

TDK-dolgozat

Fabók Judit Bíborka



36. ORSZÁGOS
TUDOMÁNYOS DIÁKKÖRI
KONFERENCIA 2023

Magyar halandósági ráták előrejelzése homogén viselkedésű országok trendjei alapján

Tudományos Diákköri Dolgozat

Konzulens:

Kovács László, egyetemi tanársegéd

Fabók Judit Bíborka

Budapesti Corvinus Egyetem

Matematikai és Statisztikai Modellezés Intézet

Statisztika Tanszék

III. évfolyam

Tartalomjegyzék

1. BEVEZETÉS	1
2. HOSSZÚ ÉLETTARTAM-KOCKÁZAT JELENTŐSÉGE A NYUGDÍJRENDSZEREKBE	3
3. LEE-CARTER MODELL BEMUTATÁSA	5
3.1. A modell rövid leírása	5
3.2. A modell változói	5
3.3. Paraméterbecslés	6
3.4. A modell bírálata	7
4. HASONLÓSÁGI MODELL – SAJÁT MÓDSZER ELJÁRÁSÁNAK BEMUTATÁSA	9
4.1. Miért lehet szükséges egy másfajta megközelítés?	9
4.2. A hasonlósági modell módszertana	11
4.3. Illeszkedésvizsgálat	12
5. ALKALMAZOTT KERETRENDSZER BEMUTATÁSA	19
5.1. Adatok ismertetése	19
5.2. Az R szoftverkörnyezete	19
5.3. Magyarországi adatok beolvasása	20
5.4. Egységes adattábla kialakítása hasonlósági modell alkalmazásához	21
5.4.1. Hasonlósági modell implementációja	22
5.4.2. A legjobb ország.....	23
5.4.3. Az előrejelzés kiértékelése	24
5.5. Lee-Carter modell előrejelzése R-ben	26
5.6. Hasonlósági és Lee-Carter modellek összehasonlítása	27
6. EREDMÉNYEK ÉS ÉRTÉKELEÉSÜK.....	29
6.1. A vizsgált, előrejelezni kívánt 3 időszak	29
6.2. Az 1991-1999 közötti időszak előrejelzése	29

6.3.	Az 1991-2017 közötti időszak előrejelzése	32
6.4.	A 2000-2017 közötti időszak előrejelzése.....	34
7.	ÖSSZEFOGLALÁS	37
IRODALOMJEGYZÉK		40
MELLÉKLETEK		42

1. BEVEZETÉS

A XXI. század nyugdíjrendszereinek fenntarthatóságának vizsgálata során megkerülhetetlen jelenség az élettartam-kockázat. A javuló életkörülmények, az egészségügyi ellátás színvonalának emelkedése, az orvostudomány fejlődése jelentős mértékben hozzájárult ahhoz, hogy a halálozási valószínűségek immáron több évtized óta csökkenő tendenciát mutatnak. Napjainkban a korábbiakhoz képest eddig még soha nem tapasztalt élettartamnövekedés figyelhető meg, így a hosszú élettartam-kockázat kérdésköre egyre nagyobb prioritású problémát jelent, mind társadalmi, mind gazdasági szempontból (Májor – Kovács, 2011).

A fejlett országok előregedő népessége miatt kiemelkedően fontos a megfelelő nyugdíjrendszer kialakítása. A felosztó-kirovó elven működő állami nyugdíjpénztárak, nyugdíjrendszerek statikus halandósági táblák adataira támaszkodva határozzák meg a várható élettartamokat. Ezen táblák azonban, az élettartam-kockázat miatt, csak elég pontatlanul jelezhetik előre a jövő trendjeit, hiszen nem veszik figyelembe a halálozási valószínűségek folyamatos csökkenését. Következésképpen a pénzügyi intézmények jelentősen alábecsülik jövőbeli kifizetéseiket. Mindez komoly pénzügyi kockázatot jelent majd a nyugdíjrendszer fizetőképességének, fenntarthatóságának szempontjából (Vékás, 2016).

A probléma megoldására léteznek matematikai modellek. Az egyik legelterjedtebb és legismertebb ezek közül Lee és Carter (1992) modellje, mely matematikai-statisztikai alapon képes a halandósági ráták előrejelzésére. A Lee-Carter modellt a szerzőpáros alapvetően az USA néphalandóságának előrejelzésére fejlesztette ki. A modell algebrai és idősorelemzési technikák segítségével jelzi előre a mortalitási rátákat. A szerzők a pontosság érdekében, a mortalitás időbeli javulását leíró paraméter újrabecslését javasolják, melyet autoregresszív integrált mozgóátlagolású (Autoregressive Integrated Moving Average, ARIMA) modell segítségével jeleznek előre. A Lee-Carter modell a legszélesebb körben alkalmazott módszerek egyike halandósági ráták előrejelzésére, népszerűségét elsősorban paramétereinek jó interpretálhatósága adja.

Ugyanakkor számtalan kutatás, tanulmány bírálja a módszer előfeltételeit és javasolja a modell módosítását, valamint annak kiterjesztését. A modell egyik legfőbb bírálata, hogy a halandóságjavulás mértékét a modell időben változatlan paraméterként kezeli. A probléma

megoldására egyes szerzők (Carter – Prskawetz, 2001; Li – Gerland, 2011) a halandóságjavulási paraméter időbeli változását vezették be, mely jelenségre szokás rotációként hivatkozni.

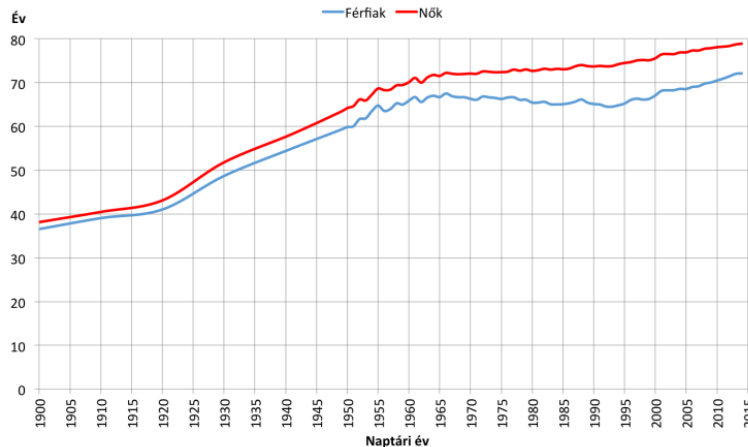
Látható, hogy bár egy mai napig használatban lévő módszerről beszélünk, alkalmazása nem minden esetben jelent hibátlan megoldást. A modell hiányosságaira léteznek részleges korrekciók, melyek segítségével javítható a modell működése, pontossága, ám ugyanakkor a módszer egyre nagyobb volumenű kiterjesztésével párhuzamosan válik egyre bonyolultabbá az eljárás.

Kutatásom célja egy olyan modell kialakítása, mely a Lee-Carter módszerrel szemben pontosabban tudja előrejelezni a magyar halandósági értékeket, azok jövőbeli alakulását. Továbbá, vizsgálom azon hipotézisemet, miszerint egy fejlődő állam mortalitási rátáinak jövőbeli alakulása előrejelezhető egy fejlett ország múltjából. A feltételezés hátterében az áll, hogy várható élettartam tekintetében alapvetően megkülönböztetünk fejlett és fejlődő államokat. Magyarország és a közép-kelet-európai régió ez utóbbi kategóriába sorolható. Fejlődő országok esetében felmerül annak a gondolata, hogy a mortalitási rátáik alakulása konvergál egy fejlett, nyugati ország mortalitása által már korábban bejárt pályához. Ezen logika alapján megvizsgálható, hogy melyik fejlett ország múltbéli halandósága hasonlít leginkább a jelenlegi magyar értékekhez, és a fejlett ország múltbéli alakulása feltételezhető-e a magyarországi mortalitás jövőjének.

2. HOSSZÚ ÉLETTARTAM-KOCKÁZAT JELENTŐSÉGE A NYUGDÍJRENDSZEREKBEN

A 2. fejezetben Májer és Kovács (2011) tanulmánya alapján ismertetem a hosszú élettartam-kockázatot és annak jelentőségét, hogy miért fontos a nyugdíjrendszerek, nyugdíjpénztárak fenntarthatóságának és fizetőképességének szempontjából.

A születéskor várható átlagos élettartam emelkedése Magyarországon nem volt mindig folyamatos. A nyugat-európai államokhoz képest az ország állandó lemaradása figyelhető meg várható élettartam tekintetében. A lemaradás ellenére is számíthatunk az élettartam növekedésének gyorsulására, mely emelkedést az 1. ábra szemléltet.



1. ábra: Nemenkénti születéskor várható élettartamok Magyarországon
Forrás: Vékás, 2016

A várható élettartam növekedését érdemes nyugdíjak számításánál, előretervezésénél figyelembe venni. Ennek oka, hogy a jelenlegi magyar állami nyugdíjrendszerben szolgáltatással meghatározott nyugdíjat állapítanak meg, mely elsősorban a ledolgozott évek számától függ. Ennek következtében az egyre hosszabbodó élettartam-kockázat teljes mértékben a társadalomra hárul. A nyugdíjbiztosítók és -pénztárak, bár más elvek alapján működnek, mivel alapvetően befizetéseket határoznak meg előre, működésük stabil alapfeltétele a be- és kifizetések közötti egyensúly. Ezért számukra is kiemelkedő fontossággal bír a különböző életkorokban vett várható élettartamok minél pontosabb előrejelzése.

A nyugdíjbiztosítók alapvetően 3 különböző, mortalitáshoz köthető kockázatot különböztetnek meg kalkulációik során: egyéni halálozási, várható élettartam- és modellkockázatot. Fontos

megjegyezni, hogy a várható élettartam-kockázat, szemben az egyéni halálozási kockázattal, nem diverzifikálható. A várható élettartam-kockázat arra utal, hogy a halandóság és a várható élettartam hosszú távon el fog térni annak előrejelzett mértékétől, trendjétől. További problémát okoz, hogy egyes halálozási kockázatokkal (például egyéves halálozási valószínűségek) szemben a veszélyközösség elve nem tudja megszüntetni a várható élettartam-kockázatot. Veszélyközösségnek nevezzük a homogén, ugyanazon veszéllyel szembenező csoportot, mely tagjainak száma kellően nagy ahhoz, hogy megvalósuljon a kockázatkiegyenlítődés (Bányár, 2016).

A hosszú élettartam-kockázat (longevity risk) annak kockázata, hogy nem tudjuk pontosan meghatározni a várható élettartam növekedését. Ennek oka, hogy a jövőbeli túlélési valószínűségek változása részben véletlenszerűen alakul. Az élettartam-kockázat kifejezés alapvetően a halandóságjavulási folyamat erősen sztochasztikus jellegére utal, melyet még komplex matematikai-statisztikai eszközök segítségével is nehéz megragadni, egyszerűen múltbeli adatokra hagyatkozva.

A fejlett országok előregedő népessége miatt kiemelkedően fontos a megfelelő nyugdíjrendszer kialakítása. A felosztó-kirovó elven működő állami nyugdíjpénztárak, nyugdíjrendszerek jellemzően statikus halandósági táblák adataira támaszkodva határozzák meg a várható élettartamokat. Ezen táblák azonban, az élettartam-kockázat miatt, csak elég pontatlanul jelezhetik előre a jövő trendjeit, hiszen nem veszik figyelembe a halálozási valószínűségek folyamatos csökkenését. Következésképpen a pénzügyi intézmények jelentősen alábecsülik jövőbeli kifizetéseiket. Mindez komoly pénzügyi kockázatot jelent majd a nyugdíjrendszer fizetőképességének, fenntarthatóságának szempontjából. Így a hosszú élettartam-kockázat napjainkra szükségessé teszi a modern, halandóság – előrejelző modellek alkalmazását (Vékás, 2016).

3. LEE-CARTER MODELL BEMUTATÁSA

3.1. A modell rövid leírása

A halandósági táblák előrejelzésének egyik legismertebb és legelterjedtebb módszere a Lee-Carter modell. A módszer matematikai-statisztikai alapú, úgynevezett extrapolatív eljárás, melynek lényege, hogy a múltban megfigyelt trendek meghosszabbításával jelzi előre a várható mortalitási rátákat, a változások háttérében álló (orvostudományi, egészségügyi, társadalmi) okok figyelembevétele nélkül (Vékás, 2016).

A modell naptári időszakról és életkortól függő paraméterekkel dolgozik. A naptári évtől függő paraméterek, azaz mortalitási indexek újrabecslésével nagyobb pontossággal lehet meghatározni a halandósági rátákat. Mindezt azon feltevés teljesülése mellett, hogy a modell alapján várt és a tényleges halálesetek számai megegyeznek (Vékás, 2016). Ez a feltevés természetesen nem teljesül maradéktalanul, máskülönben a modell hibája 0 lenne.

A továbbiakban Lee és Carter (1992) szerzőpáros munkája alapján ismertetem a kutatásom során benchmarkként használt Lee-Carter modellt.

3.2. A modell változói

A Lee-Carter modell feltételezi, hogy minden x korcsoportra és t időszakra ismertek egy populáció $m_{x,t}$ korcsoport és időszakfüggő központi halandósági rátái. A modell alapegyenlete, a_x és b_x életkor-, k_t időszakfüggő és a 0 várható értékű, $\varepsilon_{x,t}$ hibatag függvényében, a (3.2.1.) képlet alapján írható fel:

$$\ln(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t} \quad (3.2.1.)$$

Az egyenletben a_x jelöli az átlagos log-mortalitást az adott életkorban, a naptári évek átlagában. A k_t az úgynevezett mortalitási index, mely a halandósági szint időbeli javulását, alakulását reprezentálja, kortól függetlenül. Ez az egyetlen időtől függő változó az egyenletben. A b_x életkorfüggő érzékenységi mutató, azt mutatja meg, hogy mennyire érzékeny a logmortalitási ráta az egyes életkorokban a halandóság általános javulására. Azaz megadja, hogyan reagál $\ln(m_{x,t})$ a k_t változására. Bár a tanulmányban nem szerepel kikötésként, szokás az $\varepsilon_{x,t}$ hibatagokra független, 0 várható értékű, normális eloszlású valószínűségi változókként tekinteni.

A (3.2.1.) egyenlet logaritmikus transzformációját alapvetően két körülmény indokolja: egyrészt, ezáltal biztosítható, hogy a modellben nem fog előfordulni negatív becsült mortalitási ráta. Másrészt, a különböző korcsoportok és életkorokhoz tartozó halandósági ráták nagyságrendekkel is eltérhetnek egymástól, ezen eltérések logaritmizált értékei közötti eltérések azonban sokkal jobban kiszámíthatóak.

3.3. Paraméterbecslés

A modell paramétereinek becslése a legkisebb négyzetek módszerével történik:

$$\min_{\alpha, b, k} \sum_{x,t} (\ln(m_{x,t}) - a_x + b_x k_t + \varepsilon_{x,t})^2 \quad (3.3.1.)$$

A továbbiakban az egyszerűbb jelölés érdekében érdemes bevezetni az $\tilde{m}_{x,t}$ jelölést a centrált logaritmikus központi halandósági rátákra:

$$\tilde{m}_{x,t} = \ln(m_{x,t}) - a_x \quad (3.3.2.)$$

Látható, hogy a (3.3.1.) – ben felírt egyenletrendszer aluldefiniált. Az a_x értékét a mátrix x – edik sorának átlagaként megválasztva az $\tilde{m}_{x,t}$ mátrixból szingulárisérték-felbontás (Singular Value Decomposition, SVD) segítségével végtelen számú megoldás adódna b_x és k_t becsült értékére. Ezen probléma megoldására a (3.3.3.) -ban adott két megkötést bevezetve

$$\sum_{x=1}^T b_x = 1 \text{ és } \sum_{t=1}^T k_t = 0 \quad (3.3.3.)$$

b_x és k_t paraméterek becslése már egyértelműen megvalósítható, továbbá ezen két megkötésből következik, hogy a_x az x életkorhoz tartozó logaritmált halandósági ráták hosszmetzeti átlaga.

Ezt követően egyetlen lépés maradt csak hátra, a k_t paraméter kiigazítása. A korrekció szükségességét egyrészt a mortalitási ráta kiszámításánál keletkező információveszteség indokolja, a becslés ugyanis egyenlő súllyal veszi figyelembe az egyes korcsoportokat, figyelmen kívül hagyva azok kitettségét, azaz éveleji népességét és haláleseteinek a számát. További problémát jelent az a tény, hogy maguk a k_t értékek alapvetően nem magának a mortalitási rátának, hanem annak logaritmusának becslési hibáját minimalizálják.

A kiigazított k_t mortalitási indexek a

$$\sum_{x=1}^N D_{x,t} = \sum_{x=1}^N E_{x,t}^c e^{a_x + b_x k_t} \quad (3.3.4.)$$

egyenletrendszer segítségével határozhatóak meg, ahol az a_x és b_x becsült paramétereket adottnak tekintjük az SVD felbontás eredménye alapján. A $D_{x,t}$ jelöli a t – edik évben az x – edik életévükben elhunytak számát, $E_{x,t}$ pedig ugyanezen egyedek kitettségét. Látható, hogy a megfelelően korrigált k_t értékeket behelyettesítve a (3.3.4.) képletbe, az egyes időszakokban megfigyelt halálozási számok pontosan megegyeznek majd a modell által becsült halálesetek értékeivel.

Ezt követően az előrejelzéshez elsősorban k_t paramétert jelezzük előre egy idősoros modell segítségével, melyet lehet átlagos növekedési ütemmel, lineáris/exponenciális trenddel vagy ARIMA modellekkel. Az előrejelzett k_t értékeket visszaírva a modell (3.2.1.) alapegyenletébe, behelyettesítve a_x és b_x paramétereket, kiszámíthatóak az előrejelzett $m_{x,t}$ központi halandósági ráták.

3.4.A modell bírálata

A Lee-Carter modell a legszélesebb körben alkalmazott módszerek egyike halandósági ráták előrejelzésére. Ugyanakkor számtalan kutatás, tanulmány bírálja a módszer előfeltételeit. Az eljárás egyik kiemelkedő megszorítása, hogy az életkorfüggő érzékenységet (b_x) időben változatlan paraméterként kezeli. Elemzések igazolták, hogy ezen feltevés hosszú távon nem felel meg a gyakorlatban tapasztaltakkal, a halandóságjavulás mértéke ugyanis eltérő fiatalabb, illetve idősebb korosztályok esetén. Ez a jelenség a rotáció: a halandóságjavulás egyre inkább az idősebb korosztályokat érinti az idő előrehaladtával, mint a fiatalokat. Azaz, rotáció történik időben a halandóság javulásában az idősebb korosztályok felé. A probléma kiküszöbölésére több javaslat is született, többek között Carter és Prskawetz (2001) halandóságjavulási paraméter rotációja, vagy éppen Li és Gerland (2011) robusztus rotációs Lee-Carter modellje.

Az egyik legkörültekintőbb változata az eredeti modellnek Li, Lee és Gerland (2013) által bevezetett, rotációval kiterjesztett Lee-Carter modell. A módszer lényege, hogy ezen megoldásban az időváltozó nemcsak a k_t mortalitási indexben, hanem a b_x életkorfüggő halandóságjavulási rátában is jelen van. Ezáltal nemcsak a halandóságjavulás, hanem egyúttal a rotáció üteme is beépül a modellbe.

Összességében elmondható, hogy bár egy széles körben, mai napig használatban lévő módszerről beszélünk, alkalmazása nem minden esetben jelent hibátlan megoldást. Hiányosságaira, mint láthattuk, léteznek részleges korrekciók, melyek segítségével javítható a modell működése,

pontossága, ám ugyanakkor a módszer egyre nagyobb volumenű kiterjesztésével párhuzamosan válik egyre bonyolultabbá az eljárás.

További érdekességként megemlítendő, hogy a legtöbb témával foglalkozó tanulmány elsősorban gazdaságilag, társadalmilag fejlett országok esetében számolt be arról, hogy a modell alkalmazása helyesen működött. Wang (2007) esettanulmányában eredményesnek találja a modell predikcióit a svédországi adatok elemzése során. A G7 országok esetében készült tanulmány igazolta, hogy ezekben az országokban is helytállóan működik a Lee-Carter modell (Tuljapurkar, Li, – Boe, 2000). A Lee-Carter egyik kiterjesztett változatának alkalmazásával Hyndman és Ullah (2007) szintén sikeres előrejelzésekről számolnak be, francia, illetve ausztrál adatokra alkalmazva a módszert. Ezzel szemben jelen dolgozat írásakor nem tudtunk azonosítani olyan szakirodalmat, mely hasonló tapasztalatokról írna fejlődő, elmaradottabb államok tekintetében.

Fenti gondolatmenetből következő adódik az az elképzelés, hogy a jelenség nem véletlen, háttérében a fejlett és fejlődő országok közötti alapvető társadalmi, gazdasági, egészségügyi különbségek állhatnak, melynek következtében az elmaradottabb államok esetén a Lee-Carter modell előrejelzése pontatlan lenne. Így felmerül annak a gondolata, hogy ezen országok esetében érdemes lenne másfajta módszert alkalmazni a halandósági ráták előrejelzésére, modellezésére.

4. HASONLÓSÁGI MODELL – SAJÁT MÓDSZER ELJÁRÁSÁNAK BEMUTATÁSA

A 3. fejezetben bemutatott, gyakorlatban leginkább alkalmazott Lee-Carter módszer ismertetését követően kitérek arra, hogy miért, milyen esetekben érdemes másik módszertant választani, ehhez milyen eljárás lehet megfelelő. Továbbá, bemutatom a saját módszerem elméleti háttérét adó statisztikai próba felépítését, működését.

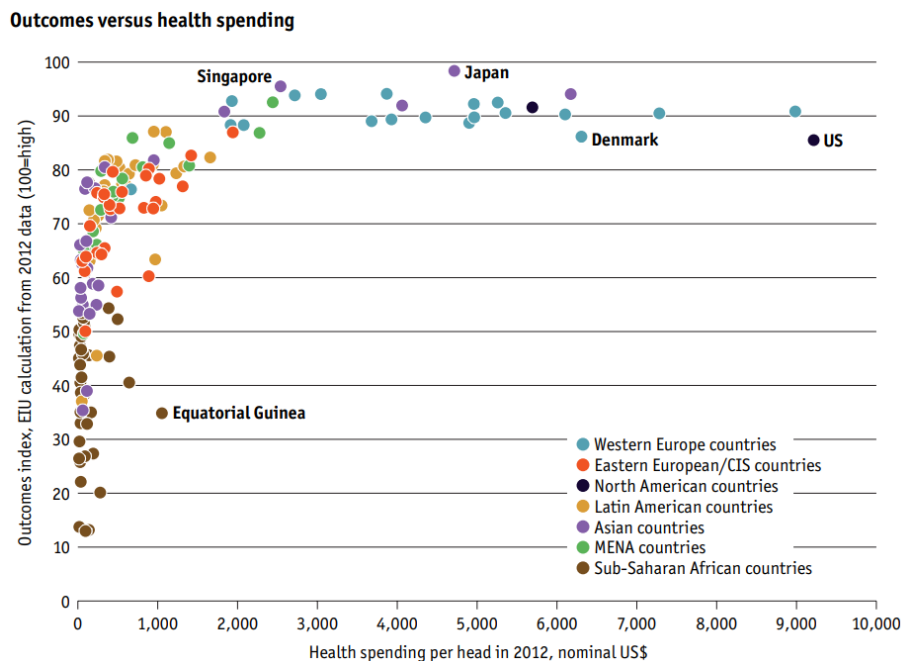
4.1. Miért lehet szükséges egy másfajta megközelítés?

Több tanulmány is arról számol be, hogy Lee-Carter modellel, illetve annak kiterjesztett változataival végzett elemzések, előrejelzések meglehetősen meggyőző eredményeket hoztak: Tuljapurkar és munkatársai (2000) tanulmánya a G7 országok adataira alkalmazva a modellt helytállónak találta azon feltételezést, miszerint a mortalitási index lineárisan csökken. Nagy-Britannia népességének alakulását a modell kiterjesztett változatának és bayes-i modellkeret segítségével jelzik előre (Wiśniowski, Smith, Bijak, Raymer, – Forster, 2015), valamint az Egyesült Államok Népszámlálási Hivatala (U.S. Census Bureau) is ezt a módszert használja népességelőrejelzés céljából.

Kicsit figyelmesebben elolvasva a felsorolt példákat, megfigyelhető egy közös jellemző, nevezetesen, hogy valamennyi esetben fejlett, nyugati államokról van szó. Jelen dolgozat írásakor nem tudtunk azonosítani olyan szakirodalmat, mely fejlődő országok mortalitási rátáinak Lee-Carter modell alapján történő előrejelzéséről írna. Az eltérés háttérében a fejlett és fejlődő országok közötti különbségek állhatnak, mind várható élettartam, mind halandósági ráta tekintetében.

Az *Economist Intelligence Unit* egyik esettanulmányában arról ír, hogy igen erős korreláció figyelhető meg az egyes országok egy főre eső egészségügyi kiadásai és egy, általa kalkulált eredményindex között. Ezt az eredményindexet alapvetően az egészségesen várható élettartam és felnőtt halálozás alapján határozza meg. A 2. ábra alapján látható, hogy főleg a nyugat-európai országok, valamint fejlett ázsiai államok (Szingapúr, Japán) költenek sokat az egészségügyre. Ebből kifolyólag ezekben az országokban magasabb a várható élettartam és alacsonyabbak a halálozási ráták (The Economist Intelligence Unit, 2014). Érdemes megfigyelni a kelet-európai országok elhelyezkedését is a 2. ábrán. Szemmel láthatóan elmaradnak a fejlett országoktól, hiszen

amellett, hogy kevesebbet költenek az egészségügyre, eredményindexük is kisebb. Azaz alacsonyabb a várható élettartam és magasabbak a mortalitási ráták.



2. ábra: Az egy főre eső egészségügyi kiadások és eredményindexek

Forrás: The Economist Intelligence Unit, 2014

Ennek értelmében a fejlett, élenjáró országok esetén egészen jól alkalmazható a Lee-Carter modell, mindennemű hibájával együtt. Hiszen mindaz, amit a hosszú élet kockázata feltételez, ezen országokban maradéktalanul teljesülni látszik, kezdve a fejlett egészségüggyel, egészen a javuló életkörülményekig.

Ugyanakkor az is érthető, hogy azon államok mortalitási rátái, melyek fejletlenebbek (kezdetleges egészségügy, alacsonyabb életszínvonal jellemző), elmaradottságukból adódóan másként fognak alakulni, nem úgy, mint egy fejlett országban. Továbbá, ha az adott fejlődő országban egy hirtelen fejlődés következik be az egészségügyben, akkor ott gyorsabb fejlődés történik majd, mint amit a Lee-Carter modell becsül. Ebből kifolyólag arra a következtetésre juthatunk, hogy fejlődő országok esetében érdemes lenne másfajta módszert választani halandósági ráták előrejelzése céljából.

Tekintve, hogy fejlődő országokról beszélünk, feltételezhető, hogy ezek idővel a fejlett, nyugati országok színvonalához fejlődnek fel. Ezt a gondolatmenetet folytatva, ha egy fejlődő ország fejlettsége az idő múlásával folyamatosan tart egy fejlett állam színvonalához, akkor ebből implicit módon következik, hogy gazdasági, társadalmi viszonyok mellett demográfiai szempontból is

közelít egy fejlett ország szintjéhez. Ezen okfejtést követve, felmerül annak a gondolata, hogy a fejlődő országok mortalitási rátáinak alakulása konvergál egy fejlett, nyugati állam mortalitása által már korábban bejárt pályához. Ezek alapján egy fejlett állam halandósági múltja megfeleltethető egy fejlődő ország jelenjének, és a fejlett ország múltbéli halandósági rátáinak alakulásának segítségével előrejelezhető a fejlődő állam mortalitása.

Ezesetben olyan megoldásra lenne szükségünk, mely képes arra, hogy megkeresse, a múltbéli halandósági táblák alapján melyik ország mortalitása illeszkedik legjobban egy adott fejlődő ország jelenlegi értékeihez, és annak az országnak a halandósági rátáinak alakulásából előrejelezni a fejlődő ország jövőjét. Látható, hogy ebben az esetben egyfajta hasonlósági modellt kellene alkalmazni, majd ez alapján az előrejelzést megvalósítani. Az általam felépített hasonlósági modell pontosan ezt teszi, melynek struktúráját, működését a 4.2.-es alfejezetben ismertetem.

4.2.A hasonlósági modell módszertana

A modell felépítése előtt alapvetően két paraméter értékéről kell döntenünk: y évről, amelytől kezdődően majd szeretnénk előrejelezni, valamint a p időszakok értékéről, azaz, hogy mennyi évre szeretnénk az előrejelzést elkészíteni. Ezen két paraméter megválasztása után kezdhethetünk hozzá a modellépítéshez.

A feladat az, hogy találjunk egy olyan országot, melynek múltbéli halandósági tábla adatai megfelelően illeszkednek egy másik (fejlődő) ország jelenlegi értékeihez. Ha sikerül ilyen államot találni, akkor feltételezhető, hogy a fejlett állam későbbi éveinek alakulása megfeleltethető az elmaradottabb ország jövőjének, mely értékeket így tekinthetjük a fejlődő állam előrejelzett mortalitási rátáinak.

Ha y évből szeretnénk előrejelezni p időszakot, akkor a fejlődő országnak az $y-1$ – edik évét fogom viszonyítási alapnak venni (hiszen az előrejelzés y évtől indul). Ehhez kell keresnem olyan fejlett országot, melynek múltbéli halandósági rátái kor függvényében homogénnek vehetők a fejlődő ország $y-1$ – edik évi értékeihez. Ha a múltbéli adatokból p időszakot szeretnénk előrejelezni, az azt jelenti, hogy a fejlett ország múltja közül az $y-1-p$ – edik évet fogom kiindulási alapnak tekinteni.

Ezen a ponton ismert, hogy melyik fejlődő ország melyik évéből mennyi időszakra szeretnénk előrejelezni, és ehhez melyik múltbéli év halandósági táblájára van szükségem egy fejlett országból (adott paraméterek mellett). Egyetlen feladat maradt csupán, annak meghatározása, hogy a fejlett

országok közül, az adott $y-1-p$ – edik évben melyik állam halandósági rátái hasonlítanak leginkább a fejlődő ország $y-1$ – edik évéhez. A kérdés megválaszolásához pedig a statisztikai eszköztár egyik módszeréhez folyamodunk, nevezetesen az illeszkedésvizsgálathoz.

4.3. Illeszkedésvizsgálat

Az illeszkedésvizsgálat a statisztikai nagymintás nemparaméteres próbák egyike. Az illeszkedésvizsgálat célja, hogy ellenőrizzük egy alapsokaság eloszlására tett hipotézisünket. A próbafüggvény értékének meghatározása a

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - f_i^*)^2}{f_i^*} \quad (4.3.1.)$$

képlet segítségével történik. A (4.3.1.) képletben az f_i jelöli a tényleges, míg f_i^* az elméleti, feltételezett megoszlásokat. Az eloszlás jobboldali χ^2 – eloszlást követ, szabadságfokainak száma $v = k - b - 1$, ahol k az osztályközök száma, b pedig a becsült paraméterek száma. A nullhipotézis állítása, hogy az alapsokaság eloszlása megegyezik a feltételezett elméleti eloszlással, az alternatív hipotézis pedig ennek ellenkezőjét állítja (Hunyadi – Vita, 2019).

A halandósági ráták előrejelzéséhez használt hasonlósági modell ezen illeszkedésvizsgálat elméleti alapjait használja majd fel, némi módosítással.

Az illeszkedésvizsgálatot az adott problémakörre interpretálva, a következő nullhipotézis fogalmazható meg: feltételezzük, hogy egy fejlődő ország halandósági rátáinak eloszlása (életkoronkénti bontásban) az $y-1$ – edik évben megegyezik egy másik, fejlett ország $y-1-p$ – edik évében a halálozási valószínűségek életkor szerinti eloszlásával. Vagyis esetünkben a tényleges eloszlások a fejlődő, míg a feltételezett eloszlások a fejlett ország halálozási rátáinak értékei lesznek.

A próbafüggvény értékének meghatározásához néhány új fogalmat és jelölést kell bevezetnünk. A próba elvégzéséhez szükséges paraméter a Θ , a vizsgált életkorok tartománya, E_x , mint a vizsgált időszakra jutó teljes adott x éves kori kitettség, h_x , az adott időszakban x éves korban elhunytak száma, valamint q_x , a halálozási valószínűség mértéke adott időszakra x éves korban (Bland – Altman, 2004). Ezek ismeretében immáron felírható a próbafüggvény meghatározásához használható képlet:

$$\chi^2 = \sum_{x \in \Theta} \frac{(h_x - q_x E_x)^2}{q_x(1 - q_x)E_x} \sim \chi^2(\#\Theta) \quad (4.3.2.)$$

Fontos megjegyezni, hogy a (4.3.2.) képletben szereplő paraméterek közül E_x és h_x értékek a fejlődő ország $y-1$ – edik évi, míg q_x – ek a fejlett ország adott $y-1-p$ – edik évi adataiból származnak.

A (4.3.2.) formula számlálójában szereplő képlet elméleti háttére ugyanaz, mint a (4.3.1.) képlet esetében, a tényleges (h_x) és a feltételezett eloszlások ($q_x E_x$) közötti eltérésnégyzetösszegeket számol, a nevező és a szabadságfokok száma azonban változott (a szabadságfokok értéke a vizsgált életkorok tartományának száma).

A (4.3.2.) képletbe a megfelelő adatokat behelyettesítve, megkapjuk az illeszkedésvizsgálat próbafüggvényének értékét, az úgynevezett χ^2 -et, mely érték alapján azonban még nem tudunk döntést hozni arról, hogy a tényleges és feltételezett eloszlások megegyeznek-e, teljesül-e a nullhipotézis. Ennek ellenőrzésére kell bevezetnünk a kritikus érték fogalmát.

Kritikus értéknek nevezzük az elfogadási és kritikus tartományt egymástól elhatároló alsó és felső értékeket, mely értékeket a kritikus tartomány részének tekintjük. A kritikus értékek egyértelműen meghatározhatóak a szignifikanciaszint és a próbafüggvény eloszlásának ismeretében. A szignifikanciaszint (jelölése α) annak a bekövetkezési valószínűsége, hogy a nullhipotézis teljesülése esetén a próbafüggvény a kritikus tartományba esik. Az α valószínűsége jellemzően 0-hoz közeli érték, százalékos formában megadva (Hunyadi – Vita, 2019).

Tekintve, hogy jobboldali χ^2 – próbát végzünk, csupán egy felső kritikus értéket kell meghatároznunk. Az érték kiszámításához ismert a próbafüggvény eloszlása (jobboldali χ^2 – eloszlást követ), szignifikanciaszintnek pedig érdemes a gyakorlatban is legtöbbször használt $\alpha = 5\%$ -ot választani. Jobboldali χ^2 – próba esetén felső kritikus érték számításánál meg kell adnunk még a szabadságfokok számát, melynek értéke a (4.3.2.) képlet alapján Θ , azaz a vizsgált életkorok tartományának száma.

Ezen paraméterek ismeretében már kiszámítható a felső kritikus érték, melyen keresztül már döntést tudunk hozni a nullhipotézisről. Amennyiben a próbafüggvény értéke meghaladja ezt a felső kritikus értéket (nagyobb/egyenlő nála), úgy elvetjük a nullhipotézist, és az alternatív hipotézist fogadjuk el. Ellenkező esetben, ha a próbafüggvény értéke kisebb, mint a felső kritikus

érték, akkor elfogadjuk azon feltételezésünket, miszerint a tényleges és a feltételezett eloszlások megegyeznek.

Mindezt az adott problémakörben vizsgálva, ha a próbafüggvény értéke kisebb, mint a felső kritikus érték, akkor elfogadhatjuk a nullhipotézist, azaz egy fejlődő állam $y-1$ – edik évének halandósági rátáinak eloszlása megegyezik egy fejlett állam $y-1-p$ – edik évében a halálozási valószínűségek eloszlásával.

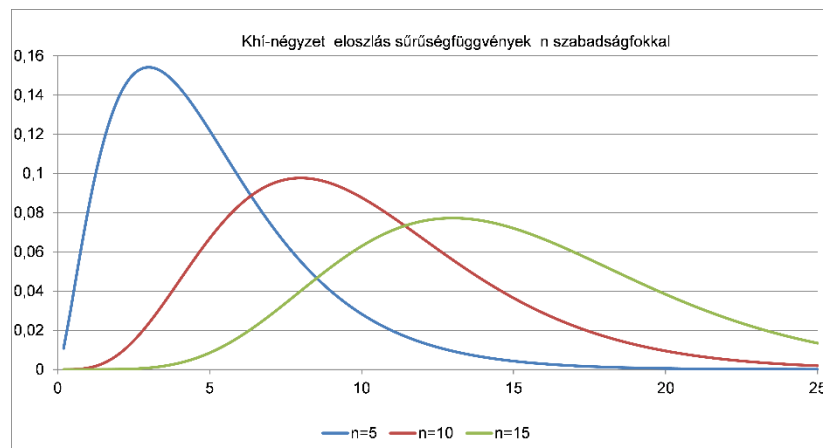
Hipotézisről azonban az úgynevezett p -érték alapján is dönthetünk, amit a mi eljárásunk is alkalmaz. A p -érték az a legkisebb szignifikanciaszint, melyen a nullhipotézist már éppen elvetjük az alternatív hipotézissel szemben. Értéke a próbafüggvény eredménye alapján határozható meg. Amennyiben p -értékre nagyobb számot kapunk, mint a beállított $\alpha = 5\%$ -os szignifikanciaszint, úgy elfogadhatjuk a nullhipotézist. Ennek oka, hogy a p -érték értelmezhető úgy is, hogy megmutatja, mekkora elsőfajú hibát (nullhipotézis elvetése, miközben az igaz) követnénk el azáltal, ha a próbafüggvény értéke alapján a nullhipotézist elvetnénk. Vagyis, ha nullhipotézis elvetéssel nagyobb hibát követnénk el, mint a megengedett $\alpha = 5\%$ -os szignifikanciaszint, akkor inkább a nullhipotézis elfogadása mellett döntünk (Hunyadi – Vita, 2019).

Mivel a hasonlósági modell során jobboldali χ^2 – próbát végzünk, ebből kifolyólag jobboldali p -értéket is kell számolnunk. Ennek meghatározásához szükséges a próbafüggvény értéke, valamint a szabadságfokok száma, majd az így kapott eloszlásfüggvény szerinti eredményt 1-ből kivonva kapjuk meg a jobboldali p -értéket.

Az illeszkedésvizsgálat azonban ezen a ponton még nem ért véget, tekintve, hogy halandósági ráták hasonlóságát vizsgáljuk. Érdeemes lenne alaposabban megvizsgálni ezen számítások eredményét, különösen a p -érték tekintetében.

Fennáll annak a lehetősége, hogy p -értékre egy viszonylag magas, 50-100%-os értéket kapunk. Azon országok esetében, ahol a p -érték ilyen magas számot eredményez, ott a χ^2 – próbafüggvény értéke, nagyságrendileg nézve, egy 0-hoz közeli szám. Ez a tény pedig a túlillesztés veszélyét hordozza magával. Ugyanis, a khi-négyzet eloszlás a próbafüggvény eloszlása igaz nullhipotézis esetén. Ebből adódik, hogy ha a χ^2 – eloszlás szerint a próbafüggvény értéke nagyon közel van a 0-hoz, az az esemény a χ^2 – eloszlás sűrűségfüggvénye alapján csupán csak nagyon kis valószínűséggel fog bekövetkezni még a valóságban igaznak vélt nullhipotézis esetén is.

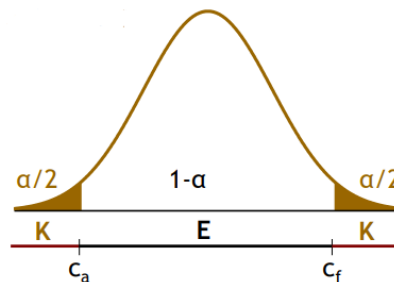
Ahogy az a 3. ábrán is látható, minél magasabb a szabadságfokok száma, annál kisebb a 0-hoz közeli értékek bekövetkezésének a valószínűsége. Ebből adódóan, ha egy ország p-értéke csak azért magas, mert a hozzá tartozó χ^2 – értéke nagyon alacsony, akkor hiba lenne azt feltételezni, hogy az adott ország mortalitási rátáinak eloszlása azonosnak tekinthető a fejlődő országgal, mivel a 0 közeli próbafüggvény értékének bekövetkezése még igaz nullhipotézis esetén is igen alacsony.



3. ábra: χ^2 -eloszlás sűrűségfüggvénye
Forrás: Csallner és Vincze, n.d.

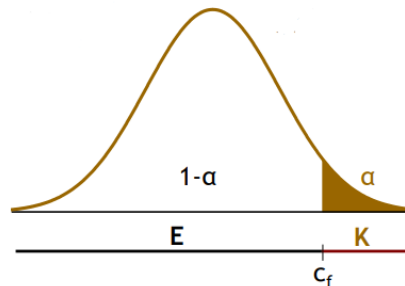
Ahol a p-értéke ilyen nagy számot vesz fel, érdemes ellenőrizni, hogy nemcsak a túlillesztés miatt kaptunk-e ilyen magas eredményt. Ennek kontrolljára a legkézenfekvőbb eszköz, ha ezen országok esetében nem jobboldali, hanem kétoldali p-értéket számolunk, ezáltal elkerülhető a túlillesztés veszélye.

Jobb-, illetve kétoldali p-érték számításának módja különböző, a különbözőség okának hátterében pedig az alsó, illetve felső kritikus értékek állnak. Különböző p-értékek számításának szemléltetéséhez a standard normális eloszlás sűrűségfüggvényét használom. Tekintsük tehát a 4. ábrán adott standard normális eloszlás sűrűségfüggvényét:



4. ábra: Kétoldali próba
Forrás: Balázs Móczai, 2020

Kétoldali p-érték számítása során a kritikus tartományba esés teljes $\alpha = 5\%$ -os valószínűségét egyenlő arányban kell megosztani a kritikus tartomány két széle között oly módon, ahogyan azt a 4. ábra is szemlélteti. Látható, hogy ebben az esetben egy alsó kritikus tartomány is létezik, szemben az 5. ábrával, amely mindössze egy felső kritikus tartományt tartalmaz.



5. ábra: Jobboldali próba
Forrás: Balázs MÓCSAI, 2020

Pontosan ezen okból kifolyólag kerülnek elfogadásra azok az esetek, amikor a próbafüggvény értéke nagyon alacsony, hiszen nincsen alsó korlát. Kétoldali p-érték számítása azonban megköveteli az alsó határ feltételét. Ebből kifolyólag azokhoz a számított χ^2 – értékekhez, melyek nagyságrendileg már az alsó kritikus tartomány alatt helyezkednek el, alacsony kétoldali p-értéket fog megállapítani, ezáltal ösztönözve arra, hogy a nullhipotézis elvetésével kisebb hibát véthetünk.

Ennek függvényében a kétoldali p-érték meghatározása is módosul. A próbafüggvény értékéből számolt p-érték kétszeresét kell venni, mégpedig az alábbiak figyelembevételével: ha a jobboldali p-érték meghaladja a teljes tartomány mediánját (50%-át), a kétoldali p-értéket úgy határozzuk meg, hogy a χ^2 – értékében egy baloldali p-értéket számolunk, majd annak vesszük a kétszeresét. Amennyiben pedig a jobboldali p-érték kisebb, mint 50%, ugyanakkor nagyobb, mint a beállított $\alpha = 5\%$ -os szignifikanciaszint, az azt jelenti, hogy a próbafüggvény értéke közelít a felső kritikus értékhez. Tehát ebben az esetben a kétoldali p-értéket úgy határozzuk meg, hogy az eloszlásfüggvény szerinti eredményt kivonjuk 1-ből, majd felszorozzuk kettővel.

Fontos megjegyezni, hogy az adott esetben nem számolhatunk azonnal kétoldali p-értéket minden országra. Ennek oka, hogy a jobboldali p-érték segítségével azonosítani tudjuk azokat az országokat, melyekre, ezen értékük alapján, nem fogadhatjuk el a nullhipotézist. Azonnali kétoldali p-érték számítás azonban nem szűrné ki ezeket az országokat, mivel, ahogyan az a 4. és 5. ábra összevetéséből látható, a 4. ábra felső kritikus értéke nagyobb, magasabb felső kritikus értéket enged meg, mint az 5. ábra felső kritikus értéke. Így a kétoldali p-érték alapján olyan országok is

elfogadásra kerülnének, melyek próbafüggvény értéke magasabb, mint a jobboldali p -értékhez tartozó felső kritikus érték. Ezen hiba elkerülése végett kell először jobboldali p -értéket meghatároznunk, ezáltal kizárva a lehetőségét annak, hogy a kétoldali p -érték számítás által olyan országok is bekerüljenek a modellbe, melyek eloszlása alapesetben nem tekinthető azonosnak a fejlődő országéval.

Azokban az országokban, ahol indokolt a magas p -érték, nem fog problémát jelenteni az átszámított kétoldali p -érték, az is a hasonlóságot fogja alátámasztani. Amennyiben ez a magas érték csak a túlillesztés műve, azt a kétoldali próba már azonosítani fogja, és alacsony p -értéket fog megállapítani a tényleges és feltételezett eloszlások tekintetében, mely azt jelenti számunkra, hogy érdemesebb elvetni a nullhipotézist, mivel így a hipotézisvizsgálat értelmében kisebb hibát vétünk.

Ezen eljárás segítségével már kiszűrhető, hogy valóban azonosnak tekinthető 2 ország mortalitási rátáinak eloszlása. A hasonlósági modellben, ha a jobboldali p -érték 5-50% között mozog, akkor elfogadható a nullhipotézis, miszerint a tényleges és feltételezett eloszlások azonosnak tekinthetőek. Amennyiben ez az érték 50% feletti értéket vesz fel, akkor folytatjuk az illeszkedésvizsgálatot, kezdve a kétoldali p -érték meghatározásával. Ha ez a szám kisebb/egyenlő, mint az 5%-os α szignifikanciaszint, akkor az alternatív hipotézist fogadjuk el, miszerint a 2 állam mortalitási rátái nem azonos eloszlásúak. Ez az az eset, amikor a magas (jobboldali) p -értéket túlillesztés eredményezi a modellben, ám ezt a hatást a kétoldali p -érték segítségével már azonosítani lehet.

Az ismertett módszertan segítségével immáron megállapítható, hogy egy fejlődő ország és egy adott fejlett ország mortalitási rátáinak eloszlása tekinthető-e azonosnak. A következő felmerülő kérdés, hogy az $y-1-p$ – edik évben létezik több ilyen fejlett ország is, és ha igen, azok közül melyiket érdemes kiindulási alapként megválasztani, azaz melyik ország múltjának alakulását használjuk előrejelzésként.

A megfelelő keretrendszer megválasztásával (melynek részletes ismertetésére az 5. fejezetben térek ki), elkészíthető egy olyan adattábla, melybe összegyűjthető valamennyi fejlett ország $y-1-p$ – edik évéhez tartozó, illeszkedésvizsgálat segítségével meghatározott jobboldali p -érték.

Elhagyva azokat az államokat, melyek jobb-, vagy átszámított kétoldali p -értékük alapján nem tekinthetőek azonos eloszlásúnak a fejlődő ország $y-1$ – edik évi adataival, előállítható egy olyan

tábla, mely az illeszkedő országok és a hozzájuk tartozó jobboldali p -értékeket tartalmazza. A korábban ismertetett túlillesztés veszélyét elkerülendő megközelítés értelmében érdemes minden országhoz meghatározni a kétoldali p -értéket.

Feltételezhető, hogy minél magasabb (az immáron átszámított kétoldali) p -érték, annál nagyobb ez a hasonlóság. Ebből adódóan azt a fejlett államot érdemes a hasonlósági modellbe beválasztani, ahol a legnagyobb a kétoldali p -érték. Ekkor remélhetjük ugyanis azt, hogy az előrejelzés a lehető legpontosabb lesz.

Ezzel gyakorlatilag elkészült a hasonlósági modell, mely alapján már előrejelezhetjük a fejlődő ország jövőjét, halálozási rátáit. Mivel az összehasonlításra az $y-1-p$ – edik évben került sor, az előrejelzés a fejlett ország múltjának $y-p$ – edik évétől kezdődően indul, ez lesz az első előrejelzett időszak, mely a fejlődő ország y – adik évének előrejelzett értéke. Ezt követően $y-p+1$ – edik év lesz a második előrejelzett időszak, mely a fejlődő ország $y+1$ – edik évi előrejelzése lesz, és így tovább, egészen p időszakon keresztül. Fontos megjegyezni, hogy az utolsó, p – edik előrejelzést a fejlett ország $y-1$ – edik évi adataiból készítjük és a fejlődő ország $y+p-1$ – edik évére vonatkozik.

Ezen eljárás segítségével jelezhető előre egy adott fejlődő ország mortalitási rátáinak jövőbeli alakulása, homogén viselkedésű, fejlett ország(ok) múltbeli trendje(i) alapján. Amellett, hogy kézenfekvő és kevés matematikai előfeltételre építő módszertant alkalmazunk, amelyben nem kell bonyolult változókkal, komolyabb matematikai ismereteket igénylő komplexebb számításokat elvégezni, látható, hogy az előrejelzés logikája jól értelmezhető, interpretálható eredményeket szolgáltat.

A hasonlósági modell és az előrejelzésének módszertana elméleti síkon már ismertetett. Annak bemutatására, hogy az elméleti modell hogyan működik a gyakorlatban, szükség van egy jól megválasztott keretrendszerre, mely magába foglalja többek között a megfelelő adathalmazokat, a számítások elvégzéséhez alkalmas programnyelvet. Az általam választott keretrendszer és a hasonlósági modell gyakorlati alkalmazásának bemutatására, leírására az 5. fejezetben térek ki.

5. ALKALMAZOTT KERETRENDSZER BEMUTATÁSA

Az 5. fejezetben mutatom be, hogy a 4. fejezetben ismertetett hasonlósági modell hogyan építhető fel, valósítható meg a gyakorlatban. Továbbá, ismertetem a felhasznált adatbázis struktúráját, az alkalmazott keretrendszert és annak működését.

5.1. Adatok ismertetése

A modell gyakorlati felépítésének és megfelelő illusztrálásának elengedhetetlen előfeltétele, hogy egy gondosan megválasztott adatbázisból dolgozzunk. Az elemzésemhez felhasznált adatok forrása a Human Mortality Database (HMD), mely a mortality.org címen keresztül érhető el. Az adatbázis jelenleg 41 ország (vagy terület¹) részletes népességi, születési és halálozási adatait tartalmazza. Kiemelkedően fontos megjegyezni, hogy az adatállomány mindezen adatokat külön-külön tárolja férfiakra és nőkre. A nem szerinti szeparációt a férfiak és nők életkor szerinti halandósági mintázatának jelentős eltérése indokolja. Az adatbázis többek között magába foglalja a január 1-jei populációt, azaz a kezdeti kitettséget. Ebből az értékből, az évközi halálozásokkal korrigálva, állítható elő a halálozási kockázatnak kitett népesség, az úgynevezett központi kitettség. Ezek mellett megtalálhatóak az adott évi, x évesen bekövetkezett halálesetek számai (D_x), valamint a centrális mortalitási ráták ($m_{x,t}$). Ezek értékeit az adott évi halálesetek és a központi kitettség hányadosaként számítják ki, majd ebből kalkulálják vissza a kezdeti halálozási $q_{x,t}$ valószínűséget a

$$q_{x,t} = \frac{m_{x,t}}{1 + \frac{1}{2}m_{x,t}} \quad (5.1.1.)$$

képlet segítségével (Human Mortality Database, n.d.).

5.2. Az R szoftverkörnyezete

Látható, hogy a megfelelő adatbázisok, adattáblák rendelkezésre állnak, most már csak egy jól választott alkalmazásra, programra van szükség, mely képes ezen adatok beolvasására, kezelésére, feldolgozására. A témakörből adódóan kézenfekvő megoldásnak tűnik az R szoftvercsomag

¹ A bemutatott programban 50 ország szerepel, ennek az eltérésnek az oka, hogy egyes országok (pl. Egyesült Királyság, Németország, stb.) több tagállamból állnak (Wales, Észak-Írország, Skócia) vagy éppen több egységre bontva az országot (Kelet-, Nyugat-Németország) tárolják az adatokat.

használata, mely egy nyílt forráskódú, folyamatos fejlesztés alatt álló szoftverkörnyezet, elsősorban statisztikai számításokhoz és grafikus megjelenítésekhez. A szoftver különböző csomagokkal bővíthető, melyek száma napról-napra növekszik. Ezeket a csomagok az úgynevezett CRAN (The Comprehensive R Archive Network) hálózat tömöríti magába. A CRAN ftp (file transfer protocol) és webszerverek hálózata a világ körül, mely az R kódjának és dokumentációjának naprakész verzióit tárolja (What is R?, n.d.).

A hatékony adatkezelési és tárolási lehetőség, valamint az adatelemzési eszközök széles skálája ideális választássá teszi az R szoftver használatát az adott kérdéskörben. Ebből kifolyólag kutatási munkám során az R program, valamint néhány kiegészítő csomag segítségével dolgoztam. Ezek közül külön kiemelném az úgynevezett *MortalityLaws* csomagot, melynek segítségével közvetlen API-n keresztül beolvashatóak a HMD halandósági táblái.

5.3. Magyarországi adatok beolvasása

Ismerve az elemzéshez felhasználandó adatok forrását (HMD) és a választott szoftvert (R), elkezdődhet az adatok feldolgozása. Első lépésként be kell olvasni a programba a halandósági táblákat a HMD adatbázisából. Érdemes lenne először a magyarországi adattáblákat letölteni, ügyelve arra, hogy a férfi-női adatokat külön-külön tároljuk. Ezen bontáson belül 3 különböző adattáblára lesz szükségünk: egy központi kitettségi táblára (E_x), egy halálozási számok táblára (D_x) és végül, de nem utolsósorban egy halálozási valószínűségeket tartalmazó táblára (q_x).

A halandósági táblák R-ben egy úgynevezett *ReadHMD* beépített függvény segítségével tölthetők le a HMD oldaláról. A függvénynek bemeneti paraméterként meg kell adni az adatok típusát, a vizsgált ország kódját, valamint a kor- és időintervallum szerinti bontás mértékét². Ezek közül érdemes kiemelni az adatok típusát. A HMD oldaláról különböző adattáblák tölthetők le, a megfelelő kulcsszó segítségével. Az E_x központi kitettség és a D_x halálozási számok tábla jelölésüknek megfelelő megnevezéssel tölthetők le. Itt arra kell figyelni, hogy ezen táblák összevontan tartalmazzák a férfi-női értékeket, azokat külön-külön kell elmenteni 1-1 változóba. A q_x halálozási valószínűségeket tartalmazó tábla betöltése másképp valósítható meg. Ezek az értékek magukban az élettartam táblázatokban (life table) találhatóak, mely táblázatok külön-külön

² Utolsó két paraméterként meg kell adni egy, a HMD adatbázis oldalára előzetesen regisztrált emailcímet és jelszót.

lekérdezhetőek férfi és női adatok vonatkozásában, a szó angol megfelelőjének kezdőbetűjével (male – férfi, female – nő).

Az így beolvasott adattáblákon némi módosítást, konvertálást elvégezve, a táblák struktúrája a következőképpen fog kinézni: minden évre (oszlop), életkoronkénti bontásban (sor) ismertek az adott, lekérdezett értékek (q_x , E_x , D_x)³. Így a kiinduló 3-3 adattábla már adott mind férfiak, mind nők esetében, magyar adatok tekintetében.

5.4. Egységes adattábla kialakítása hasonlósági modell alkalmazásához

Tekintsük adottnak y évet, amelytől kezdődően majd szeretnénk előrejelezni, valamint az előrejelezni kívánt p időszakok számát. Ekkor a feladatunk az adott $y-1-p$ – edik évi életkoronkénti halálozási valószínűségek összegyűjtése valamennyi HMD ország esetében. Ennek megvalósításához első lépésként szükséges egy olyan függvény megírása, mely a HMD adatbázisában fellelhető összes (külföldi) ország/tagállam kódját kimentí egy listába. Ennek segítségével létrehozható egy üres lista, melynek hossza a külföldi tagállamok kódjainak száma. Az így létrehozott listába fogjuk elmenteni minden ország $y-1-p$ – edik évi q_x mortalitási rátáit.

A megvalósítás egy *for* ciklusba ágyazott, úgynevezett *AbroadM* és *AbroadFe* függvények segítségével történik. A ciklus pontosan annyiszor fog lefutni, ahány külföldi országot kigyűjtöttünk. A cikluson belüli függvénynek mindössze két bemeneti paraméterre van szüksége, az ország kódjára, valamint az adott $y-1-p$ – edik évre. Az ország kódja mindig a soron következő cikluselem, azaz a kigyűjtött országcódok listájának i . tagja lesz. Letöltve az adott i – edik országot, leszűrjük a táblát úgy, hogy csak az adott $y-1-p$ – edik évi q_x mortalitási rátákat tartalmazza. Az így kapott táblát kimentjük a már korábban létrehozott üres listának az i – edik helyére, egy gondosan beépített *tryCatch* függvény kíséretében. Ennek alkalmazását elsősorban az indokolja, elképzelhető, hogy nem minden kigyűjtött országhoz létezik adat az adott $y-1-p$ – edik évben. Ennélfogva a függvény hibaüzenetet jelezne a szűrést követően. Így ezen listaelemekhez egyszerűen NULL értéket rendelünk, mely értékeket a továbbiakban elhagyjuk a listából.

Ezzel a ciklusmag végére érünk, és a folyamat kezdődik előlről, egészen addig, amíg a függvény az összes külföldi ország adatait össze nem gyűjti. A folyamat végén, az egyszerűbb és átláthatóbb

³ Az elemzésem során a magyar adatok 1950 és 2017 közötti értékeit használok fel, 0-tól 100 éves korig, mind férfiak, mind nők esetében.

értelmezés érdekében, némi átalakítást végzünk el az adattáblán: a *tryCatch* függvény által detektált, NULL értékű listaelemeket elhagyjuk, az egyes oszlopneveket pedig átnevezzük az adott oszlophoz tartozó ország kódjára. Ezt követően már rendelkezésünkre áll az a kiinduló tábla, mely adott évre, életkoronkénti bontásban tartalmazza valamennyi országra vonatkozóan azok halálozási valószínűségeit.

Fontos megjegyezni, hogy az itt leírt függvény során is külön figyelmet kell fordítani a nemekre. A konzisztens működés érdekében külön függvényt írtam mind a férfi, mind a női adatok letöltésére (*AbroadM* és *AbroadFe*).

5.4.1. Hasonlósági modell implementációja

A következő feladat annak meghatározása, hogy melyik ország $y-1-p$ – edik évi halálozási valószínűségei hasonlítanak leginkább a magyarországi $y-1$ – edik évi értékekhez.

Ennek megoldásához mindenekelőtt meg kell határoznunk az adott $y-1$ – edik évi magyar központi kitettségek és halálozási számokat. Ehhez a korábban beolvasott adattáblákat alkalmazzuk, melyekben, hasonlóan a *for* ciklusban alkalmazott módszerhez, leszűrjük az adatokat az adott évre. Az ehhez létrehozott *KozosTabla* (valamint *KozosTablaFe*) függvény szintén két bemeneti paramétert vár: egy évszámot, amely alapján majd szűrünk, valamint egy úgynevezett alsó korhatárt. Ennek bevezetésére azért van szükség, mivel alapvetően az idősebb korosztály értékeinek alakulására fókuszál a kutatás, így ezen paraméter értéke valamennyi esetben 60 lesz, azaz a 60 éves korosztálytól kezdődően vizsgáljuk az adatokat. A függvény ezen két szűrést követően összefűzi egy táblává a magyar központi kitettségek és halálozási számok $y-1$ – edik évi értékeit a már korábban előállított összes többi ország $y-1-p$ – edik évi mortalitási rátáival.

Ezen tábla segítségével már elvégezhető a 4.3. fejezetben részletesen ismertetett illeszkedésvizsgálat, azaz, hogy tekinthető-e azonosnak a magyar és egy adott külföldi ország mortalitási rátáinak eloszlása. Ennek megállapítására írt *MenOfYear* függvény működésének bemutatása következik.

A függvény az összefűzött táblával dolgozik, egyetlen bemeneti paramétert vár, nevezetesen azt a táblát, melybe valamennyi ország releváns évre vonatkozó halálozási rátáit gyűjtöttük össze. A függvény először létrehoz 5 listát, melyek hossza a bemeneti paraméterként adott tábla oszlopainak száma. Azaz az országok száma, mely országokhoz az adott évben meg tudtuk határozni a

mortalitási rátákat. Ezeket a listákat a függvény belsejében megírt *for* ciklus segítségével fogjuk feltölteni.

A ciklus 3-tól indul és a bemeneti paraméterként megadott tábla hosszának kettővel növelt értékéig tart. Az adott ciklusban az *i*. elem az összevont tábla *i*. oszlopát fogja jelölni. A 3-as kezdést az indokolja, hogy az összefűzött táblában az első két oszlopban magyar adatok találhatóak. A külföldi országok adatai, melyekre ténylegesen végezzük a vizsgálatot, csak a 3. oszloptól kezdődnek. Ebből pedig az is látszik, hogy ha nem növelnénk kettővel a *for* ciklus iterációinak a számát, úgy a 3-as kezdés miatt minden esetben pontosan két adatsort, pontosabban adatoszlopot veszítenénk.

A ciklus magjában, a megfelelő matematikai és statisztikai eszközöket alkalmazva, megkezdődik az egyes listaelemek kiszámításának és feltöltésének a folyamata. A ciklus rendre a következő értékeket határozza meg az egyes országokra vonatkozóan: χ^2 – érték, jobboldali *p* – érték, kritikus felső érték, kétoldali *p* – érték, illeszkedési hipotézisről szóló döntés (*illeszkedik* vagy *nem illeszkedik*).

A ciklus magja, a 4.3. fejezetben ismertetett elméleti háttér miatt, néhány *if* – *else* ággal is kiegészül. Kétoldali *p* – érték meghatározására csak abban az esetben kerül sor, ha a jobboldali *p* – érték nagyobb, mint 50%. Annak megállapítására, hogy a nullhipotézis teljesül-e, a következő feltételrendszert kell vizsgálnunk: ha számított kétoldali *p* – érték nagyobb, mint a beállított $\alpha = 5\%$ -os szignifikanciaszint, akkor igaznak tekinthető a nullhipotézis. Továbbá abban az esetben is teljesül a nullhipotézis, ha a jobboldali *p* – érték nagyobb, mint a beállított $\alpha = 5\%$ -os szignifikanciaszint és nem nagyobb 50%-nál. Ilyen esetekben az adott országhoz tartozó hipotézis *illeszkedik* feliratot fog eredményezni, minden egyéb kimenet esetén *nem illeszkedik* szöveg fog megjelenni.

Az így feltöltött listákat, néhány módosítást követően, a függvény egy közös táblává fűzi össze, melyből már egyértelműen leolvasható, mely országok esetében állapítható meg, hogy mortalitási rátaiknak eloszlása azonosnak tekinthető a magyarországi értékek eloszlásával.

5.4.2. A legjobb ország

A legjobban illeszkedő ország meghatározása az *illeszkedes* függvény segítségével történik. Ismerve az országok *y-1-p* – edik évi hipotézisének kimenetét, leszűrhető az illeszkedő országok

listája, valamint a hozzájuk kiszámított jobb-, illetve kétoldali p – értékek. A konzisztens megoldás végett ebben a táblában érdemes ahhoz az országhoz is meghatározni a kétoldali p – értéket, ahol eddig csak jobboldali p – értéket számítottunk, mivel nem állt fenn a túlillesztés veszélye. Ennek végrehajtásához egy *for* ciklussal végigjárva a táblát, ahol még nincs kétoldali p – érték számítva, ott ezt pótoljuk. Az így módosított tábla tehát tartalmazza az illeszkedő országok nevét, valamint a hozzájuk tartozó jobb- és kétoldali p – értékeket.

A korábban ismertetett hasonlósági modell elmélete alapján ebből a listából azt az országot kellene a hasonlósági modell alapjául választani, ahol a legnagyobb a kétoldali p – érték, ezáltal biztosítva azt, hogy az előrejelzés a lehető legpontosabb legyen. A fentiek ismeretében, a megfelelő R függvények segítségével, kikereshető a legmagasabb kétoldali p – értékkel rendelkező ország, mely ország kódját egy változóban eltároljuk.

A kikeresett ország ismeretében már előrejelezhetőek a halálozási valószínűségek. Az előrejelzés megvalósításához írt *elorejelzes* függvénynek összesen 5 bemeneti paraméterre van szüksége. Egyrészt arra az adott országra, melyből az előrejelzést szeretnénk elkészíteni. Ez az az ország lesz, amit az *illeszkedes* függvénnyel kerestünk ki a maximális kétoldali p – érték alapján. Itt is meg kell adni az alsó korhatárt (60 év), az első előrejelzett évet, ami alapesetben az $y-p$ – edik év, valamint az előrejelezni kívánt p periódusok számának eggyel csökkentett értékét⁴. Továbbá kötelezően meg kell adni a beolvasni kívánt adattábla típusát, annak megfelelően, hogy milyen nemre szeretnénk az előrejelzést elkészíteni („LT_m” [férfi] vagy „LT_f” [nő] jelöléssel).

Így már rendelkezésünkre áll az adott ország múltja alapján az $y-p$ – edik évtől kezdődő, $y-1$ – edik évig tartó előrejelzés, melyet a hasonlósági modell értelmében Magyarország y – adik és $y+p-1$ – edik közötti évek előrejelzett értékeinek feleltethetünk meg.

5.4.3. Az előrejelzés kiértékelése

Az előrejelzés ezáltal megvalósítható, ugyanakkor érdemes lenne az előrejelzett és a tényleges értékeket összehasonlítani. Ennek végrehajtásához hoztam létre egy úgynevezett *kiertekeles* függvényt, mely szintén 5 bemeneti paramétert vár: az alsó korhatárt, az y – adik évet, az előrejelezni kívánt p periódusok számának eggyel csökkentett értékét, a hasonlósági modell alapján

⁴ Matematikai okokból kifolyólag, valamint az észszerűbb függvényírás megvalósításának érdekében.

készített előrejelzett mortalitási rátákat, valamint a tényleges magyar halálozási valószínűségeket tartalmazó táblát férfiak vagy nők vonatkozásában a vizsgált előrejelzési időszakra.

A függvény a magyar halálozási valószínűségek tábláját leszűri a bemeneti paraméterek függvényében, y – adik évtől kezdődően p periódusra, a 60+ fölötti korosztályra. Ezt követően összefűzi egy közös táblává a tényleges és előrejelzett értékeket. Ezután egy *for* ciklus segítségével végigmegy valamennyi adaton, mely ciklus a (5.7.1.) – ben bemutatott konvertálást végzi el az értékeken. Mind a tényleges, mind az előrejelzett q_x halálozási valószínűségeket a

$$m_{x,t} = \frac{q_{x,t}}{1 - \frac{1}{2}q_{x,t}} \quad (5.7.1.)$$

áttérő képlet segítségével átszámítja $m_{x,t}$ centrális mortalitási rátákká, majd az így kapott értékek logaritmusát veszi. A logaritmizálást elsősorban az indokolja, hogy az 1-1 életkorban mért esetleges kilógóan nagy vagy kicsi hiba ne befolyásolja majd túl érzékenyen az összesített hibát.

A konvertálást követően már vizsgálható a tényleges és az előrejelzett értékek közötti eltérések mértéke. Ennek megvalósításához az átlagos négyzetes hibák (Mean Squared Error, MSE) módszerét alkalmaztam, mely az egyik legelterjedtebb statisztikai eszköz előrejelzések kiértékelésekor (Hunyadi – Vita, 2019). Ahogy a neve is mutatja, alapvetően a becslt (előrejelzett) és a tényleges értékek átlagos négyzetes eltéréseit adja eredményül. Kiszámítása a

$$MSE = \frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n} \quad (5.7.2.)$$

képlet segítségével történik, ahol t jelzi az egyes időszakokat, $\hat{y}_t$ a t – edik időszak előrejelzett, y_t pedig a tényleges t – edik időszak értékeit.

Az (5.7.2.) képletet beépítve a függvénybe, könnyedén kiszámíthatóak az egyes MSE-k minden p évre. A függvény belsejében létrehozunk egy p hosszúságú listát, melybe majd az egyes évekhez tartozó MSE értékeket tároljuk. Ezt követően egy *for* ciklussal végigjárva az összevont adattáblát, az (5.7.2.) formula segítségével minden időszakra meghatározzuk az MSE értékét. Az így meghatározott MSE értékeket elmentjük a létrehozott listába. Ezzel ismertek egy-egy évre, hogy egy adott ország múltjából előrejelzett értékek átlagosan (négyzetesen) mennyire tértek el a tényleges magyar centrált, logaritmizált halálozási valószínűségektől. Figyelembe véve, hogy négyzetesen számított eltéréseket számolunk, a listában tárolt MSE értéket egyszerűen

átlagolhatjuk, így meghatározhatjuk a hasonlósági modell alapján készített előrejelzés egészére vonatkozó átlagos négyzetes hibát.

5.5. Lee-Carter modell előrejelzése R-ben

Egyetlen lépés maradt hátra, nevezetesen annak vizsgálata, hogy valóban jobb predikciót kapunk-e a hasonlósági modell alapján, mint amit a Lee-Carter módszerrel értünk volna el. A kérdés megválaszolására, ellenőrzésére egy Lee-Carter modell alapján történő előrejelzést is kell végeznünk az R-ben.

A Lee-Carter előrejelzést megvalósító *LCmodell* függvény alapvetően 4 bemeneti paramétert vár: elsősorban a tanító időszak végét jelző $y-1$ – edik évet, az előrejelezni kívánt p időszakok számát, valamint a magyarországi központi kitettségi és halálozások száma táblát, a vizsgálni kívánt nem függvényében megadva.

A függvény elején egy változóba elmentjük a tanító időszakok számát⁵. Ezt követően az R program Lee-Carter implementációi közül az úgynevezett *StMoMo* (Stochastic Mortality Modelling) csomag segítségével létrehozunk egy üres Lee-Carter modellt. Ennek és a megfelelő paraméterek felhasználásával, készíthető el a Lee-Carter modell egyes paramétereinek becslése. Az így meghatározott paraméterek alapján jelezzük majd előre a következő időszaki magyar halálozási valószínűségeket.

A paraméterbecsléshez a függvénynek meg kell adnunk a halálozási számokat, a központi kitettséget, a tanító időszakok számát, valamint az életkorok tartományát. Ezek felhasználásával a program a háttérben elkészíti a modell paramétereinek becslését a Lee-Carter elméletét bemutató, 3. fejezetben ismertetett módszer alapján. A becsült paraméterek felhasználásával és az előrejelezni kívánt p időszakok megadásával, a függvény elvégzi az előrejelzést. Eredményül egy olyan táblát kapunk vissza, mely életkoronkénti bontásban (60+ korosztályra szűrve), p időszakra tartalmazza az előrejelzett, magyar mortalitási rátákat.

Természetesen ebben az esetben is érdemes leellenőrizni a tényleges és az előrejelzett értékek közötti eltérést. A hibát ezúttal is a már korábban ismertetett MSE mutatóval fogjuk mérni. Ennek ellenőrzésére egy hasonló kiértékelő függvényt készítettem, mint amivel a hasonlósági modell

⁵ A HMD adatbázis adatainak ismeretében tudjuk, hogy a tanító időszak mindenképp 1950-től indul, ettől korábbi adattábla nincsen magyarországi adatok vonatkozásában.

során is dolgoztam (*LCkiertekelo*). Bemeneti paraméterként ugyanúgy a tényleges és előrejelzett halálozási valószínűségek tábláit kell megadni, mindegyikre a megfelelő szűréseket elvégezve, valamint az előrejelezni kívánt p időszakok számát. A halálozási valószínűségeket itt is az (5.7.1.) formula szerint kell átváltani logaritmizált $m_{x,t}$ centrális mortalitási rátákká. Az átszámított értékekre az (5.7.2.) képlet alapján meghatározzuk az MSE mutató értékeit minden egyes évre. Ezt követően pedig az MSE mutatókat egy előzetesen létrehozott listába kimentjük. A lista értékeire egy átlagot számítva megkapjuk a Lee-Carter modell által készített előrejelzések átlagos négyzetes hibáját a tényleges értékekhez viszonyítva.

5.6. Hasonlósági és Lee-Carter modellek összehasonlítása

Ezekután egyetlen egyszerű összehasonlítást kell tennünk: meg kell vizsgálnunk, hogy melyik modell esetében lett kisebb az átlagolt átlagos négyzetes eltérés. Amennyiben az adott előrejelzés során a hasonlósági modell átlagos MSE értéke kisebb, mint a Lee-Carter modell alapján számolt átlagos MSE, úgy elmondható, hogy adott esetben valóban jobban teljesített a hasonlósági modell, jobb predikciót tudott készíteni, mint a Lee-Carter modell.

A keretrendszerhez készítettem egy grafikus felhasználói felületet (graphical user interface, GUI) Shiny applikációként. Az alkalmazás célja az egyes előrejelzések eredményeinek vizuális összehasonlítása a két modell esetén. A felhasználónak a felületen 4 paramétert kell beállítania az előrejelzés megvalósításához: az évszámot, az országot, az előrejelezni kívánt periódusok számát és a nemet. Az évszám az előrejelezni kívánt magyar év, amelynek az értéke 1991-2017 közötti lehet. Az applikáció a hasonlósági modell érzékenységvizsgálatához biztosít egy grafikus eszközt a demográfiai kutatóknak, ezért a program nem keresi ki a legjobban illeszkedő országot. A felhasználó maga állíthatja be a paramétert aszerint, hogy melyik országból szeretne előrejelezni. A periódusok száma az a p paraméter, mely azt jelzi, hogy hány évvel korábbi halandóságot keresünk vissza a kiválasztott fejlett országból (lásd 4.2. fejezet). Továbbá, a felhasználó kiválaszthatja, hogy férfiakra vagy nőkre szeretné az előrejelzést elkészíteni. Valamennyi beállítást követően az *Előrejelzés* gombra kattintva, a program háttérében megkezdődik az előrejelzések futtatása és néhány másodpercnyi várakozás után megjelennek a vizualizált előrejelzések. Az alkalmazásról készített képernyőképeket az 1. és 2. számú mellékletekben jelenítettem meg.

Fontos megjegyezni, hogy a ShinyApp program részben az R kód által generált munkaállomány eredményeiből dolgozik. A keretrendszert megvalósító R kód, az elemzésem eredményének munkaállománya és a Shiny applikáció a

https://unicorvinus-my.sharepoint.com/:f:/g/personal/judit_fabok_stud_uni-corvinus_hu/EiXO_lJP4ixCvcKSNYEOWOABR-mGoIzRjejqdLR01-bcg?e=bq6mMR

linken érhető el.

Ismerve a hasonlósági modell elméleti hátterét, valamint gyakorlati megvalósítását, működését, a 6. fejezetben ismertetem a ténylegesen elvégzett előrejelzések eredményeit, mind a hasonlósági, mind a Lee-Carter modell esetén. Elemzésem során kitérek majd az eltérések hátterében húzódó okok magyarázatára is.

6. EREDMÉNYEK ÉS ÉRTÉKELÉSÜK

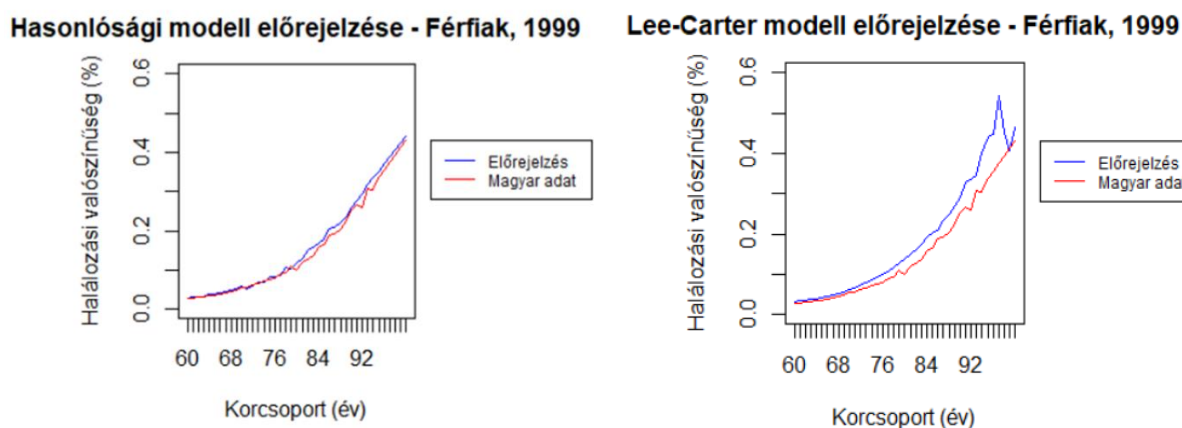
6.1. A vizsgált, előrejelezni kívánt 3 időszak

Kutatásom során alapvetően 3 különböző időszak előrejelzéseit vizsgáltam: az 1991-1999-es, az 1991-2017-es, valamint a 2000-2017-es évek mortalitási rátáinak alakulását. Tekintve, hogy a férfiak és nők halálozási rátái jelentősen eltérnek egymástól, mindhárom időszakra külön-külön előrejelzéseket készítettem mindkét nem esetében. Továbbá, figyelembe véve a tényt, hogy az élettartam-kockázat elsősorban a nyugdíjrendszer szempontjából releváns, kutatásom során én is a nyugdíjas vagy ahhoz közel álló korosztályra fókuszáltam. Tehát, az egyes predikciók a 60 év feletti korosztályokra vonatkoznak. Fontos megjegyezni, hogy a hasonlósági modellben a 100 év feletti korcsoportokat – kis gyakoriságuk miatt – összevontuk a 100 korévhez, és ebből kifolyólag a 100 éves korosztály halálozási valószínűségét 100%-nak vettük. Ezen összevonások alapján keresi meg a modell a legjobban illeszkedő országot, melyből majd előrejelzünk. A 100 éves korra előrejelzett halálozási valószínűségeket a Lee-Carter modellel való összehasonlíthatóság miatt azonban már nem 100%-ként jeleztük előre.

6.2. Az 1991-1999 közötti időszak előrejelzése

Elemzésem során először az 1991-1999 közötti időszak halálozási rátáinak alakulását jeleztem előre a saját hasonlósági modellem segítségével. Az előrejelzés megvalósításához első lépésként meg kellett határoznom, hogy 1981-ben melyik külföldi ország mortalitási rátái hasonlítottak leginkább az 1990-es évi magyar értékekhez. Férfiak esetében ez az ország Szlovákia lett. Az 1982-1990-es szlovák adatok alakulása megfeleltethető az 1991-1999-es évi magyar halandósági ráták alakulásának.

Az így előrejelzett értékeket összevetve a tényleges 1991-1999 közötti magyar adatokkal, a centrált logaritmizált halálozási valószínűségekre ($m_{x,t}$) számolt MSE értéke 0,0103. Ezzel szemben a Lee-Carter modell segítségével készített előrejelzések esetén ugyanezen mutató értéke 0,0393. Tehát, az MSE érték a hasonlósági modell alapján jobb, hiszen közel 3 századdal jobban teljesített. Továbbá érdemes a két módszer által előrejelzett halálozási valószínűségeket vizuálisan is összevetni a tényleges magyar adatokkal. A 6. ábra jeleníti meg, hogy az egyes modellek esetében hogyan alakultak a halálozási valószínűségek az első előrejelzés utolsó időszakában, 1999-ben.



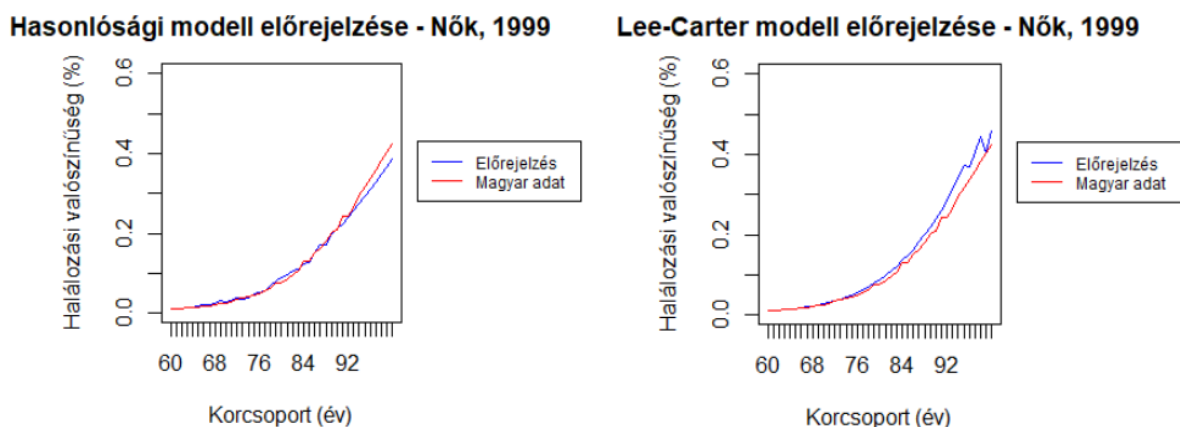
6. ábra: Előrejelzések alakulása férfiak esetében 1999-ben, hasonlósági és Lee-Carter modell alapján
 Forrás: saját szerkesztés

Látható, hogy a hasonlósági modell esetén az előrejelzések szinte végig, minden korosztály esetében jól közelítik a tényleges adatok alakulását, míg a Lee-Carter modell már a 70 éves korosztálytól kezdődően jelentősen felülbecsli a férfiak halálozási valószínűségeit. Ezen előrejelzés esetében elmondható tehát, hogy érdekesebb lenne hasonlósági modell alapján előrejelezni egy fejlődő ország (Magyarország) mortalitási rátáinak jövőjét egy fejlett állam (jelen esetben Szlovákia) múltbeli alakulásának függvényében. A 70-100 éves korosztály halandóságát a Lee-Carter modellel ellentétben ugyanis nem becsli túl. Ellentmondónak tűnhet, hogy a hasonlósági modell 1981-ben Szlovákiát azonosította, mint fejlett állam. Az ellentmondás egy lehetséges magyarázatára a nőkre vonatkozó előrejelzést követően fogok kitérni.

Hasonlóan elvégezve az elemzést a nőkre, a hasonlósági modell alapján 1981-ben Litvánia halálozási rátái hasonlítottak leginkább az 1990-es magyar adatokhoz. Litvánia múltjának adatait Magyarország jövőjének feltételezve, a hasonlósági modellre számolt MSE mutató értéke 0,0094, míg a Lee-Carter modellel készített előrejelzés alapján számított MSE érték 0,0194. Itt a hasonlósági modell már 1 századdal teljesített jobban, mint a Lee-Carter modell. Az előrejelzett értékeket a 7. ábrán vizuálisan is megjelenítve hasonló eredmény tapasztalható.

A Lee-Carter modell előrejelzett 1999-es értékei, bár a férfiakhoz képest kisebb mértékben, de továbbra is túlbecsülik a 85-90 éves és az afölötti korosztályok halálozási valószínűségeit, ugyanakkor elég precíz predikciókat képez a 60-80 éves korcsoporthoz. A hasonlósági modell nők esetében is elég pontos előrejelzéseket készít, egyedül a 90-100 éves korosztályban figyelhető meg egy minimális alulbecslés. Érdekemes megjegyezni, hogy az alulbecslés kisebb hiba, mint a

felülbecslés, amennyiben az eredményeket nyugdíjrendszer tervezéshez szeretnénk használni. Az alulbecslés ebben az esetben azt jelenti, hogy a valóságban többen haltak meg, mint amennyi halálozást a modell előrejelzett. Ekkor a valós, kifizetett nyugdíj kevesebb lesz, mint a tervezett. Ez pedig valóban kisebb hibát jelent a nyugdíjak kifizetőjének, mintha nem lenne elég a tervezett pénz, mert kevesebben haltak meg az előrejelzéshez képest. Azaz, bár összességében a Lee-Carter modell előrejelzései is eléggé közelítik a tényleges magyar adatokat, nők esetében is jobb előrejelzést képes adni a hasonlósági modell.



7. ábra: Előrejelzések alakulása nők esetében 1999-ben, hasonlósági és Lee-Carter modell alapján
Forrás: saját szerkesztés

Az 1991-1999-re készített előrejelzések összefoglalását az 1. táblázat tartalmazza:

	férfi	női
Ország	Szlovákia	Litvánia
MSE	0,0103	0,0094
Lee-Carter MSE	0,0393	0,0194

1. táblázat: Az 1991-1999 közötti mortalitási ráták összefoglaló táblázata
Forrás: saját szerkesztés

Érdemes kitérni kutatásom kiindulópontjául szolgáló hipotézisemre, miszerint a fejlődő államok mortalitási rátái egy fejlett ország értékeihez hasonlíthatnak majd. Ebben az esetben azonban más tapasztalható, a legjobban illeszkedő ország ugyanis Szlovákia és Litvánia lett. A két állam közös jellemzője, hogy az egykori Szovjetunióhoz köthetőek: Szlovákia, mint a volt szocialista blokk része és Litvánia, mint egykori szovjet tagállam. Ennek értelmében nem sorolhatóak maradéktalanul a fejlett államok közé 1981-ben. Az ellentmondás feloldásához kisebb történelmi kitekintést kell tennünk a két ország vonatkozásában az adott időszakra.

A Csehszlovák Szocialista Köztársaság számára a hetvenes évek az erősen centralizált politikai élet mellett gazdasági fellendülést hozott. A nyolcvanas évek végére pedig Csehszlovákia is nyitni kezdett a külvilág felé, valamint lépéseket tett a piacgazdaság és a demokrácia irányába (Horbulák, 2018). Látható tehát, hogy alapvetően Csehszlovákia jobb helyzetben volt, szemben a szovjet blokkot képező tagállamokkal. Ezzel indokolható az, hogy a hasonlósági modell Szlovákiát, mint fejlett államot, adta eredményül 1981-ben. Érdeemes megjegyezni, hogy Magyarországon is hasonló folyamatok mentek végbe a 80-as években, mint Szlovákiában. A hasonlósági modell alapján azonban úgy tűnik, hogy ez a nyugati nyitás Szlovákiában már a 80-as években tükröződött a férfi halandóságban, míg Magyarországon ez a javulás csak a 90-es években ment végbe.

Litvánia, mint balti állam, gazdaságilag az egyik legfejlettebb köztársaságnak számított a szovjet tömbön belül (Kenéz, 2008). Ennek következtében alapvetően jobb helyzetben volt, mint a Szovjetunió legtöbb tagállama. Ezzel magyarázható az, hogy a hasonlósági modell Litvániát fejlett államként azonosította 1981-ben.

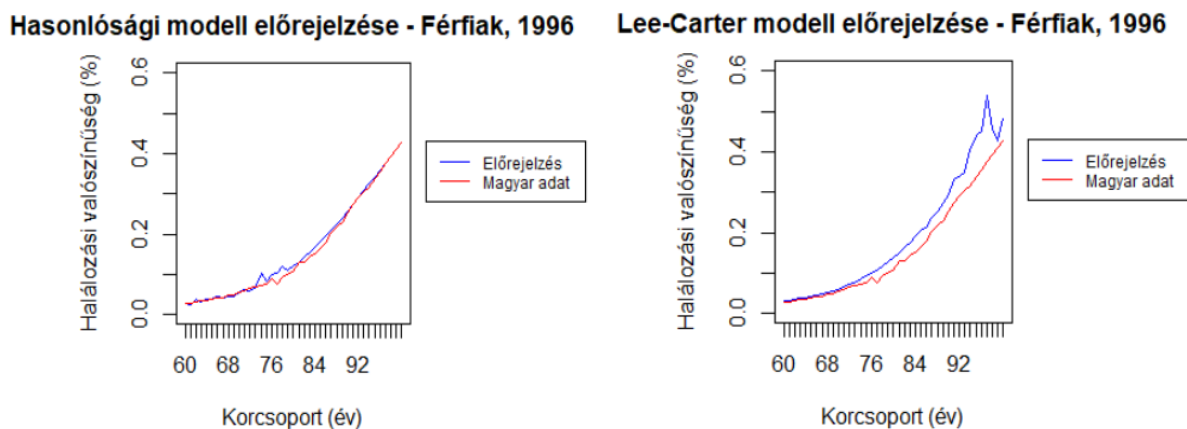
6.3. Az 1991-2017 közötti időszak előrejelzése

Kutatásom során a következő vizsgált és előrejelezni kívánt időszak az 1991-2017 közötti periódus volt. A hasonlósági modellem alapján meghatároztam, hogy férfiak vonatkozásában az 1990-es magyar adatokhoz 1963-ban leginkább Luxemburg halandósági rátái hasonlítottak, melynek értelmében az 1964-1990 közötti luxemburgi értékek megfeleltethetők a magyar 1991-2017 közötti mortalitási rátáknak. Az így meghatározott előrejelzéseket összevetve a tényleges magyar adatok alakulásával, a centrált logaritmizált halálozási valószínűségekre ($m_{x,t}$) számolt MSE értéke 0,0339. Ugyanezen előrejelzés elkészítéséhez Lee-Carter modellt alkalmazva a magyar adatokra, az MSE mutató értéke 0,1096. Ebben az esetben is tehát, az MSE mutató értéke alapján elmondható, hogy a hasonlósági modell pontosabb előrejelzést képes készíteni, mint a Lee-Carter modell.

Érdeemes összehasonlítani az eddigi két előrejelzés Lee-Carter modellre számolt MSE értékeit: az első, 1999-ig tartó előrejelzés MSE értéke férfiak esetén 0,0393 volt, míg a második, 2017-ig tartó előrejelzés során ugyanezen mutató értéke 0,1096. Látható, hogy a második esetben a Lee-Carter modell nagyobb hibát vétett. A nagyobb hibázás a Lee-Carter modell azon tulajdonságából adódik, hogy a modellben kiemelkedően fontos a tanító időszakok száma. Mindkét esetben az 1950-1990

közötti évek töltötték be a tanító időszak szerepét. Az első esetben mindössze 9 évre készítettünk előrejelzéseket, míg a második előrejelzés során már jóval hosszabb, 27 időszakra jeleztük előre a mortalitási rátákat, ugyanannyi tanító időszak mellett. Érthető tehát, hogy a második esetben nagyobb az MSE érték, hiszen a Lee-Carter modell ugyanannyi tanító időszak ismeretében egyre hosszabb távú előrejelzések elkészítéséhez egyre pontatlanabb, bizonytalanabb predikciókat képez. Ez is tekinthető a modell egyfajta hátrányának, hiszen a tanító időszakok száma korlátozza a Lee-Carter modellel készített előrejelezhető évek számát, szemben a hasonlósági modellel, ahol ez a korlát közvetlenül nem áll fenn.

A két modell által előrejelzett mortalitási rátákat és a tényleges magyar adatok alakulását, valamint ezek összehasonlítását a 8. ábra szemlélteti 1996-ra.



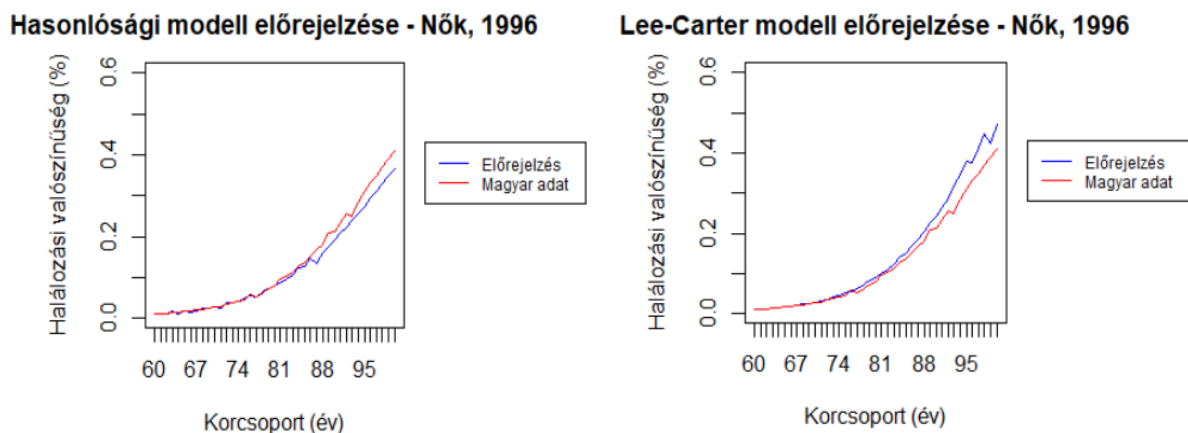
8. ábra: Előrejelzések alakulása férfiak esetében 1996-ban, hasonlósági és Lee-Carter modell alapján
Forrás: saját szerkesztés

Látható, hogy amíg a Lee-Carter modell a 70-75 éves korosztálytól kezdődően felülbecsli a halálozási valószínűségeket, addig a hasonlósági modell egész jól közelíti a tényleges férfi mortalitási rátákat, mindössze egy kisebb felülbecslést végez a 75-80 éves korcsoportban.

A nők vonatkozásában, a hasonlósági modell alapján 1963-ban Új-Zéland halálozási rátái hasonlítottak leginkább az 1990-es magyar adatokhoz. A hasonlósági modellre számolt MSE értéke 0,0147, mely 2 századdal jobb, mint a Lee-Carter modellre számolt MSE mutató 0,0355-ös értéke. Tehát ebben az esetben is az MSE értékek alapján a hasonlósági modell alkalmazása tűnik célszerűbbnek. Az előrejelzett halálozási valószínűségek alakulását a 9. ábra szemlélteti 1996-ra.

A vizualizált értékek alapján azonban úgy tűnik, nem tehető jelentős különbség a két módszer között. Mindkét modell jól becsli a 60-85 év közötti korcsoport értékeit, az idősebb

korosztályokban azonban eltérés tapasztalható. Míg a hasonlósági modell alul-, addig a Lee-Carter modell túlbecsli a 85-100 éves korosztály halálozási rátáit. Ahogy a 6.2. fejezetben bemutattuk, nyugdíjrendszer tervezés szempontjából az alulbecslés jelenti a kisebb hibát, így e tekintetben a hasonlósági modellt érdemes alkalmazni. Továbbá, szemmel láthatóan, a Lee-Carter modell nagyobb mértékben becsül túl, mint a hasonlósági modell alul. Azaz elmondható, hogy ebben az esetben is a hasonlósági modell képez pontosabb predikciókat, de gyakorlatilag a két módszer közül bármelyik modell alkalmazása megfelelő a nők 1991-2017 közötti mortalitási rátáinak előrejelzésére.



9. ábra: Előrejelzések alakulása nők esetében 1996-ban, hasonlósági és Lee-Carter modell alapján
Forrás: saját szerkesztés

Az 1991-2017-re készített előrejelzések összefoglalását a 2. táblázat tartalmazza:

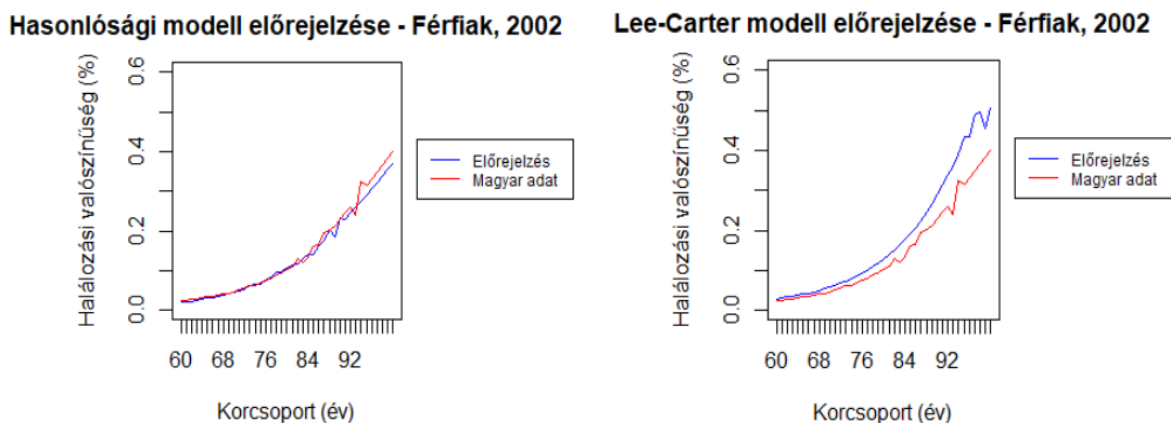
	férfi	női
Ország	Luxemburg	Új-Zéland
MSE	0,0339	0,0147
Lee-Carter MSE	0,1096	0,0355

2. táblázat: Az 1991-2017 közötti mortalitási ráták összefoglaló táblázata
Forrás: saját szerkesztés

6.4. A 2000-2017 közötti időszak előrejelzése

Kutatásom utolsó előrejelzése a 2000-2017-es évek mortalitási rátáinak alakulása. Férfiak esetében a hasonlósági modell alapján 1981-ben Finnország mortalitási rátái hasonlítanak leginkább a magyarországi férfiak 1999-es értékeihez. Ennek értelmében Finnország 1982-1999 közötti időszaka megfeleltethető a magyar mortalitási ráták 2000-2017 közötti alakulásának.

Férfiak esetében a hasonlósági modellre számolt MSE értéke 0,0205, míg ugyanezen mutató értéke a Lee-Carter modell esetén már 0,1036. Látható, hogy MSE mutató tekintetében ez esetben is a hasonlósági modell teljesített jobban. Az előrejelzést vizuálisan is megjelenítve 2002-re, az MSE értékeket reprezentáló eredmény tapasztalható, melyet a 10. ábra szemléltet.



10. ábra: Előrejelzések alakulása férfiak esetében 2002-ben, hasonlósági és Lee-Carter modell alapján
Forrás: saját szerkesztés

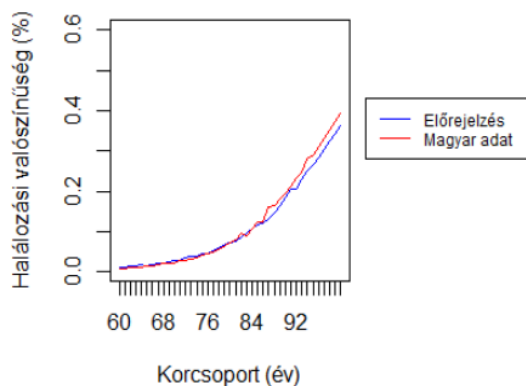
A hasonlósági modell előrejelzett értékei férfiak esetében 85-90 éves korosztályig elég szépen közelítik a tényleges magyar adatok alakulását, e felett pedig némileg alulbecsül. A Lee-Carter modellel készített predikciók gyakorlatilag minden korcsoportban felülbecsülnek, méghozzá jelentős mértékben. Tehát a 2000-2017 közötti időszak halálozási rátáinak előrejelzését – férfiak esetén – egyértelműen érdemesebb hasonlósági modell alapján elvégezni.

Nők esetében a hasonlósági modell alapján 1981-ben Skócia értékei hasonlítanak leginkább a magyarországi nők 1999-es mortalitási rátáihoz. Ennek alapján, a skót nők halálozási valószínűségeinek 1982-1999 közötti alakulása megfeleltethető a magyar nők 2000-2017 közötti értékeinek. A hasonlósági modellre számolt MSE értéke 0,0277, míg Lee-Carter modell esetén 0,0342. Látható, hogy ez esetben már nincsen számottevő különbség a két módszer között MSE érték alapján, hiszen a hasonlósági modell mindössze 6 ezreddel teljesített jobban. A két modell által előrejelzett mortalitási rátákat és a tényleges magyar adatok alakulását, valamint ezek összehasonlítását a 11. ábra szemlélteti 2002-re.

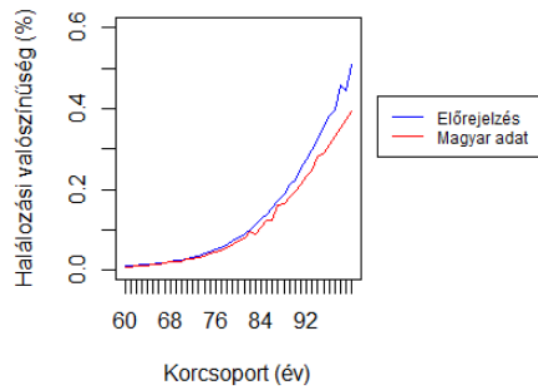
A hasonlósági modell 85 éves korosztályig elég pontos predikciókat képez, az idősebb korosztályokban némileg alulbecsül (nyugdíjrendszer tervezéshez ez jelenti a kisebb hibát, lásd 6.2. fejezet), de továbbra is jól közelíti a tényleges adatok alakulását. A Lee-Carter modell 80 éves korosztályig szintén elég precíz előrejelzéseket készít, a 80-100 éves korcsoportban azonban

túlbecsüli a tényleges értékeket. Ennek következtében, bár MSE érték alapján úgy tűnt, nincs jelentős különbség a két módszer között, nők esetében is érdekesebb a hasonlósági modellel előrejelezni.

Hasonlósági modell előrejelzése - Nők, 2002



Lee-Carter modell előrejelzése - Nők, 2002



11. ábra: Előrejelzések alakulása nők esetében 2002-ben, hasonlósági és Lee-Carter modell alapján
Forrás: saját szerkesztés

Az egyes modellek segítségével készített, 2000-2017-re vonatkozó előrejelzések összefoglalását a 3. táblázat tartalmazza:

	férfi	női
Ország	Finnország	Skócia
MSE	0,0205	0,0277
Lee-Carter MSE	0,1096	0,0342

3. táblázat: Az 2000-2017 közötti mortalitási ráták összefoglaló táblázata
Forrás: saját szerkesztés

7. ÖSSZEFOGLALÁS

A 7. fejezetben a 3 időszakra tett előrejelzések eredményeinek kiértékelése következik. Beszámolok arról, hogy hogyan teljesített az általam készített hasonlósági modell az egyes előrejelzések során. Továbbá, kitérek kutatásom kezdeti feltételezésére is, miszerint fejlődő államok halálozási rátái előrejelezhetőek egy fejlett állam múltjából. Bemutatom, hogy hipotézisem az elemzésem során igaznak bizonyult-e, illetve amennyiben nem, ott mi okozta az eltérést.

Először egy összefoglaló tábla segítségével szemléltetem, hogy a két modell az egyes előrejelzések során mennyire készített pontos előrejelzéseket, összehasonlítva a két modellre számolt átlagos négyzetes hibák értékeit. A két modellre számított MSE értékeket a 4. táblázat foglalja össze.

	1. előrejelzett időszak: 1991-1999		2. előrejelzett időszak: 1991-2017		3. előrejelzett időszak: 2000-2017	
	férfi	nő	férfi	nő	férfi	nő
Illeszkedő ország	Szlovákia	Litvánia	Luxemburg	Új-Zéland	Finnország	Skócia
Hasonlósági modell MSE	0,0103	0,0094	0,0339	0,0147	0,0205	0,0277
Lee-Carter MSE	0,0393	0,0194	0,1096	0,0355	0,1096	0,0342
Eltérés Lee-Carter modellhez képest	-0,0290	-0,0100	-0,0757	-0,0208	-0,0891	-0,0065

4. táblázat: A hasonlósági és Lee-Carter modellre számolt MSE értékek összefoglaló táblázata

Forrás: saját szerkesztés

A 4. táblázatban dőlten jelöltem azokat az MSE mutatókat, ahol a hasonlósági modell jobban teljesített, mint a Lee-Carter modell. Látható, hogy mindhárom előrejelzés során, mindkét nem esetében a hasonlósági modellre számolt MSE értékek alacsonyabbak. Elmondható tehát, hogy az általam épített hasonlósági modell valamennyi esetben jobban teljesített, mint a Lee-Carter modell. Az eltérés nagyságrendileg egyedül a 3. időszakban, nők esetén nem számottevő, hiszen ott mindössze 6 ezrednyi különbségről beszélhetünk a két modell tekintetében. A többi predikció során azonban a hasonlósági modell már akár 1-2 századdal is jobban teljesített. A női populációnak volt

Magyarországon egy normálisabb mortalitásjavulása az 1990-es és 2000-es években, míg a férfiaké nagyon hirtelen gyors volt. Így érthető, hogy a férfiak esetében teljesít rosszabbul a matematikai alapú modell.

A 6. fejezetben illusztratív ábrákkal szemléltettem, hogyan alakultak az egyes modellek előrejelzései a tényleges adatokhoz képest. Láthattuk, hogy a Lee-Carter modell a 60-70-80 éves korcsoportokra elég precíz előrejelzéseket tudott készíteni, ugyanakkor az idősebb korosztályokban jelentős mértékben túlbecsülte a mortalitási rátákat. Bár a Lee-Carter modell a nőknél is felülbecsült, férfiak esetében sokkal nagyobb mértékű eltérések figyelhetők meg. Ennek oka, hogy a Lee-Carter modell extrapolatív eljárás, mely predikcióit a háttérben álló (orvostudományi, egészségügyi, társadalmi) okok figyelembevétele nélkül teszi (Vékás, 2016). A hasonlósági modell esetében viszont nem egy matematikai-statisztikai alapú algoritmus predikcióira hagyatkoztunk, hanem tényleges múltbeli adatokat használtunk fel, mely adatok magukban hordozzák a társadalmi tényezőket is. Ennek következtében, a hasonlósági modell sosem becsülte túlságosan alul vagy felül a tényleges adatok alakulását.

Összességében elmondható tehát, hogy a hasonlósági modell jobban teljesített, mint a Lee-Carter modell. Ezt alátámasztják az eltérések mértékét mérő MSE értékek, valamint a grafikonokon vizuálisan megjelenített értékek is. Egyetlen ellenőrizendő állítás maradt hátra, nevezetesen kutatásom azon hipotézise, miszerint egy fejlődő ország mortalitási rátáinak alakulása egy fejlett állam múltjából jelezhető előre.

Feltételezésemet az eredményeim igazolták. Bár hipotézisem az első, 1991-1999 közötti időszakra vonatkozó előrejelzés esetén helytelennek tűnik, hiszen a modell két olyan államot határozott meg, melyek nem sorolhatóak maradéktalanul a fejlett államok sorai közé. Az eltérés azonban megindokolható az adott 1981-es év háttérében zajló történelmi körülmények figyelembevételével, melynek ismertetésére a 6.2. fejezetben tértem ki. A hasonlósági modell azonban a 2. (1991-2017 közötti) és 3. (2000-2017 közötti) előrejelezni kívánt időszakra egyértelműen fejlett országokat határozott meg, mind férfiak, mind nők esetén. Ennek értelmében, fejlett államok által már korábban bejárt mortalitási rátáinak pályája megfeleltethető egy 10-20-30 évvel későbbi, fejlődő ország jövőbeli alakulásának.

Ugyanakkor érdemes megvizsgálni, hogy milyen korlátai vannak a modellnek és ezzel párhuzamosan milyen fejlesztési lehetőségek állhatnak rendelkezésre. A hasonlósági modellnek

egyértelmű korlátja, ha nagyon hosszú időszakot, például 100 évet szeretnénk előrejelezni. Ekkor ugyanis egyáltalán nem biztos, hogy létezik 100 évvel korábbi halandósági adat más országokra, illetve, ha létezik is, akkor sem igazán reális azon hipotézis, miszerint a mai magyar halandóság megfeleltethető egy másik ország 100 évvel ezelőtti múltjának. A problémát úgy lehetne orvosolni, hogy az előrejelzés első 30 évét megadjuk a hasonlósági modellből és utána ezen adatokat figyelembe véve felépítünk egy Lee-Carter modellt, és abból jelezzük előre a maradék 70 évet. Ez lehet a hasonlósági modell egyik továbbfejlesztési iránya.

Másik továbbfejlesztési lehetőség, hogy a modellt nemcsak Lee-Carter modellel hasonlítjuk össze, hanem az eljárás rotációval bővített változatával is, például Li, Lee és Gerland (2013) tanulmányával. Emellett érdemes olyan újfajta megközelítéseket is figyelembe venni, melyek statisztikai modellek helyett gépi tanulást alkalmaznak mortalitási ráták előrejelzésére. A szakirodalom egyik ilyen jelentős munkája Richman és Wüthrich (2019) szerzőpárhoz köthető, akik rekurrens neurális hálózatokkal (recurrent neural networks, RNNs) jelezték előre svájci mortalitási ráták alakulását. Ezért fejlesztési lehetőségként azt is lehetne vizsgálni, hogyan teljesít a hasonlósági modell a neurális hálózathoz képest.

További fejlesztési irány lehetne, hogy a modellt nem csak Magyarországra, hanem a közép-kelet-európai régió egyéb országaira is leteszteljük, például Szlovákiára, Szlovéniára, Horvátországra, Romániára. Ezáltal több fejlődő országra is ellenőrizni tudnánk, hogy a hasonlósági modell megfelelő predikciókat készít-e, valamint ténylegesen fejlett államok mortalitási rátáiból jelez-e előre.

IRODALOMJEGYZÉK

- Balázsne Mócsai, A. (2020). *Hipotézisvizsgálat alapok*. Prezentációs diasor. Budapest: Budapesti Corvinus Egyetem
- Banyár, J. (2016). *Életbiztosítás*. Budapesti Corvinus Egyetem
- Bland, J. M., & Altman, D. G. (2004). The logrank test. *British Medical Journal*, 328(7447), 1073. <https://doi.org/10.1136/bmj.328.7447.1073>
- Carter, L. R., & Prskawetz, A. (2001). *Examining structural shifts in mortality using the Lee–Carter method*. Working paper. Rostock: Max Planck Institute for Demographic
- Csallner, A. E., & Vincze, N. (n.d). *Bevezetés a valószínűség-számításba és a matematikai statisztikába*. Dr. Csallner András Erik, Vincze Nándor: Bevezetés a valószínűség-számításba és a matematikai statisztikába. http://www.jgypk.hu/tamop15e/tananyag_html/statisztika/index.html
- Horbulák, Zs. (2018). Gazdasági reformkísérletek Csehszlovákiában a szocializmus idején. *Világtörténet*, 8(3), 399 – 414.
- Human Mortality Database, (n.d.). *The Human Mortality Database*. www.mortality.org
- Hunyadi, L., & Vita, L. (2019). *Statisztika II..* Akadémiai Kiadó. <https://doi.org/10.1556/9789634542223>
- Hyndman, R. J., & Ullah, M. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, 51(10), 4942 – 4956.
- Kenéz, P. (2008). *A Szovjetunió története - A kezdetektől az összeomlásig*. Akkord Kiadó Kft.
- Lee, R. D., & Carter, L. R. (1992). Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, 87(419), 659-671. <https://doi.org/10.2307/2290201>
- Li, N., & Gerland, P. (2011). *Modifying the Lee-Carter method to project mortality changes up to 2100*. Paper. Washington, DC: Population Association of America.

- Li, N., Lee, R., & Gerland, P. (2013). Extending the Lee-Carter Method to Model the Rotation of Age Patterns of Mortality Decline for Long-Term Projections. *Demography*, 50(6), 2037- 2051. <https://doi.org/10.1007/s13524-013-0232-2>
- Májér, I., & Kovács, E. (2011). Élettartam-kockázat – a nyugdíjrendszerre nehezedő egyik teher. *Statisztikai Szemle*, 89(7 – 8), 790 – 812.
- R Foundation, (n.d.). *What is R?*. <https://www.r-project.org/about.html>
- Richman, R. & Wüthrich, M.V. (2019). Lee and Carter go Machine Learning: Recurrent Neural Networks. *Neuroeconomics eJournal*. <https://dx.doi.org/10.2139/ssrn.3441030>
- The Economist Intelligence Unit. (2014). *Health outcomes and cost: A 166-country comparison*. White paper. <https://stateofreform.com/wp-content/uploads/2015/11/Healthcare-outcomes-index-2014.pdf>
- Tuljapurkar, S., Li, N., & Boe, C. (2000). A Universal Pattern of Mortality Change in the G7 Countries. *Nature*, 405(6788), 789 – 792.
- Vékás, P. (2016). *Az élettartam-kockázat modellezése*. Doktori értekezés. Budapest: Budapesti Corvinus Egyetem
- Wang, J. Z. (2007). *Fitting and Forecasting Mortality for Sweden: Applying the Lee–Carter Model*. Working paper. Stockholm: Stockholm University
- Wiśniowski, A., Smith, P. W. F., Bijak, J., Raymer, J., & Forster, J. J. (2015). Bayesian Population Forecasting: Extending the Lee – Carter Method. *Demography*, 52(3), 1035 –1059. <https://doi.org/10.1007/s13524-015-0389-y>

MELLÉKLETEK

1. melléklet

Shiny App - Mortalitási ráták előrejelzése

Válasszon évszámot!

Válassza ki az országot!

LUX

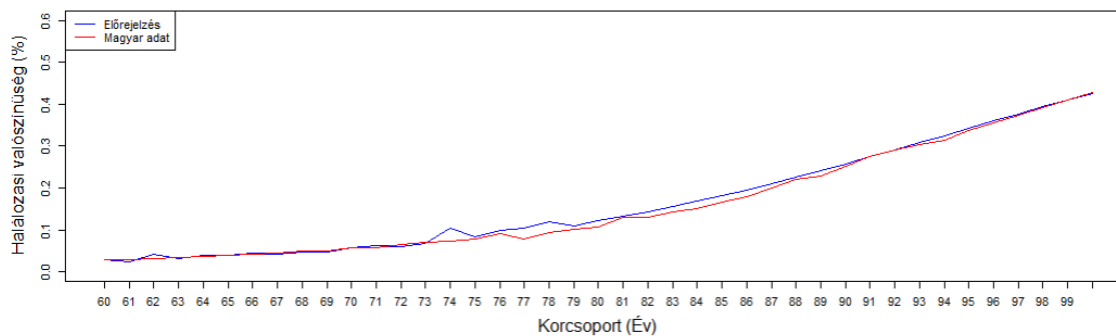
Válassza ki az előrejelzni kívánt periódusok számát!

Válasszon nemet!

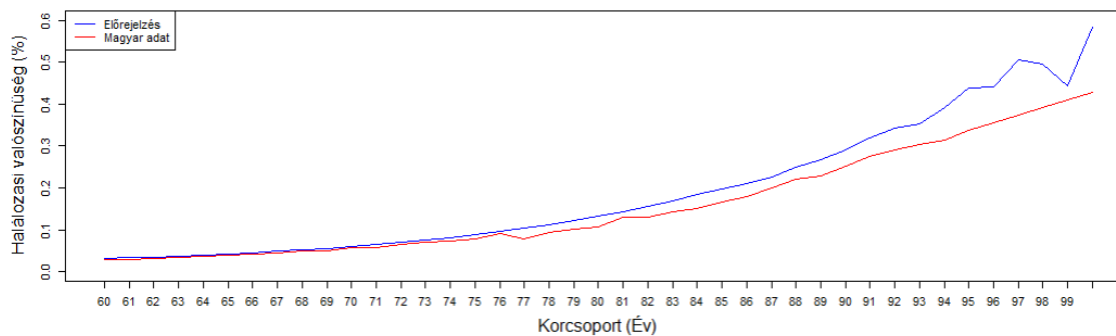
férfi

Előrejelzés

Hasonlósági modell előrejelzése - Férfiak 1996



Lee-Carter modell előrejelzése - Férfiak 1996



Az appot készítette: Fabók Judit Bőborka, 2022

1. melléklet: R Shiny applikáció grafikus felhasználói felülete
Előrejelzések alakulása férfiak esetében 1996-ban, hasonlósági és Lee-Carter modell alapján
Forrás: saját szerkesztés

2. melléklet

Shiny App - Mortalitási ráták előrejelzése

Válasszon évszámot!

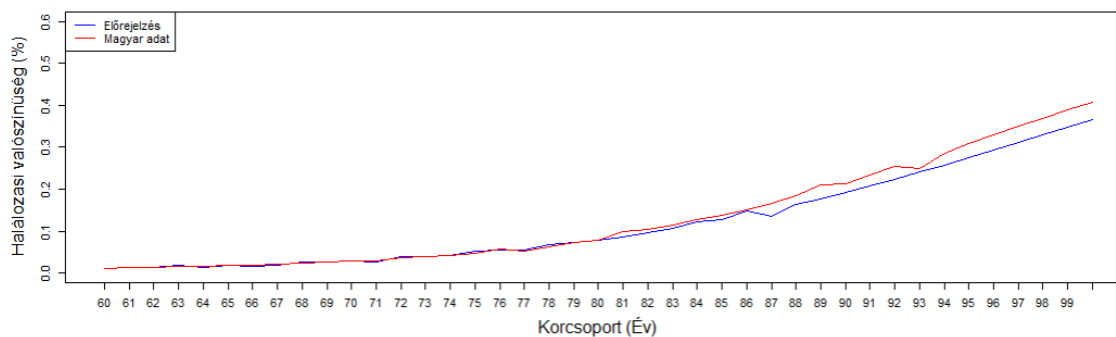
Válassza ki az országot!

Válassza ki az előrejelzni kívánt periódusok számát!

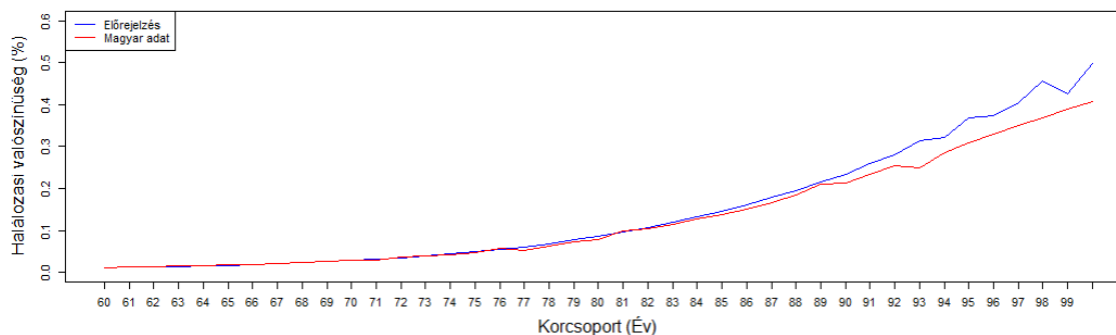
Válasszon nemet!

Előrejelzés

Hasonlósági modell előrejelzése - Nők 1996



Lee-Carter modell előrejelzése - Nők 1996



Az appot készítette: Fabók Judit Bőrkork, 2022

2. melléklet: R Shiny applikáció grafikus felhasználói felülete
Előrejelzések alakulása nők esetében 1996-ban, hasonlósági és Lee-Carter modell alapján
Forrás: saját szerkesztés