részpopulációk halandósága esetén a SAINT módszer alkalmazásának, azonban saját számításaim alapján nem mutat minden esetben jobb illeszkedést. Érdekes lehetne hosszabb időszoron és frissebb adatokon is tesztelni a SAINT modell pontosságát, mivel sajnos számomra csak korlátozott hozzáférés lehetősége adott biztosítotti adatokhoz. Továbbá érdemes lehetne tesztelni a módszert más témájú adatokon is a halandóságon túl.

7. Hivatkozások

- Arató, M. (2001). Nem-életbiztosítási matematika. *Egyetemi tankönyv*.
- Banyár, J. (2003). *Életbiztosítás*. Budapest: Aula Kiadó.
- Basellini, U., Camarda, C. G., & Booth, H. (2023). Thirty years on: A review of the Lee–Carter method for forecasting mortality. *International Journal of Forecasting*, 1033-1049. Forrás: <https://www.sciencedirect.com/science/article/pii/S0169207022001455>
- Booth, H. (2020). Coherent mortality forecasting with standards: Low mortality serves as a guide. *Developments in demographic forecasting*, Springer International Publishing, 153-178.
- Booth, H., Maindonald, J., & Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population Studies*, 325-336.
- Brouhns, N., Denuit, M., & Van Keilegom, I. (2005). Bootstrapping the Poisson log-bilinear model for mortality forecasting. *Scandinavian Actuarial Journal*, 212-224.
- Camarda, C., & Basellini, U. (2021). Smoothing, decomposing and forecasting mortality rates. *European Journal of Population*, 569-602.
- Currie, I. (2006). Smoothing and forecasting mortality rates with P-splines. *Talk given at the Institute of Actuaries*.
- de Jong, P., Tickle, L., & Xu, J. (2016). Coherent modeling of male and female mortality using Lee–Carter in a complex number framework. *Insurance: Mathematics & Economics*, 130-137.
- Haberman, S., & Renshaw, A. (2011). A comparative study of parametric mortality projection models. *48(1)*, 35-55. *Insurance: Mathematics and Economics*. Forrás: <https://www.sciencedirect.com/science/article/abs/pii/S0167668710001022?via%3Dihub>
- Human Mortality Database*. (dátum nélk.). Forrás: <https://www.mortality.org/Data/DataAvailability>
- Human Mortality Database: Methods Protocol*. (2021). Forrás: <https://www.mortality.org/File/GetDocument/Public/Docs/MethodsProtocolV6.pdf>
- Janssen, F., Bardoutsos, A., El Gewily, S., & De Beer, J. (2021). Future life expectancy in Europe taking into account the impact of smoking, obesity, and alcohol. *ELife*.
- Janssen, F., van Wissen, L., & Kunst, A. (2013). Including the smoking epidemic in internationally coherent mortality projections. *Demography*, 1341-1362.
- Kannisto, V., Lauritsen, J., Thatcher, A., & Vaupel, J. (1994). Reductions in mortality at advanced ages: Several decades of evidence from 27 countries. *Population and Development Review*, 793-810.
- Lee, R., & Miller, T. (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography*, 537-549.

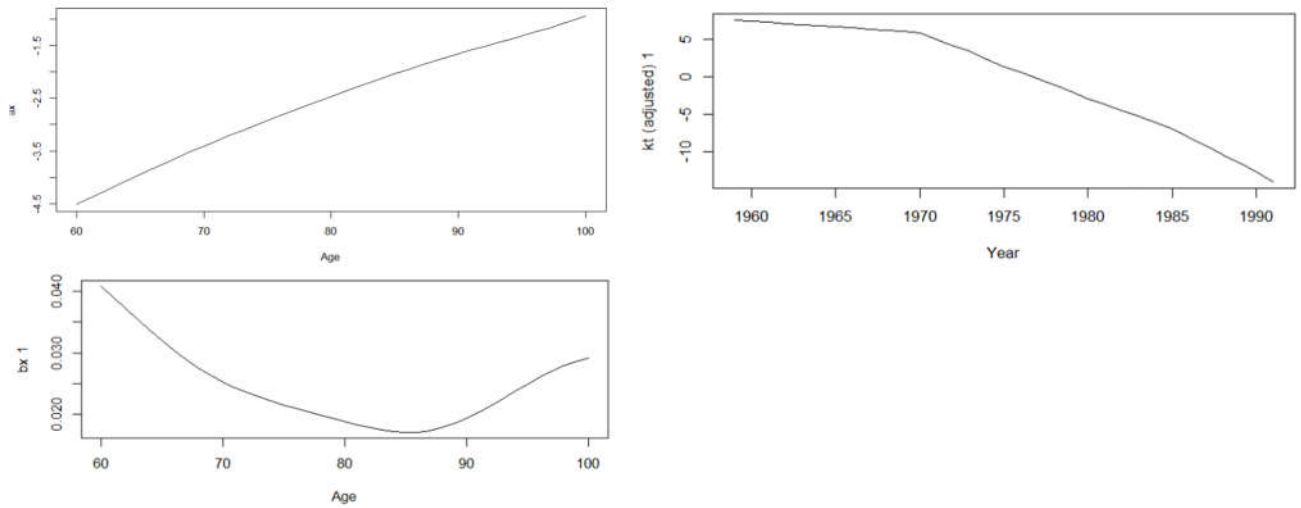
- Li, J.-H., Hardy, M., & Tan, K. (2009). Uncertainty in mortality forecasting: An extension to the classical Lee-Carter approach. *Astin Bulletin*, 137-164.
- Li, N., & Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, 575-594.
- Li, N., Lee, R., & Gerland, P. (dátum nélk.). Extending the Lee-Carter Method to Model the Rotation of Age Patterns of Mortality Decline for Long-Term Projections.
- Lyu, P., De Waegenare, A., & Melenberg, B. (2021). A multi-population approach to forecasting all-cause mortality using cause-of-death mortality data. *North American Actuarial Journal*, 421-456.
- Májér, I., & Dr. Kovács, E. (2011). Élettartam-kockázat – a nyugdíjrendszerre nehezedő egyik teher. *Statisztikai Szemle*(7-8), 790-812.
- Menziatti, M., Morabito, M., & Stranges, M. (2019). Mortality Projections for Small Populations: An Application to the Maltese Elderly. *Risks*, 7, 35.
- MNB. (2015). 54/2015. (XII. 21.) MNB rendelet a technikai kamatláb maximális értékéről.
- Rabbi, A., & Mazzuco, S. (2021). Mortality forecasting with the Lee–Carter method: Adjusting for smoothing and lifespan disparity. *European Journal of Population*, 97-120.
- Rau, R., Soroko, E., Domantas, J., & Vaupel, J. (2008). Continued reductions in mortality at advanced age. *Population and Development Review*, 747-768.
- Renshaw, A., & Haberman, S. (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38(3), 556-570. Forrás: <https://www.sciencedirect.com/science/article/abs/pii/S0167668705001678>
- Society of Actuaries: Mortality and other rate tables.* (dátum nélk.). Forrás: <https://mort.soa.org/Default.aspx>
- Stoeldraijer, L., van Duin, C., van Wissen, L., & Janssen, F. (2018). Comparing strategies for matching mortality forecasts to the most recently observed data: exploring the trade-off between accuracy and robustness. *Genus*, 1-20.
- Szentkereszt, G., & Vékás, P. (2022). Magyar halandósági ráták előrejelzése visszacsatolt neurális hálózatokkal. *Statisztikai Szemle*, 905-922.
- Tolsztoj, L. N. (1886). Ivan Iljics halála. Oroszország. doi:ISBN 963-07-7933-1
- Van Baal, P., Peters, F., Mackenbach, J., & Nusselder, W. (2016). Forecasting differences in life expectancy by education. *Population Studies*, 201-216.
- Vékás, P. (2016). *Az élettartam-kockázat modellezése (doktori értekezés)*. Budapesti Corvinus Egyetem, Általános és Kvantitatív Közgazdaságtan Doktori Iskola.
- Villegas, A., Millossovich, P., & Kaishev, V. (2016). StMoMo: An R Package for Stochastic

Mortality Modeling.

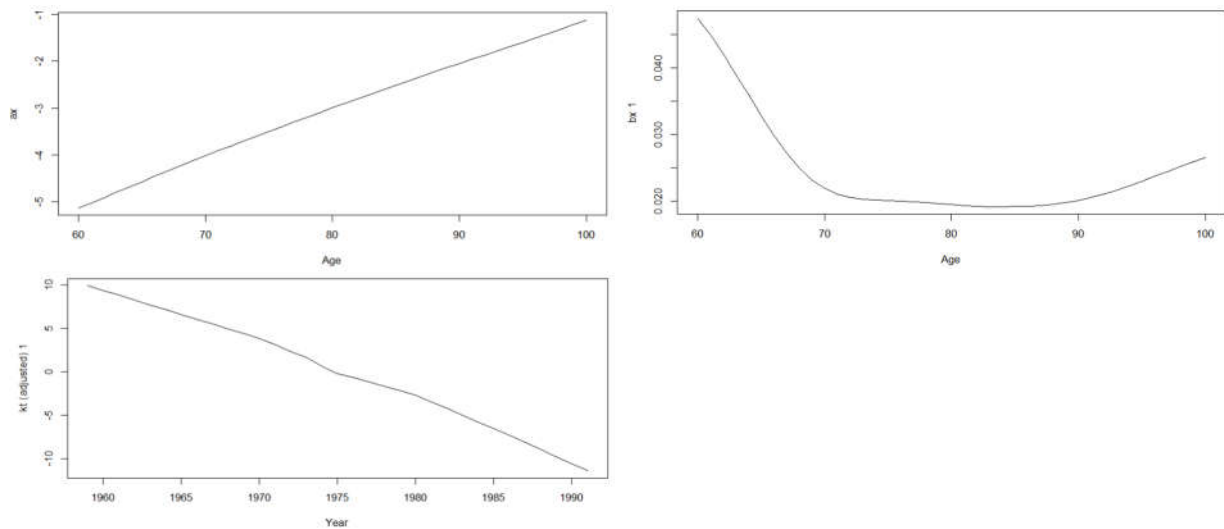
- Wan, C., & Bertschi, L. (2015). Swiss coherent mortality model as a basis for developing longevity de-risking solutions for Swiss pension funds: A practical approach. *Insurance: Mathematics and Economics*, 63, 66-75.
- Wang, H.-C., Yue, C.-S., & Chong, C.-T. (2018). Mortality models and longevity risk for small populations. *Insurance: Mathematics and Economics*, 78, 351-359. Forrás: <https://www.sciencedirect.com/science/article/pii/S0167668717304365?via%3Dihub#sec1>
- Wilmoth, J. (1993). Computational methods for fitting and extrapolating the Lee-Carter model of mortality change. Department of Demography, University of California, Berkeley.

8. Melléklet

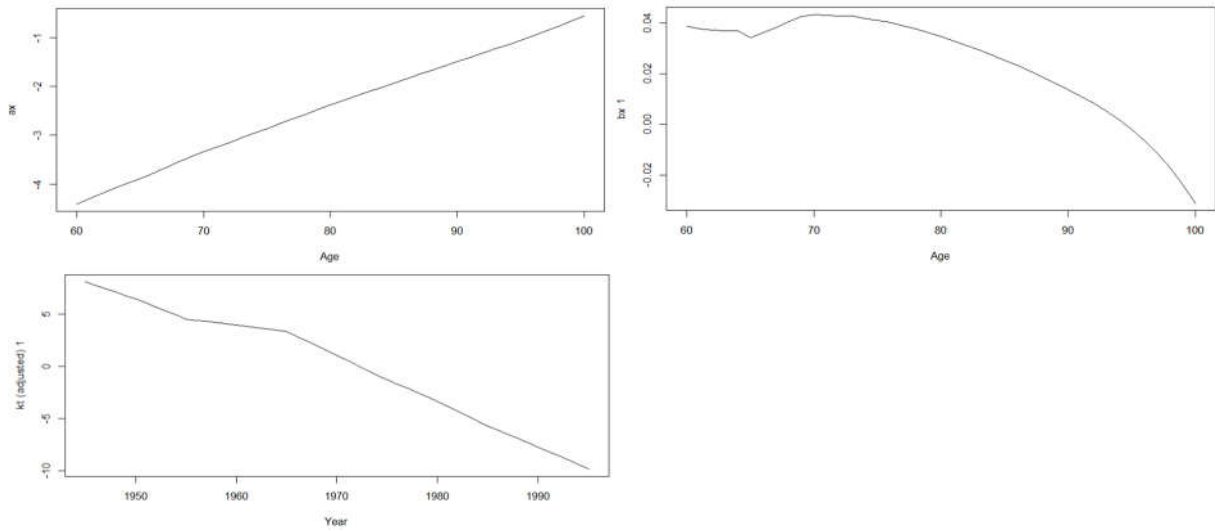
1959-1991 között a férfiak Lee-Carter modelljének paraméterei 60-100 éves kor között az amerikai biztosítotti adatokon



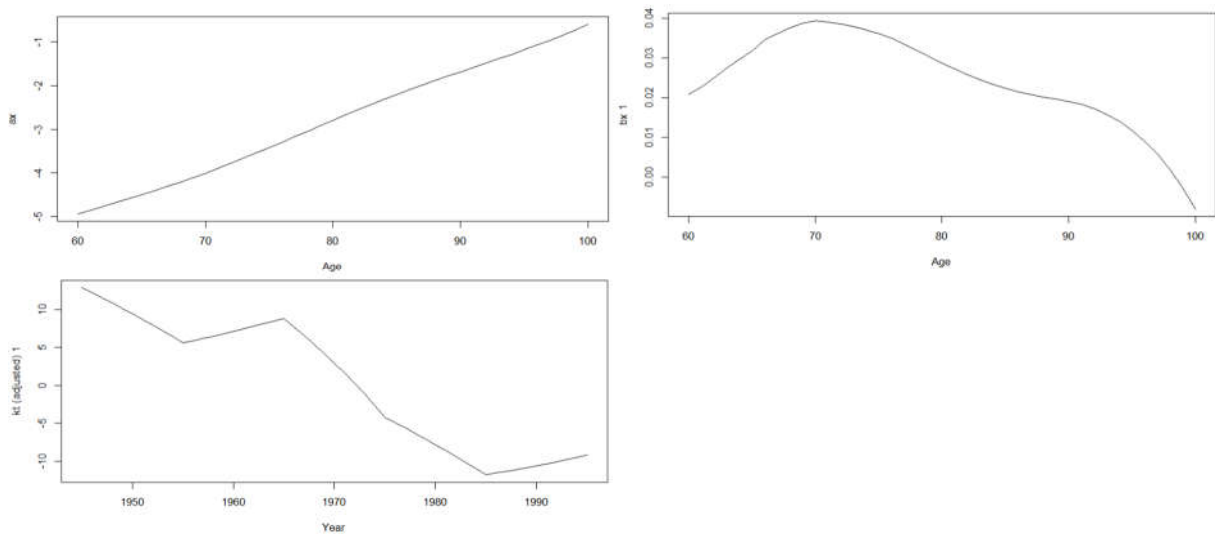
1959-1991 között a nők Lee-Carter modelljének paraméterei 60-100 éves kor között az amerikai biztosítotti adatokon



1945-1995 között a férfiak Lee-Carter modelljének paraméterei 60-100 éves kor között a svájci biztosított adatokon



1945-1995 között a nők Lee-Carter modelljének paraméterei 60-100 éves kor között a svájci biztosított adatokon



NYILATKOZAT

Név: Kelemen Viktória

ELTE Természettudományi Kar, szak:

Biztosítási és pénzügyi matematika MSc

Neptun azonosító: FZ3LV0

Szakedolgozat címe:

A SAINT halandóság-előrejelző modell életbiztosítási alkalmazása

A **szakedolgozat** szerzőjeként fegyelmi felelősségem tudatában kijelentem, hogy a dolgozatom önálló munkám eredménye, saját szellemi termékem, abban a hivatkozások és idézések standard szabályait következetesen alkalmaztam, mások által írt részeket a megfelelő idézés nélkül nem használtam fel.

Budapest, 2024.



a hallgató aláírása