

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
BUDAPESTI CORVINUS EGYETEM

Horváth Roland

MAGYAR HALANDÓSÁGI TÁBLÁK
ELŐREJELZÉSE MULTIPOPULÁCIÓS
MODELLEKKEL

MSc Diplomamunka

Témavezető:

Vékás Péter

Budapesti Corvinus Egyetem
Operációkutatás és Aktuáriustudományok Tanszék



Budapest, 2016. december

Köszönetnyilvánítás

Ezúton is szeretném megköszönni Vékás Péternek, hogy elvállalta a konzulensi feladatokat. A téma kiválasztása mellett a szakmai beszélgetések és tanácsok nagyban segítettek szakdolgozatom elkészültét. Hálával tartozom továbbá családomnak, barátaimnak és csoporttársaimnak, akik egyetemi éveim során végig támogató és inspiráló közeget biztosítottak, felejthetlenné téve ezt az időszakot.

Tartalomjegyzék

1. A halandóság modellezésének alapjai	7
1.1. Módszertani alapok	7
1.1.1. A halandóság folytonos modellezése	7
1.1.2. Diszkrét modell és halandósági tábla	8
1.1.3. A halandóság becslése	10
1.2. Lee-Carter modell	12
1.2.1. A modell és interpretálása	12
1.2.2. Paraméterek becslése	14
1.2.3. Előrejelzés	14
2. Az általánosított modellkeret	16
2.1. A Lee-Carter modell továbbfejlődése	16
2.2. A GAPC modellcsalád	18
2.2.1. Poisson Lee-Carter modell	20
2.2.2. Renshaw-Haberman modell	21
2.2.3. APC modell	21
2.2.4. Cairns-Blake-Dowd modell	21
2.2.5. M7 modell	22
2.2.6. Plat modell	22
2.3. Becslés és előrejelzés	23
2.3.1. Paraméterbecslés és modellválasztás	23
2.3.2. Előrejelzés	23
3. Többpopulációs módszerek	26
3.1. Motiváció	26
3.2. A standard modellek kiterjesztése két populációra	28
3.2.1. Modellfelírások és paraméterbecslés	29
3.2.2. Sztochasztikus faktorok	31

3.3. További többpopulációs módszerek	33
4. Numerikus elemzés	35
4.1. Adatok és a modellezés menete	35
4.2. Kétpopulációs modellezés és eredményei	39
4.2.1. Magyar-osztrák férfi	40
4.2.2. Magyar-szlovák férfi	43
4.2.3. Magyar-osztrák női	44
4.2.4. Magyar-szlovák női	46
4.2.5. Magyar férfi-magyar női	47
4.3. Összehasonlítás az egypopulációs modellek eredményeivel, alkalmazások . .	48
4.3.1. A legjobb magyar mortalitási modellek	48
4.3.2. Magyar, osztrák és szlovák mortalitás előrejelzések 2040-ig	50
4.3.3. Életjáradék árazása és várható élettartam	52
4.4. Kitekintés és konklúzió	53
Összefoglalás	55
Irodalomjegyzék	56

Bevezetés

Mennyi ideig fogok még élni? – teszi fel a kérdést magában sok ember akár nap, mint nap, legyen fiatal vagy idős. És teszik fel időről időre a kérdést élet- és nyugdíjbiztosítók aktuáriusai, akik az ügyfeleikről szeretnék tudni, hogy – legalábbis várhatóan – meddig élnek, és így milyen jövőbeli kiadásokkal és bevételekkel számolhatnak.

A kérdés megválaszolása azonban nem csak egyéni szempontból reménytelen, de sokévi megfigyeléseken alapuló statisztikai előrejelzések sem teljesítenek még mindig olyan bizonyossággal, melyet bátran fel lehetne használni egy több évtizedes halandósági vagy más néven mortalitási projekciónál. Az aktuáriusi szakirodalomban *longevity risk*-ként ismert¹, a várható élettartamban rejlő ilyen jellegű kockázat modellezése ma a biztosítási terület kockázati modellezésének egyik legizgalmasabb témája.

Az életbiztosítások kapcsán nem túl gyakori a pozitív gazdaság-politikai jelenségként fellépő kockázat, a *longevity risk* viszont ilyen. A várható élettartam ugyanis rohamos léptékben nőtt lényegében minden fejlett országban az elmúlt évtizedekben. Magyarországon 1950-ben a nők születéskor várható élettartama 64, a férfiaké 59 év volt. 2014-ben ez a két szám 79 és 72 év. Ez a trend ráadásul globális jelenség. Wilson (2001) tanulmánya szerint egy egyértelmű konvergencia figyelhető meg a halandósági folyamatokban, különösen azon nemzetek populációi között, melyek a globalizáció révén egyre szorosabb kapcsolatban élnek és fejlődnek egymással.

Magyarország esetén érdekesnek látszik megvizsgálni, hogy felírható-e bármilyen konvergencia a nyugati vagy szomszédos országok mortalitásaival, különösen más Európai Unió tagországokéval. Több populáció halandóságának együttes modellezése viszont nem egyszerű feladat, a vonatkozó modellek csupán néhány éves múlttra tekintenek vissza. Legjobb tudomásom szerint dolgozatomban lesz az első olyan írás, mely többpopulációs módszerekkel vizsgálja hazánk mortalitását. Ehhez végül Ausztria és Szlovákia férfi illetve női népességének halandóságát „vettem segítségül”, és több esetben is sikerült jobban teljesítő előrejelzést adni, mint az egypopulációs modellekkel.

¹A magyar szakirodalomban használatos még a Kovács és Májer (2011) által meghonosított *élettartam-kockázat* szakzsargon is.

A nemzetközi modellezés különösen hatékony módszer lehet a multinacionális életbiztosítóknak, hiszen amennyiben két biztosított részsokaság élettartam-kockázatában valóban fellelhető összefüggőség, úgy ezen kitettség kezeléséhez bizonyosan hatékonyabb módszer az együttes modellezés, mint az egymástól független. Különösen égető probléma ez a mai S2-környezetben illetve az alacsony hozamok mellett, ahol a befektetések közel sem tudják kompenzálni az élettartamok rohamos, egyelőre elég kiszámíthatatlannak bizonyuló fejlődését.

A dolgozat felépítése a következő. Az 1. fejezetben ismertetem a halandóság modellezésének alapjait, illetve az úttörő 1992-es Lee-Carter modellt. A 2. fejezetben mutatom be, hogy a mortalitási modellezés egy populációra hova is fejlődött az elmúlt bő két évtizedben, és hogy hogyan foglalhatóak a ma legjobbnak vélt modellek egy keretbe matematikailag. A 3. fejezetben szólok ezen hat ismertetett modell kiterjesztéséről kétpopulációs esetre, illetve a legfontosabb ma létező irányzatokról a témában. A 4. fejezetben mutatom be a saját elemzést, és vonom le a megfelelő következtetéseket a magyar, illetve az osztrák és szlovák mortalitásokra is. Ezt követően lezárásként összefoglalom a dolgozat eredményeit.

1. fejezet

A halandóság modellezésének alapjai

Ebben a fejezetben a halandósági modellezés alapjait mutatom be, majd a mérföldkőnek számító Lee-Carter modellt részletesen tárgyalom.¹ Ez jelenti minden ma használt legfejlettebb modell módszertani és heurisztikus alapját is. Ezeket a modelleket a későbbi fejezetekben fogom részletesebben tárgyalni, hogy aztán a kellő matematikai alapok után bemutathassam saját elemzésemet a magyar, osztrák és szlovák halandósági adatokon.

1.1. Módszertani alapok

A modellek bemutatásához első lépésben az alapfogalmakat, jelöléseket és összefüggéseket ismertetem. Dolgozatomban végig igyekszem a szakirodalomban elterjedt jelölésekkel dolgozni, és ezeket konzisztensen használni a különböző modellekben.

1.1.1. A halandóság folytonos modellezése

A mortalitási modellezés egyik legfontosabb fogalma a *túlélési függvény* (angolul *survival function*), mely az egyén élettartamának eloszlását leíró függvény. Jelölje L az élettartamot, ami tehát egy nemnegatív, folytonos valószínűségi változó. Ezen valószínűségi változó túlélési függvényén azt a $G : \mathbb{R}^+ \rightarrow [0, 1]$ függvényt értjük, melyre:

$$G(y) = \mathbb{P}(L \geq y) \quad (y \geq 0). \quad (1.1)$$

Nyilvánvalóan fennállnak az alábbi összefüggések:

$$G(0) = 1,$$

¹Az itt olvasható elméleti alapok Vékás (2016) munkájára épülnek.

$$G(y) = 1 - F(y) \quad (y \geq 0),$$

ahol F jelöli L eloszlásfüggvényét. Kiemelt fontosságú emellett a hátralévő élettartam valószínűségeloszlása: ha az egyén megélt $x \geq 0$ életkort, akkor a számára még hátralévő $L - x$ élettartam eloszlását az $L \geq x$ feltétel melletti feltételes eloszlás adja. Ezt a valószínűséget fejezi ki a *reziduális túlélési függvény*:

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

Ahogy azt a Bevezetésben említettem, számunkra különösen érdekes lesz a *várható hátralévő élettartam*. Az x kort megélt egyén számára ez valószínűségelméletileg természetesen a feltételes várható értékkel fejezhető ki, ami pedig továbbírható integrál segítségével:

$$e_x = \mathbb{E}(L - x | L \geq x) = \int_x^\infty G_x(y) dy \quad (x \geq 0).$$

e_x tehát az az időmennyiség, melyet várhatóan egy jelenleg x éves egyén még élni fog, amennyiben élettartamának reziduális túlélési függvénye G_x .

Kitüntetett szerepet kap a *születéskor várható élettartam*:

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

Folytonos modellkeretnél maradva a pillanatnyi halálozási valószínűség nyilván 0. Hátértékben azonban vizsgálhatjuk a halálozási valószínűséget. Ehhez szokás definiálni az ún. *halálozási intenzitást* (angolul *hazard rate* vagy *force of mortality*):

$$\mu(y) = \lim_{\epsilon \rightarrow 0^+} \frac{\mathbb{P}(L < y + \epsilon | L \geq y)}{\epsilon} \quad (y \geq 0).$$

A halálozási intenzitás tehát felfogható úgy, mint a pillanatnyi halálozási valószínűség évesített értéke. Felhasználva a feltételes valószínűség definícióját, és bővítve a túlélési függvénnyel kaphatjuk a következő, gyakorlati szempontból hasznos összefüggést:

$$\mu(y) = \lim_{\epsilon \rightarrow 0^+} \frac{F(y + \epsilon) - F(y)}{\epsilon G(y)} = \frac{f(y)}{G(y)} \quad (y \geq 0, G(y) > 0),$$

ahol továbbra is F az élettartam eloszlásfüggvénye, G a túlélési függvény, illetve f jelöli a sűrűségfüggvényét.

1.1.2. Diszkrét modell és halandósági tábla

A gyakorlatban aktuáriusi alkalmazásoknál lényegében kivétel nélkül diszkrét halálozási adatokkal dolgozunk, a legtöbbször egyéves időtávokat tekintve. Ehhez az alábbiakban

definiálom a *koréves túlélési* illetve a *koréves halálozási valószínűségeket*, melyek azt a valószínűséget fejezik ki, hogy az x éves kort éppen betöltő egyén a következő születésnapját megéli-e:

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

vagy sem:

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

A definícióból nyilvánvalóan $p_x + q_x = 1 \quad \forall x \in \mathbb{N}$ esetén. Továbbá a túlélési függvényt felhasználva is meghatározható a koréves túlélési valószínűség:

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

Az egészértékű felírás haszna a halandósági táblák használatakor látszik. Ezen táblák kor- és naptári évekre bontva tartalmazzák többek közt a koréves túlélési és halálozási valószínűségeket, melyre példát az alábbi ábrán láthatunk:

Kor	qx	lx	dx	ex
0	0,00502	100000	502	72,26
1	0,00043	99498	43	71,62
2	0,00022	99455	22	70,65
3	0,00024	99433	24	69,67
4	0,00015	99410	15	68,68
5	0,00014	99395	14	67,69
6	0,00014	99381	14	66,7
7	0,00006	99367	6	65,71
8	0,00016	99362	15	64,72
9	0,00006	99346	6	63,73
10	0,00014	99340	14	62,73
11	0,00016	99326	16	61,74
12	0,00014	99310	14	60,75
13	0,00018	99296	18	59,76
14	0,00022	99278	22	58,77
15	0,00026	99256	26	57,78
16	0,00033	99229	33	56,8
17	0,00035	99197	35	55,81
18	0,00033	99162	32	54,83
19	0,00051	99130	51	53,85
20	0,00061	99079	61	52,88

1.1. ábra. A 2014-es magyar férfi halandósági tábla 0-20 éves korig. (Adatok forrása: mortality.org)

A halandósági tábla minden esetben el van vágva egy felső korláttal, melyet ω -val szokás jelölni. A mortality.org táblái esetén ez 110, vagyis a 110 évesekre már azt tesszük fel, hogy mindannyian ebben a korban halnak meg, azaz $q_{110} = 1$. Ez nem egy erős megszorítás, hiszen például hazánkban 2014-ben két férfi és három nő élt ennél magasabb életkorban, vagyis modellszempontról elhanyagolható létszámban.

A halandósági tábla tartalmazza még az alábbi mutatókat:

- $l_x = l_0 \cdot G(x)$ *továbbélési rendet*, mely 100.000 főre vetítve mutatja meg, várhatóan hányan élnek meg az x éves kort. Itt súlyos feltételezés – és ezzel később még foglalkozom –, hogy az idő előrehaladtával a korévenkénti halálozási valószínűségeket változatlanoknak tekintjük ezen mutatónál.
- $d_x = l_{x+1} - l_x$ *kihalási rendet*, mely az adott korosztály várható elhunytjainak számát jelöli szintén 100.000 fős kezdeti népességre vetítve, és az említett feltételezéssel élve.
- Végül $e_x = \frac{1}{l_x} \sum_{i=x+1}^{\omega} l_i + \frac{1}{2}$ várható hátralévő élettartamot.

A mortalitási modellezés célja az, hogy a jövőbeli q_x korévenkénti halálozási valószínűségeket megbecsüljük. Ebből pedig már többek közt várható élettartamokat vagy például egy járadéktermék díját is könnyen kiszámolhatjuk. Mielőtt a konkrét modelleket elkezdném bemutatni, tekintsük át ennek a valószínűségbecslésnek az alaptechnikáját.

1.1.3. A halandóság becslése

A múltbeli halálozási adatok számszerűsítéséhez leggyakrabban használt statisztika a *mortalitási ráta* (vagy angolul *mortality rate*). Ez a mutató a választott időszak (naptári év), kor és populáció esetén a megfigyelt halálos esetek számának és a populáció létszámának arányaként írható fel:

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

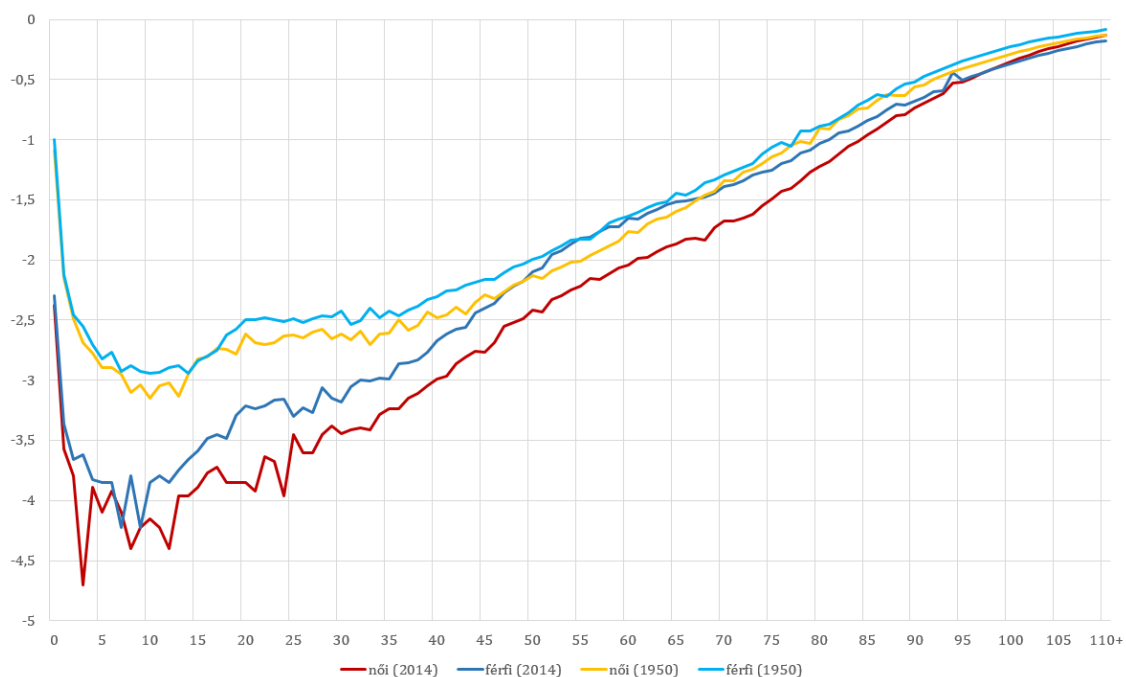
A halálos esetek száma egy egyértelmű mennyiség adott x és t mellett. A populáció létszáma viszont már egy dinamikus mennyiség, hiszen a halálos esetek vagy szerződések megszűnése miatt ez a megfigyelt időszakon belül csökkenhet. Ha az időszak kezdőpontjában vett létszámot vesszük figyelembe, akkor azt *kezdeti kitettségnak* nevezzük (*initial exposure to risk*), jelölje $E_{x,t}^0$. A másik használatos definíció az időszak alatt élt egyének átlagos számát veszi figyelembe, ez a *központi kitettség* (*central exposure to risk*), aminek $E_{x,t}^c$ a szokásos jelölése. Ez az „átlagos” szám persze attól függ, hogy mennyi időt éltek átlagosan a megfigyelt időszakban azok, akik ekkor meg is haltak. Jelölje ezt A . Ez a következő összefüggést implikálja:

$$E_{x,t}^c = E_{x,t}^0 - (1 - A)D_{x,t}. \quad (1.2)$$

Természetesen a leggyakoribb egyszerűsítő feltevés – és én is ezzel fogok élni a dolgozat elemzése során – az $A = \frac{1}{2}$ paraméterezés.

A kitettség definíciójának függvényében beszélhetünk tehát kezdeti ($m_{x,t}^0$) és központi ($m_{x,t}^c$) halandósági rátáról. Az már a modellezés mikéntjétől függ, hogy melyiket használjuk. A definíció választásától függetlenül elmondható viszont, hogy a modellezés során

a fiatal és idős korosztály rátái közötti nagyságrendi különbség miatt ezek logaritmusát vagy logit függvényét fogjuk tekinteni. Az 1.2. grafikonon a magyar adatokon szemléltetem a logaritmált központi mortalitási rátákat. A korábbi klasszikus mortalitási modellek (Gompertz és Makeham vagy Heligman és Pollard konstrukciója) ezen keresztmetszeti függvények formájára próbáltak paraméteres modelleket felírni. Dolgozatomban ezzel nem foglalkozom, de az Olvasónak ajánlom további részletekért a téma egyik alapművét: Dickson, Hardy és Waters (2009) könyvét.



1.2. ábra. Logaritmált központi mortalitási ráták a magyar férfi illetve női populációra 1950-ben és 2014-ben. (Adatok forrása: mortality.org)

A fenti ábrán jól kivehető a fejlődés, vagyis a mortalitások erőteljes csökkenése. Ennek mértéke korosztálonként és nemenként változó: például a csecsemőhalandóság jelentősen visszaszorult, ugyanakkor a nyugdíjazás előtt álló férfiak halandósága alig alacsonyabb a 64 évvel ezelőttinél. Érdekes jelenség még, hogy a huszonéves korban látható halandósági púp –, melyet a fiatalok veszélyesebb, sokszor motorozási szokásainak tulajdonítanak – a nőknél teljesen eltűnt. Továbbá megjegyzendő, hogy 80 éves kor felett az adatokat szolgáltató *Human Mortality Database* a kevés megfigyelés miatt kiegyenlítő eljárást alkalmaz.

A számunkra fontos összefüggés természetesen az lesz, hogy a statisztikailag könnyebben megfogható mortalitási ráta és a $q_{x,t}$ valószínűségek hogyan feleltethetőek meg egymásnak. Többek közt az Ágoston-Kovács (2000) vagy a Vékás (2016) munkákban megtalálható

a könnyű, de hosszadalmas levezetés; az összefüggések pedig:

$$\begin{aligned} q_{x,t} &= \frac{m_{x,t}}{1 + \frac{1}{2}m_{x,t}}, \\ m_{x,t} &= \frac{q_{x,t}}{1 - \frac{1}{2}q_{x,t}}. \end{aligned} \tag{1.3}$$

1.2. Lee-Carter modell

A mortalitás modellezésben mára már klasszikussá vált struktúrárt Ronald D. Lee és Lawrence Carter alkotta meg 1992-ben. Alapötletük az volt, hogy egy log-bilineáris paraméteres modellt állítanak fel a központi mortalitási rátákra mint függő változóra. A következőkben bemutatandó modell meglepően jól írta le az 1933 és 1987 közötti USA-beli mortalitási adatokat, mára azonban már jóval szofisztikáltabb modellekkel készítik az előrejelzéseket, az elméleti alapok viszont ezeknél a struktúráknál is hasonlóak.

1.2.1. A modell és interpretálása

A dolgozat ezen szakaszában végig központi mortalitási rátákkal dolgozom, ezért a ^c indexelést elhagyom. Jelölje tehát $m_{x,t}$ az $x \in \{1, 2, \dots, X\}$ korévhez és $t \in \{1, 2, \dots, T\}$ naptári évhez tartozó központi halálozási rátát. Feltesszük, hogy ezeket az értékeket minden x és t esetén ismerjük, alakulásukat pedig a következő egyenlet írja le:

$$\ln m_{x,t} = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t} \quad (x = 1, 2, \dots, X, \quad t = 1, 2, \dots, T). \tag{1.4}$$

ahol $X \geq 2$ és $T \geq 2$ a korcsoportok és naptári évek darabszámai. A három paraméter közül α_x és β_x fogják meg a kortól való függést, míg κ_t a halandóság időbeli fejlődését testesíti meg. Utóbbit *mortalitási indexnek* is nevezik. Az $\epsilon_{x,t}$ hibatag a modell által meg nem magyarázott hatásokat foglalja magában. A hibatagokról szokás megkövetelni, hogy 0 várható értékű és $\sigma^2 > 0$ varianciájú, egymástól független normális eloszlást kövessenek:

$$\epsilon_{x,t} \sim \mathcal{N}(0, \sigma^2) \quad (\forall x, t), \tag{1.5}$$

Megjegyzendő, hogy bár a szerzők az eredeti cikkben nem tesznek feltételt a peremeloszlásra – csak a 0 várható értéket és a konstans varianciát követelik meg –, a normális eloszlás lehetővé teszi, hogy a cikkben ismertetett eredmények maximum likelihood becsléssel is elérhetőek legyenek.

A normalitásból könnyen látható, hogy $\ln m_{x,t}$ is normális eloszlást követ:

$$\ln m_{x,t} \sim \mathcal{N}(\alpha_x + \beta_x \kappa_t, \sigma^2) \quad (\forall x, t). \tag{1.6}$$

Látható, hogy az (1.4) egyenlet paraméterezése nem egyértelmű, hiszen az alábbi két transzformációra is invariáns:

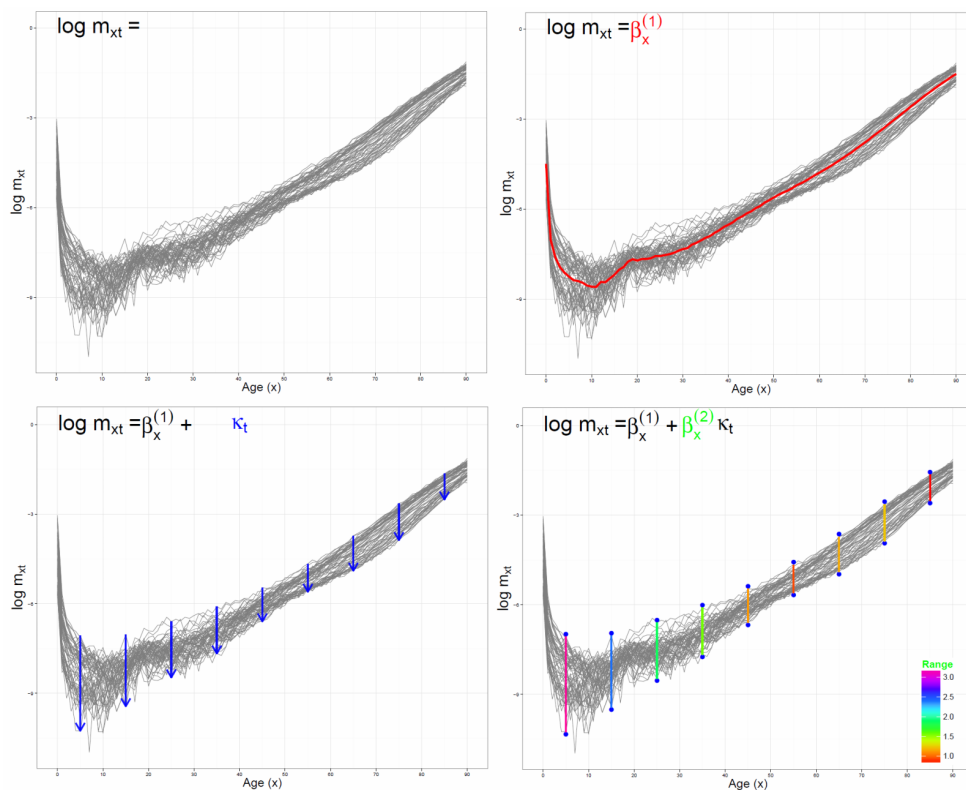
$$\{\alpha_x, \beta_x, \kappa_t\} \rightsquigarrow \left\{ \alpha_x, c \cdot \beta_x, \frac{\kappa_t}{c} \right\} \quad \forall c \in \mathbb{R}, c \neq 0, \text{ illetve}$$

$$\{\alpha_x, \beta_x, \kappa_t\} \rightsquigarrow \{\alpha_x - c \cdot \beta_x, \beta_x, \kappa_t + c\} \quad \forall c \in \mathbb{R}.$$

A szabadságfokok megszüntetése érdekében a szerzők az alábbi paramétermegkötéseket teszik:

$$\sum_{t=1}^T \kappa_t = 0, \quad \sum_{x=1}^X \beta_x^2 = 1.$$

A második megszorításban az eredeti cikkben (Lee és Carter, 1992) nem négyzetes tagok összege szerepel, de a továbbiakban ennek a használata megkönnyít bizonyos számításokat, az irodalomban is ez a variáció terjedt el. Az első egyenlet pedig pontosan azt írja elő, hogy az α_x értékek a korévenkénti log-mortalitások empirikus átlagának feleljenek meg. Hogy még szemléletesebb legyen a modell felépítése, tekintsük az alábbi ábrákat, melyek a Lee-Carter modell együtthatóit mutatják „sematikusan”:



1.3. ábra. A Lee-Carter modell együtthatóinak szemléltetése. (Forrás: Katrien Antonio (2016)). $\beta_x^{(1)}$ – ami tehát az én jelölésemnél α_x – az átlagos korévenkénti mortalitás, κ_t az időbeli fejlődés, míg $\beta_x^{(2)}$ ennek az időbeli fejlődésnek a korévenkénti érzékenysége.

1.2.2. Paraméterek becslése

A paraméterek becsléséhez a maximum likelihood (ML) módszer alkalmazása a legkézenfekvőbb. Belátható, hogy normális eloszlású hibatagok esetén ugyanerre az eredményre vezet a szerzők által is javasolt szinguláris érték felbontás (SVD), mely jóval kisebb számításgényű eljárásnak számít. Ma már – főként egy halandósági modellezésnél – ezek az időhatékonyságbeli különbségek elhanyagolhatóak. Én is azt tapasztaltam az elemzésemnél, hogy akár SVD-t akár ML-módszert – és ehhez Newton-Raphson iterációt – használtam a programozás során az R szoftverben, néhány másodperc alatt minden modell paraméterbecslése elkészült.

Az ML-módszerhez egyébként könnyen felírható a log-likelihood függvény – felhasználva a log-mortalitásokra felírt (1.6) normális eloszlást:

$$l = -\frac{1}{2}XT \ln(2\pi) - XT \ln \sigma - \frac{1}{2\sigma^2} \sum_{x=1}^X \sum_{t=1}^T (\ln m_{x,t} - \alpha_x - \beta_x \kappa_t)^2. \quad (1.7)$$

Ezt kell tehát maximalizálnunk az $\alpha_x, \beta_x, \kappa_t$ és σ paraméterek szerint úgy, hogy teljesüljön a két paramétermegkötés. A levezetéstől és az analitikus eredményektől eltekintek, de az említett két magyar cikkben ezek megtalálhatóak.

A becslés egyik fontos tulajdonsága, hogy minden mortalitási megfigyelést azonos súllyal vett figyelembe. Lee és Carter szerint ez ellentmond annak a gyakorlati tényezőnek, hogy az egyes kor- és naptári év csoportok létszámai és halálesi gyakoriságai nagyságrendileg eltérhetnek. Ezért a szerzők a következő egyenletrendszert mint kiigazítást javasolják:

$$\hat{\kappa}^{adj} = \{\kappa \in \mathbb{R}^T : D_t = \sum_{x=1}^X D_{x,t} = \sum_{x=1}^X E_{x,t} \exp(\hat{\alpha}_x + \hat{\beta}_x \kappa_t) \quad (t = 1, 2, \dots, T)\}. \quad (1.8)$$

Az egyenletrendszernek létezik megoldása és az egyértelmű, ha minden időszakban történt haláleset. Megoldásával tehát elérhető a numerikus alkalmazásnál, hogy az egyes időszakokban megfigyelt halálesetek száma megegyezzen a modelltől adódó várható értékeivel.

1.2.3. Előrejelzés

A fentiek alapján tehát meg tudjuk becsülni egy adott populációra illesztett Lee-Carter modell paramétereit. Ahhoz viszont, hogy jövőbeli előrejelzést készítsünk, az időtől függő paramétert, vagyis κ_t -t külön kell modelleznünk. Lee és Carter (1992) úgy találta számos idősor empirikus vizsgálata után, hogy a kiigazított mortalitási indexre leginkább az eltolásos véletlen bolyongás (angolul *Random Walk with Drift*) folyamat illeszthető, amit eltolásos ARIMA(0, 1, 0)-ként is ismerhetünk. A sztochasztikus modell tehát:

$$\hat{\kappa}_t^{adj} = \hat{\kappa}_{t-1}^{adj} + s + \Phi_t \quad (t = 2, 3, \dots, T), \quad (1.9)$$

ahol a kezdőérték ($\hat{\kappa}_1^{adj}$) ismert, s az eltolás paraméter, Φ_t pedig a hibatag. Utóbbiról feltesszük, hogy függetlenek egymástól, a Lee-Carter modell (1.4) alapegyenletének hibatagjaitól szintén, és az alábbi normális eloszlást követik:

$$\Phi_t \sim \mathcal{N}(0, \sigma_{RWD}^2) \quad (t = 2, 3, \dots, T). \quad (1.10)$$

A Lee-Carter modellben tehát a mortalitási index egy időben lineáris sztochasztikus folyamatot követ, növekményei függetlenek, a trend meredekségét pedig az s paraméter adja meg, mely a valóságban legtöbbször negatív a javuló historikus halálozási trendek miatt. Ahhoz, hogy előrejelzést készíthessünk, az s és a σ_{RWD} paraméterekre kell becslést adnunk. Az (1.9) és (1.10) egyenletek alapján felírhatjuk a kiigazított mortalitási index elsőrendű differenciáinak eloszlását:

$$\hat{\kappa}_t^{adj} - \hat{\kappa}_{t-1}^{adj} \sim \mathcal{N}(0, \sigma_{RWD}^2) \quad (t = 2, 3, \dots, T).$$

Ismert, hogy normális eloszlás várható értékének és varianciájának ML-becslése rendre a megfigyelt minta átlaga illetve varianciája:

$$\hat{s} = \frac{1}{T-1} \sum_{t=2}^T (\hat{\kappa}_t^{adj} - \hat{\kappa}_{t-1}^{adj}) = \frac{\hat{\kappa}_T^{adj} - \hat{\kappa}_1^{adj}}{T-1},$$

$$\hat{\sigma}_{RWD}^2 = \frac{1}{T-1} \sum_{t=2}^T (\hat{\kappa}_t^{adj} - \hat{\kappa}_{t-1}^{adj} - \hat{s})^2.$$

Érdekes, de nem túl meglepő tény, hogy a mortalitási index trendparaméterének becslése kizárólag az első és az utolsó megfigyeléstől függ.

Így tehát a jövőbeli mortalitási index pontbecslését az alábbiak szerint kaphatjuk meg:

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

A becsléshez s trendparaméter helyére – kihasználva a maximum likelihood becslés függvénytranszformációra vonatkozó invarianciáját – beírhatjuk \hat{s} -t. Az alapegyenletbe vissza-helyettesítve kapjuk a log-mortalitás előrejelzését a $T + j$. évre:

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

Érdekes még vizsgálni az előrejelzés hibáját, melyhez az analitikus bonyolultság miatt Monte-Carlo szimulációt szokás használni. Erről még bővebben lesz szó a következő fejezetben.

2. fejezet

Az általánosított modellkeret

2.1. A Lee-Carter modell továbbfejlődése

A Lee-Carter modell legalább az ezredfordulóig töretlen siker volt, számos tanulmány született különböző országok mortalitási modellezésére. Meglepően jól teljesített a modell – a tapasztalatokról Lee (2000) készített átfogó tanulmányt –, azonban sok elméleti vita folyt és folyik arról, hogy helyes-e az elmúlt évszázad elképesztő halandósági javulását feltételezni, jogosan számítunk-e azzal statisztikailag megegyező mértékű javulásra az előrejelzésnél. Ezzel az érdekes elméleti kérdéssel dolgozatomban most nem foglalkoznék, sokkal inkább a módszertanra helyezem a hangsúlyt.

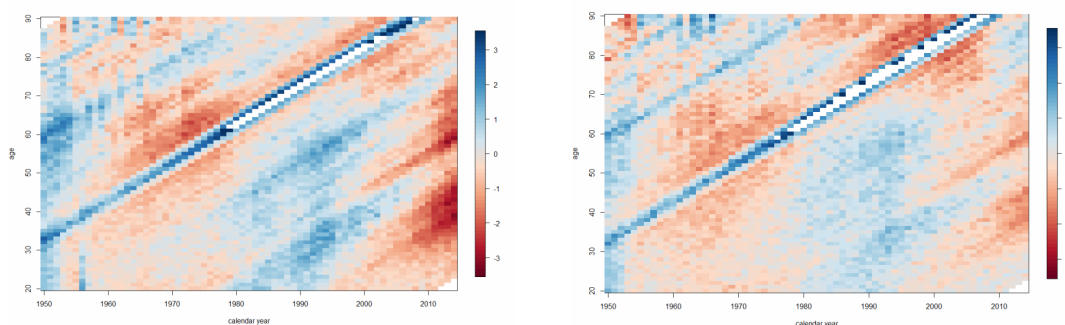
Amely tehát szintén kapott számos kritikát. Még az eredeti cikk is tartalmazott egy kiegészítést: az ugrásszerű outlier időszakokra – mint például a spanyolnátha évei – érdemes *dummy* változót bevezetni. Ennél módszertanilag sokkal erősebb kritikák is felbukkantak, melyek jelentősen hozzájárultak ahhoz, hogy mára egy egész modelles család fejlődjön ki a mortalitás modellezésére. Ebben az alfejezetben a legfontosabb ilyen modelleket fogom áttekinteni, majd a következő fejezetben ezek többpopulációs keretre való kiterjesztésével foglalkozom.

A legelső módszertani kritikai írás a cikk megjelenését követő évben jelent meg Wilmoth publikálásában (1993). Szerinte a modell egyik legtámadhatóbb pontja, hogy a log-mortalitásokra minden korév - naptári év páros esetén ugyanazt a varianciát, vagyis homoszkedaszticitást feltételez. A cikk rávilágít, hogy a variancia nagyjából fordítottan arányos a csoportokba eső halálesetek számával, ezért érdemes súlyozott legkisebb négyzetek módszerét alkalmazni a Lee és Carter által javasolt SVD helyett.

Egy Wilmoth felvetését is tartalmazó, de annál még sajátosabb modellel állt elő Brohns, Denuit és Vermunt (2002). A szerzőhármas a homoszkedaszticitás feltételének eltörlése mellett ML-módszert alkalmaz, továbbá az eredeti modell mortalitási indexre vonat-

kozó utólagos kiigazítását is elveti, lévén ez egy, a sztochasztikus modellkeretbe nehezen illeszkedő lépés. Modelljük leglényegibb eleme, hogy a normális eloszlású hibatagok helyett a mortalitási rátákhoz használt halálesetek számát tekintjük Poisson-eloszlású valószínűségi változónak. A matematikáját később részletesen ismertetem még. Nagyon hasznos ez abból a szempontból is, hogy azon korév - naptári év csoportokra, ahol nincs haláleseti megfigyelés, ezzel a lépéssel kiküszöböljük, hogy a modell bal oldalán $\ln 0$ szerepelhessen. Gyakorlati haszna ennek óriási, hiszen biztosított állományoknál tipikusan nagyon kevés, egy-egy korosztályban sokszor nulla haláleset történik, vagy érdekes például az a tény is, hogy 2014-ben Szlovákiában egyetlen 9 és 10 éves lány sem hunyt el.

A Lee-Carter modell másik legnagyobb hiányosságaként az ún. *kohorszhatást* (angolul *cohort effect*) róják fel. Ennek lényege, hogy a mortalitást nem csak a naptári év és a kor befolyásolja, de számos megfigyelés létezik rá, hogy a születési év is szignifikáns faktor. Erre az alábbi ábra nyújt kiváló példát:



2.1. ábra. Bal oldalon a férfi, jobb oldalon pedig a női magyar Lee-Carter modell reziduálisainak hő térképe láthatóak. A modellt 1950-2014 naptári éveken és 20-90 korosztályra generáltam. (Forrás: saját szerkesztés.)

Ebben a modellkeretben tehát az azonos évben születettek egy csoportot, úgynevezett kohorszt alkotnak. A 2.1. ábrán mindenki egy balról jobbra felfelé tartó diagonális mentén éli életét. A reziduálisok alapján a kohorszhatás mindkét nem esetén szembeötlő. Matematikailag ezt könnyen behozhatjuk a sztochasztikus modellbe. Ezt először Renshaw és Haberman (2006) tette meg. Az ezt követően publikált modellek lényegében mindegyike tartalmazza már a kohorszhatást is.

A Renshaw-Haberman modellt a szerzőpáros 2011-es cikkében néhány egyszerűsítéssel újraalkotta, ugyanis az eredeti modell a gyakorlatban numerikusan instabillnak bizonyult. ¹

¹Sőt, ahogy az később az elemzésemből is ki fog derülni, még így is instabil a gyakorlatban: a vizsgált hat populációból mindössze kettőben sikerült bekonvergálnia R-ben ennek a modellnek.

Az új modelljük egy speciális esete az ún. APC modell, az angol *age-period-cohort* szóhármass rövidítéséből. Az elnevezés arra utal, hogy itt a kortól nem csak a korhorszhatás, de a naptári év hatás is független, míg ugye ez eddig egyik említett modellben sem volt annak tekinthető. Az APC modell már jóval korábban használatos volt az orvostudományban (Hobcraft és szerzőtársai, 1982), mortalitási keretbe először Carstensen (2007) foglalta. Dolgozatomban ezt a modellt is fogom vizsgálni többpopulációs esetben is.

Egy másik, nagyon sikeres modellbővítési irányzat a többtényezős mortalitási modellezés. Ennek alapötlete, hogy számos mortalitási index – vagyis a korábbi egy κ_t idősor helyett kettő vagy akár több – együttesen írja le a halandóság javulását. A legismertebb ilyen modell a Cairns-Blake-Dowd (2006), illetve a Plat (2009) konstrukciók, előbbi két, míg utóbbi három tényezővel operál.

Ezek mellett még számos további irányzat létezik. Összességében elmondható, hogy az egyre kifinomultabb modellek nagyon jól teljesítenek bizonyos adathalmazokon. Kiváló kutatási terület, hogy mely ország, populáció esetén mely modell hogyan viselkedik. Az Olvasó számára többek közt Lovász (2011) cikkét ajánlanám, mely áttekintő képet ad számos európai ország esetén a numerikus tapasztalatokról. Magyarországi adatokon végzett tanulmányként a már említett Májer-Kovács (2011) cikket, ezen kívül a Baran és szerzőtársai (2007) valamint a Bajkó-Maknics-Tóth-Vékás (2015) tanulmányokat ajánlanám.

2.2. A GAPC modellcsalád

A számos modell és irányzat jól láthatóan nagyon sok matematikai vonásában közös. Így felmerült az igény néhány évvel ezelőtt, hogy egy egységes modellkeretben foglalják össze ezeket a mortalitás-előrejelző eljárásokat. Erre Hunt-Blake (2014), Currie (2016), illetve Villegas és szerzőtársai (2016) tettek kísérletet. Az általuk javasolt egységes modellkeretet *általánosított kor-időszak-kohorsz modellnek* (angolul *Generalized Age-Period-Cohort* vagy röviden GAPC modellnek) nevezzük. Ez a modellkeret egységesíti azon mortalitás-előrejelző eljárásokat, melyek a kor - naptári év csoportosításban log-bilineáris vagy logit-bilineáris, egy- vagy többtényezős, valamint korhorszhatást figyelembe vevő illetve ignoráló modellek. A következő alfejezetekben a fent már röviden bemutatott modelleket fejtem ki részleteiben, és mutatom meg, hogy hogyan illeszthetőek bele a GAPC családba. Fontos megjegyezni, hogy a fenti modellek közül mind beleillik ezen modellkeretbe, kivéve az eredeti Lee-Carter modellt (1992), így ennek csak a Poisson-változatával fogok foglalkozni.

Maradva a korábbi jelöléseknél, legyen $m_{x,t}^0$ illetve $m_{x,t}^c$ az $x \in \{1, 2, \dots, X\}$ korcsoporthoz és $t \in \{1, 2, \dots, T\}$ naptári évhez tartozó kezdeti illetve központi mortalitási ráta. Megfigyelésink a $D_{x,t} \in \mathbb{N}$ halálesetek száma, illetve az $E_{x,t}^0 \in \mathbb{N}^+$ kezdeti vagy $E_{x,t}^c > 0$

központi kitétségek. Követve Villegas és szerzőtársainak (2016) konstrukcióját, a GAPC modellcsalád négy komponensből épül fel:

Sztochasztikus komponens

Mint már említettem, az eredeti Lee-Carter modell továbbfejlesztésének egyik sarokköve, hogy a $D_{x,t}$ mennyiséget érdemesebb valószínűségi változónak tekinteni, amire vezessük be a megfigyelések számától eltérő $\bar{D}_{x,t}$ jelölést. A halálesetek számát minden naptári évre és korcsoportra egymástól független, binomiális vagy Poisson-eloszlásúnak feltételezzük:

$$\bar{D}_{x,t} \sim Bin(E_{x,t}^0, m_{x,t}^0) \quad (2.1)$$

vagy

$$\bar{D}_{x,t} \sim Poisson(E_{x,t}^c \cdot m_{x,t}^c). \quad (2.2)$$

Linkfüggvény

A kezdeti vagy központi halandósági rátákat szeretnénk becsülni. Ahogy Lee és Carter esetén láttuk, ezt egyszerűbb megtenni, ha annak valamilyen függvényét tekintjük. Legyen $g : \mathbb{R}^+ \rightarrow \mathbb{R}$ ez a függvény, melyet szokás *linkfüggvénynek* nevezni. Hunt-Blake (2014) központi kitétségek és Poisson-eloszlás használata esetén a $g(x) = \ln(x)$ logaritmikus függvényt javasolják, míg kezdeti kitétség és binomiális eloszlás mellett a $g(x) = \ln\left(\frac{x}{1-x}\right)$ logit függvényt.

Szisztematikus komponens

A megfelelő linkfüggvény argumentumába beírva a megfelelő mortalitási rátát kapjuk az ún. szisztematikus komponenszt:

$$\eta_{x,t} = g(m_{x,t}) \quad (x = 1, 2, \dots, X, \quad t = 1, 2, \dots, T). \quad (2.3)$$

A mortalitási modelleink tehát ezt a változót szeretnék előrejelezni. A GAPC modellkeret így az alábbi univerzális egyenlettel írható fel:

$$\eta_{x,t} = \alpha_x + \sum_{i=1}^N \beta_x^{(i)} \kappa_t^{(i)} + \beta_x^{(0)} \gamma_{t-x} \quad (x = 1, 2, \dots, X, \quad t = 1, 2, \dots, T), \quad (2.4)$$

ahol:

- ahogy már megismertük, α_x a kor szerinti hatás paramétere.

- $N \in \mathbb{N}$ a kor - naptári év tényezők száma. A $\kappa_t^{(i)}$ mortalitási indexek testesítik meg a mortalitási szint időbeli sztochasztikus alakulását, míg $\beta_x^{(i)}$ együttthatók a halandósági ráta érzékenységét adják meg kor szerint a mortalitási indexekre nézve.
- Végül pedig a γ_{t-x} paraméterek veszik fel a kohorszhatást, vagyis hogy a tipikus halandósági pályához képest az adott kohorsz milyen mértékben tér el.

Megszorítások

Ahogy az eredeti modellnél, az általánosított esetben is csak transzformáció erejéig egyértelmű a paraméterbecslések megoldása, ezért bizonyos megszorításokat kell tennünk a paraméterekre. Általánosan ezt úgy fogalmazhatjuk meg, hogy a

$$\Theta := (\alpha_x, \beta_x^{(1)}, \dots, \beta_x^{(N)}, \kappa_t^{(1)}, \dots, \kappa_t^{(N)}, \beta_x^{(0)}, \gamma_{t-x}) \quad (2.5)$$

paramétervektorra fennáll, hogy egy bizonyos θ transzformációja kielégíti a

$$\theta(\Theta) = \bar{\Theta} = (\bar{\alpha}_x, \bar{\beta}_x^{(1)}, \dots, \bar{\beta}_x^{(N)}, \bar{\kappa}_t^{(1)}, \dots, \bar{\kappa}_t^{(N)}, \bar{\beta}_x^{(0)}, \bar{\gamma}_{t-x})$$

egyenletet, és így a (2.5) egyenlet által definiált Θ paramétervektor $\theta(\Theta)$ transzformáltja szintén a (2.4) egyenletbeli szisztematikus komponens eredményezi.

2.2.1. Poisson Lee-Carter modell

A Brouhns és szerzőtársai (2002) által megalkotott modell lényegében az eredeti Lee-Carter modell (1992) hibatag nélküli változata, ahol a GAPC modellkerethez hűen a bal oldalon valószínűségi változót, méghozzá a Poisson-eloszlás miatt $\eta_{x,t} = \ln m_{x,t}^c$ változót tételezünk fel:

$$\eta_{x,t} = \alpha_x + \beta_x^{(1)} \kappa_t^{(1)} \quad (x = 1, 2, \dots, X, \quad t = 1, 2, \dots, T).$$

A szerzők javaslata alapján a következő megszorításokkal élünk:

$$\sum_{x=1}^X \beta_x^{(1)} = 1, \quad \sum_{t=1}^T \kappa_t^{(1)} = 0.$$

A mortalitás előrejelzéséhez továbbra is ARIMA folyamatokat használunk, ennél a modellnél is empirikusan alátámasztottnak bizonyult az RWD folyamat, azaz:

$$\kappa_t^{(1)} = \delta + \kappa_{t-1}^{(1)} + \Phi_t, \quad \Phi_t \sim \mathcal{N}(0, \sigma_{RWD}^2).$$

2.2.2. Renshaw-Haberman modell

A következő tárgyalt struktúra a Renshaw-Haberman (R-H) modell (2006) volt, mely már figyelembe vette a kohorszhatást. Ez tehát a Poisson Lee-Carter modell alábbi módon bővített változata:

$$\eta_{x,t} = \alpha_x + \beta_x^{(1)} \kappa_t^{(1)} + \beta_x^{(0)} \gamma_{t-x} \quad (x = 1, 2, \dots, X, \quad t = 1, 2, \dots, T).$$

Az identifikációs probléma megoldására javasolt paraméter-megkötések:

$$\begin{aligned} \sum_{x=1}^X \beta_x^{(1)} &= 1, & \sum_{t=1}^T \kappa_t^{(1)} &= 0, \\ \sum_{x=1}^X \beta_x^{(0)} &= 1, & \sum_{c=1-X}^{T-1} \gamma_c &= 0. \end{aligned}$$

Mivel numerikusan instabilnak találták a fenti modellt, ezért 2011-es cikkükben a $\beta_x^{(0)} = 1$ egyszerűsítést javasolták. A korábbiakhoz képest most már két idősorunk van: $\kappa_t^{(1)}$ és γ_{t-x} . Mindkettő modellezése a már bemutatott ARIMA folyamatokkal történik, fontos, hogy e két idősor független folyamat.

2.2.3. APC modell

Az Age-Cohort-Period (APC) modell – ahogy azt korábban említettem – a kor, kohorsz és naptári év dimenziók mentén függetlennek tekinti a mortalitást. Ez lényegében az R-H modell egy speciális esete $\beta_x^{(0)} = \beta_x^{(1)} = 1$ feltételek mellett, melyet először Carstensen (2007) publikált:

$$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + \gamma_{t-x} \quad (x = 1, 2, \dots, X, \quad t = 1, 2, \dots, T).$$

Az alábbi megszorításokkal elérhető az identifikációs probléma feloldása:

$$\sum_{t=1}^T \kappa_t^{(1)} = 0, \quad \sum_{c=1-X}^{T-1} \gamma_c = 0, \quad \sum_{c=1-X}^{T-1} c\gamma_c = 0.$$

2.2.4. Cairns-Blake-Dowd modell

A Cairns-Blake-Dowd (CBD) modell (2006) egy olyan kéttényezős struktúra, mely kezdeti kitettséget használ logit linkfüggvénnyel, és nem tartalmaz sem additív életkorhatást, sem kohorszhatást:

$$\eta_{x,t} = \kappa_t^{(1)} + (x - \bar{x}) \kappa_t^{(2)} \quad (x = 1, 2, \dots, X, \quad t = 1, 2, \dots, T),$$

ahol $\bar{x} = \frac{X+1}{2}$ az előforduló korcsoportok számtani átlaga. A szerzők ezt a modellt főként magasabb, $x_0 = 60$ éves kor feletti populációra találták megfelelőnek, ahol az $\eta_{x,t}$ szisztematikus faktor $x_0 + x$ korcsoportra vonatkozik. Itt nem lép fel identifikációs probléma, ezért a megszorítások halmaza üres.

2.2.5. M7 modell

Későbbi cikkükben (Cairns, 2009) alapmodelljükből kiindulva egy még szofisztikáltabb struktúrát fejlesztettek, mely már kohorszhatást is tartalmaz. Ez pedig a *kvadratikus CBD modell kohorszhatással bővített változata*, melyet a szakirodalomban csak *7-es modellként* vagy még rövidebben *M7-ként* szoktak említeni. A modell struktúrája a következő:

$$\eta_{x,t} = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + \kappa_t^{(3)}((x - \bar{x})^2 + \hat{\sigma}_x^2) + \gamma_{t-x} \quad (x = 1, 2, \dots, X, \quad t = 1, 2, \dots, T),$$

ahol $\hat{\sigma}_x^2$ a korcsoportok indexeinek kvadratikus átlagtól való eltéréseinek átlaga, vagyis az $(x - \bar{x})^2$ értékek átlaga. A szerzők az alábbi megszorításokat javasolják:

$$\sum_{c=1-X}^{T-1} \gamma_c = 0, \quad \sum_{c=1-X}^{T-1} c\gamma_c = 0, \quad \sum_{c=1-X}^{T-1} c^2\gamma_c = 0.$$

Ez a feltételhármas azért hasznos, mert biztosítja, hogy a kohorszhatás 0 körül fluktuáljon, és ne legyen lineáris illetve kvadratikus trendje. A modellnek létezik egy speciális esete (*M6*) és egy másik változata (*M8*) is a gyakorlatban, de ezekkel a dolgozatban nem foglalkozom.

2.2.6. Plat modell

Végül a ma használatos standard halandósági modellek közül Plat (2009) konstrukcióját ismertetném. Alapjaiban a CBD-modellből indul ki, de amellett, hogy már háromtényezős, a Poisson Lee-Carter modellből is vesz át elemeket:

$$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + \kappa_t^{(3)}(\bar{x} - x)^+ + \gamma_{t-x} \quad (x = 1, 2, \dots, X, \quad t = 1, 2, \dots, T),$$

ahol $(\bar{x} - x)^+$ a szokásos jelölése $\max(0, \bar{x} - x)$ -nek, és az alábbi paramétermegszorítás javasolt:

$$\begin{aligned} \sum_{t=1}^T \kappa_t^{(1)} = 0, \quad \sum_{t=1}^T \kappa_t^{(2)} = 0, \quad \sum_{t=1}^T \kappa_t^{(3)} = 0, \\ \sum_{c=1-X}^{T-1} \gamma_c = 0, \quad \sum_{c=1-X}^{T-1} c\gamma_c = 0, \quad \sum_{c=1-X}^{T-1} c^2\gamma_c = 0. \end{aligned}$$

A szerző magasabb életkorokra $\kappa_t^{(3)}$ paraméter elhagyását – mivel ekkor $x > \bar{x}$ miatt $(\bar{x} - x)^+ = 0$ – és a CBD modellnél látott logika használatát javasolja.

2.3. Becslés és előrejelzés

Az előző alfejezetben tehát összesen 6 egy populációs mortalitási modellt mutattam be. Ezek mindegyike a nemrégiben publikált GAPC modelles család tagjai. Ezt a tényt mindegyik konstrukcióra könnyű ellenőrizni, további részletekért lásd Hunt-Blake (2014) vagy Currie (2016) írását.

Ezek a modellek igen sikeresnek mondhatók a gyakorlatban. Több olyan irány létezik viszont, melyek segítségével ezek az előrejelzések még pontosabbá, megbízhatóbbá tehetők. Az egyik legérdekesebb és legújabb ilyen irányzat a többpopulációs modellezés. A következő fejezetben a fentiekben bemutatott 6 modellt fogom kiterjeszteni két populációra, majd végzek numerikus számításokat. Előtte viszont még az eredeti Lee-Carterhez hasonlóan itt is bemutatom a paraméterbecslés és az előrejelzés mikéntjét.

2.3.1. Paraméterbecslés és modellválasztás

A (2.4) általános modellfelírás paramétereit a (2.5)-ben definiált Θ vektorral fejeztük ki. A paraméterek becslése maximum likelihood módszerrel végezhető el, ahol a log-likelihood függvény az alábbi formában írható fel Poisson-eloszlású haláleset számokat feltételezve:

$$l(\Theta) = \sum_{x=1}^X \sum_{t=1}^T \omega_{x,t} \left(-E_{x,t}^c m_{x,t}^c + D_{x,t} (\ln E_{x,t}^c + \ln m_{x,t}^c) - \ln(D_{x,t}!) \right), \quad (2.6)$$

míg binomiális eloszlás esetén:

$$l(\Theta) = \sum_{x=1}^X \sum_{t=1}^T \omega_{x,t} \left(\ln \binom{E_{x,t}^0}{D_{x,t}} + D_{x,t} \ln m_{x,t}^0 + (E_{x,t}^0 - D_{x,t}) \ln(1 - m_{x,t}^0) \right), \quad (2.7)$$

ahol $\omega_{x,t}$ mindkét esetben az az indikátorfüggvény, mely rögzített (x, t) esetén 1-et vesz fel, ha van megfigyelésünk az x korcsoportra és t időszakra, illetve 0-t, ha nincs. A választott modellspecifikációban tehát a log-likelihood maximalizálásával található meg az optimális paraméterezés, figyelembe véve a modellre korábban tett identifikációs megszorításokat. Az optimalizálási feladat megoldására számos numerikus módszer létezik, én munkám során az R szoftver `StMoMo` csomagját használtam, melyet részletesen mutat be Villegas és szerzőtársainak (2016) cikke.

2.3.2. Előrejelzés

A modellezés utolsó lépése – miután már megbecsültük a paramétereket – az előrejelzés. Az egy populációs modellek mortalitási indexét tipikusan a Lee-Carter modellnél már bemutatott eltolásos véletlen bolyongással (RWD) szokás modellezni. A GAPC jelentette általános, többdimenziós esetre többek közt Villegas és szerzőtársai (2016) specifikálja ezt

a modellstruktúrát. A szóban forgó mortalitási indexeket, melyekből tehát összesen N darab van, foglaljuk egy vektorba:

$$\hat{\kappa}_t = \begin{bmatrix} \hat{\kappa}_t^{(1)} \\ \hat{\kappa}_t^{(2)} \\ \vdots \\ \hat{\kappa}_t^{(N)} \end{bmatrix}$$

Így a többdimenziós RWD felírása az alábbi egyenlettel történik:

$$\hat{\kappa}_t = \hat{\kappa}_{t-1} + \mathbf{s} + \mathbf{\Phi}_t \quad (t = 2, 3, \dots, T),$$

ahol $\hat{\kappa}_1$ a kezdőérték-vektor a mortalitási indexre, \mathbf{s} az eltolás mértékét kifejező trendparaméter, $\mathbf{\Phi}_t$ pedig az egymástól és a halálozási gyakoriságoktól független hibatagok, melyek $\mathbf{0}$ várható értékű, Σ pozitív definit kovarianciamátrixú többdimenziós normális eloszlást követnek:

$$\mathbf{\Phi}_t \sim \mathcal{N}(\mathbf{0}, \Sigma) \quad (t = 2, 3, \dots, T).$$

A (2.4) általános GAPC felírásban nemcsak a mortalitási indexek, hanem a kohorsz indexek is mint idősorok szerepelnek, így γ_{t-x} változókra is fel kell írunk valamilyen sztochasztikus modellt. Az irodalomban ez esetben is a trendparaméteres ARIMA(p, q, d)-folyamatokat szokás használni, melynek általános alakja a kohorsz indexre felírva:

$$\Delta^d \gamma_c = \delta_0 + \Phi_1 \Delta^d \gamma_{c-1} + \dots + \Phi_p \Delta^d \gamma_{c-p} + \epsilon_c + \delta_1 \epsilon_{c-1} + \dots + \delta_q \epsilon_{c-q},$$

ahol Δ^d a d -edik differenciaoperátor, δ_0 a trendparaméter, Φ_1, \dots, Φ_p az autoregresszív együtthatók ($\Phi_p \neq 0$), $\delta_1, \dots, \delta_q$ a mozgóátlag koefficiensek ($\delta_q \neq 0$), végül pedig ϵ_i -k egymástól független, 0 várható értékű, σ^2 varianciájú normális eloszlást követő hibatagok. A Cairns (2011), Lovász (2011) és Villegas és szerzőtársai (2016) cikkeket követve az alábbi paraméterezéseket ajánlatos alkalmazni a kohorszhatást tartalmazó modellek esetén:

2.1. táblázat: A kohorsz index modellparaméterezése

Mortalitás modell	Kohorsz index paraméterezése
APC	ARIMA(1, 1, 0)
RH	ARIMA(1, 1, 0)
M7	ARIMA(2, 0, 0)
Plat	ARIMA(2, 0, 0)

A modellezés után ismét felírhatóak a differenciákra a megfelelő normális eloszlások, és az azokból következő log-likelihood függvények. Ezekről most a terjedelmük miatt eltekintek, de a levezetések a korábban említett cikkekben megtalálhatóak. Ha a becslések megvannak, akkor már egyszerűen adódnak a j időszakkal későbbi pontbecslések RWD segítségével a mortalitási indexre:

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

illetve például ARIMA(1, 1, 0) esetén a kohorsz indexre:

$$\mathbb{E}(\gamma_{T-1+j}) = \hat{\delta}_0 + \hat{\Phi}_1 \mathbb{E}(\gamma_{T-2+j}) + \hat{\Phi}_2 \mathbb{E}(\gamma_{T-3+j}) \quad (j = 1, 2, \dots).$$

Ezeket a becsléseket pedig már behelyettesíthetjük a szisztematikus komponens egyenletébe, és megkapjuk a pontbecslését:

$$\begin{aligned}\hat{\eta}_{x,T+j} &= \hat{\alpha}_x + \sum_{i=1}^N \beta_x^{(i)} \mathbb{E}(\hat{\kappa}_{T+j}^{(i)}) + \hat{\beta}_x^{(0)} \mathbb{E}(\hat{\gamma}_{T+j-x}) \\ &(x = 1, 2, \dots, X, \quad j = 1, 2, \dots).\end{aligned}$$

A pontbecslés nevében is benne van, hogy ez csak a „legjobb” vélt átlagos értéket fogja figyelembe venni, és azzal számol előrejelzést, a mögöttes sztochasztikus változók bizonytalanságát nem veszi figyelembe. Mivel analitikailag ezek elég bonyolult képletek, már az első Lee-Carter modellfejlesztések óta a Monte Carlo szimulációk alkalmazása a bevett gyakorlat. Ez egy viszonylag egyszerű módszer a modellezésre: a becslések után a hibtagokat szimuláljuk, és ezekkel jelzünk előre számos scenáriót. A GAPC modellcsalád minden tagjára használható ez a népszerű módszer, a használt R-csomag szintén tartalmazza ennek numerikus megvalósítását.

Fontos viszont megjegyezni, hogy eddig nem vettük figyelembe a paraméterbecslésben rejlő bizonytalanságot. Erre először Brouhns és társai (2005) mutattak rá. Ők a *szemiparaméteres bootstrap eljárást* javasolják a probléma megoldására, melyet az *StMoMo* R-csomag is használ. Lényege, hogy a megfigyelt $D_{x,t}$ halálzási értékeket a paraméterbecslések után újra kell generálni a megfelelő feltételezett eloszlás szerint (Poisson vagy binomiális), majd az így kapott összes mintára az eredeti modellválasztást megtartva kell elvégezni újra a becsléseket. Elég sok mintagenerálással a bootstrap már jól fogja közelíteni a mortalitás vagy más vizsgált mennyiség elméleti eloszlását.

3. fejezet

Többpopulációs módszerek

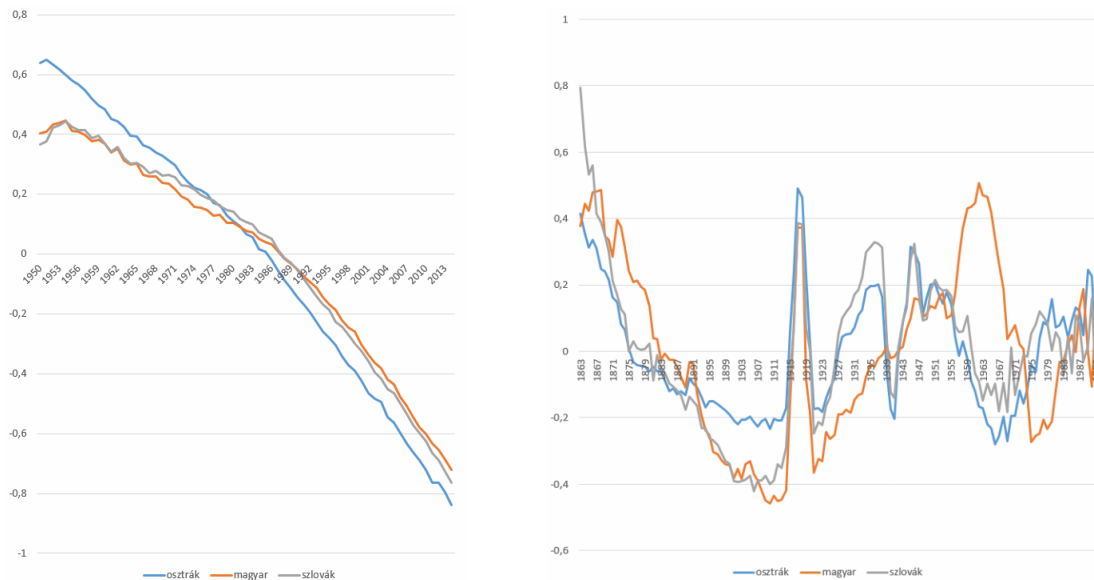
3.1. Motiváció

A korábbiakban tehát áttekintést adtam a mai mortalitási modellek legnépszerűbbjeiről, a GAPC modelles család hat kiemelt tagjáról. Ezek – bár igen sikeresnek számítanak – további statisztikai eszközökkel tovább fejleszthetőek, hogy még pontosabb halandósági előrejelzés legyen lehetséges. Az utóbbi évek egyik meghatározó irányzata a többpopulációs modellezés. Új változók bevonása a standard egypopulációs modellekbe, vagy bármilyen más, matematikailag kisebb volumenű változtatás nem igazán hozott jobb eredményeket az előrejelzésekben. Viszont két nagyon fontos megfigyelést érdemesnek találhatunk beépíteni a modellbe. Az egyik, hogy a biztosított állomány része az országos populációnak; és ugyan ismert, hogy ezek hajlamosak merőben eltérő halandóságot mutatni, fejlődéseik erősen összefüggnek. Az ebből a különbségből fakadó kockázatot *population basis risk* néven ismeri a szakma, amit magyarul *bázispopulációs kockázatnak* fordíthatunk. Dolgozatomban ezzel a populáció-szubpopuláció iránnyal nem fogok foglalkozni, de szólok róla bizonyos modelleknél, illetve a jövőben hazai modellezéseknél adott biztosított állományra ezeket a módszereket kimondottan ajánlatosnak tartom elemzésre. A bázispopulációs kockázat modellezése lényegében csak többpopulációs modellezési technikával lehetséges.

A másik fontos megfigyelés pedig, hogy az egymással szoros kapcsolatban élő populációk halandóságai összefüggőséget mutatnak. A 1.2. ábrán már láthattuk, hogy a magyar női és férfi halandóság mennyire hasonlóan fejlődött az elmúlt bő 60 évben. Elemzésem másik iránya pedig a szomszédos országok halandóságait hivatott nemenként együtt modellezni. A historikus együttmozgásra láthatunk példát a 3.1. ábrán.

Láthatjuk, hogy nem tökéletes az együttmozgás, de annyira szoros összefüggőség ötlük szembe, ami már érzésre is rengeteg plusz információt tehet hozzá a modellhez, amennyiben sikerül egy érvényes együttes struktúrát találni. Az ábrát elkészíthettem volna a férfiakra

is, vagy nemzeteken belül férfi-női párosokra, hasonló képet látnánk kibontakozni. Bizonyos, hogy az Európai Unión belül legtöbb esetben, főleg a szomszédos országoknál, nagyon hasonló jelenségre bukkanhatunk. Különösen hasznos lehet így a többpopulációs modellezési technika a multinacionális európai életbiztosítók esetén, köztük számos hazai vállalatnál is.



3.1. ábra. A bal oldali ábrán a magyar, osztrák és szlovák női APC-modellből adódó κ_t , vagyis mortalitási indexek láthatóak; míg jobb oldalon ugyanezen egypopulációs modellekből adódó γ_{t-x} kohorsz index idősorok. (Forrás: saját szerkesztés 1950-2014 adatokon futtatva a modellt a 20-90 éves korosztályokra.)

Érdekes lehet megvizsgálni még, hogy ezen egymásra ható populációkat ne csak országok mentén vagy nemzeteken belül nemek szerint különítsük el, hanem például dohányzó - nem dohányzó szubpopulációkra vagy bizonyos gazdasági-szociális helyzet szerint felszavozva a társadalmat. Ezekkel az irányzatokkal dolgozatomban nem foglalkozom, és sajnos a magyar adatbázisok szűkössége miatt a jövőben is nehézkesnek tartanám. Érdekes tanulmány ezzel kapcsolatban viszont Cairns és szerzőtársainak cikke (2016), melyben Dánia népességét tíz szociális-gazdasági csoportra osztják fel, és modellezik azok mortalitását többpopulációs módszerrel.

3.2. A standard modellek kiterjesztése két populációra

Már az újabb mortalitási modellek kidolgozását megelőzően megszületett az igény a többpopulációs modellezésre. Li és Lee 2005-ös cikkükben az eredeti Lee-Carter modellt terjesztették ki kétpopulációs esetre. Ötletük azon alapszik, hogy hosszú távon – ahogy ezt a már említett Wilson cikk (2001) is bemutatja – a mortalitási szintek konvergálnak globálisan. A későbbi modelleknél is ez az elsődleges feltételezés: a két vagy több vizsgált populáció mortalitásainak hosszú távú együttmozgása figyelhető meg. Erre a kapcsolatra kézenfekvő kointegrációs elemzéseket végezni, és teszi is ezt a többpopulációs modellek legtöbbször.

Fontos kérdés, hogy a vizsgált populációkat ezen összefüggőség esetén szimmetrikusnak feltételezzük, vagy ellenkezőleg: az egyik dominálja a másik halandóságát, vagyis meghatározza a hosszú távú trendet. Utóbbi irányzatra alkotta Dowd és szerzőtársai (2011) az ún. *gravitációs modellt*. Lényege, hogy a domináns populációt függetlenül modellezzük, majd ezt követően a másik populáció historikus eltéréseit ezektől a domináns mortalitásoktól vizsgáljuk, és a populációk egymáshoz képesti méretei gravitációs paraméterként foghatóak fel. A modell már kohorszhatást is tartalmazott, egészen pontosan az APC modellt terjesztették ki ezen elv alapján két populációra, eredményeiket pedig az angol és wales-i férfi populáción mutatták be.

A gravitációs elvet alkalmazza Jarner és Kryger (2011). Ők még nagyobb hangsúlyt fektetnek a létszámban arányaiban messze eltérő, és az idősebb korévekre fókuszáló modellezésre. Ez egy kiváló módszertan lehet biztosítóknál, hiszen az állomány méreteiben általában jóval kisebb, mint az országos populáció; emellett nyugdíjtermékeknel a 65 feletti halandóságokra kell tekintettel lenni. Dolgozatomban mégsem ezt az irányt mutatom be, ugyanis magyar adatokra épülő, bevezető többpopulációs modellezésként hasznosabbnak, érdekesebbnek tűnt országos halandósági elemzést végezni.

Térjünk ezért át a másik irány modelljeire, ahol közel szimmetrikus feltevésekkel élünk. Ez azt jelenti, hogy a hosszú távú kapcsolat vizsgálata során a sztochasztikus faktorokat populációpáronként egymással összefüggő folyamatoknak tekintjük; és ezeket az idősorokat együttesen, szimmetrikus feltételekkel élve modellezzük. Hogy erre milyen módszerek léteznek, arról a 3.2.2. szakaszban lesz szó.

Ez az ötlet és a modell kidolgozása a Yang-Wang (2012) illetve a Zhou és szerzőtársai (2012) cikkben jelent meg. A sztochasztikus folyamatok összefüggő modellezésének bonyolultsága miatt azonban csak az eredeti Lee-Carter modellre, ami, mint láttuk, már az egy populációs modellezéshez sem mondható igazán erősnek. Az irányzat egyik legújabb tanulmánya, Li, Zhou és Hardy 2015-ös írásában viszont már mindegyik korábban említett standard modellre kidolgozta a kétpopulációs eset modellfelírásait. Ezt fogom bemutatni.

ni a következő két alfejezetben, melynek kidolgozása során erősen támaszkodtam a fent említett három cikk mindegyikére. Az n populációra való általánosítás egyelőre itt még nincs levezetve, és elemzésem sem tartalmaz kettőnél magasabb dimenziójú modelleket. Jövőbeli nemzetközi analízisekhez viszont szükségessé válhat még az általánosítás.

3.2.1. Modelfelírások és paraméterbecslés

A több populációra való kiterjesztés első lépéseként írjuk fel minden populációra a standard modellt. Az egységes jelöléshez felső indexbe fogom írni a populáció sorszámát. Így tehát a következő 6 konstrukciót tekintjük:

3.1. táblázat: Multipopulációs modellek

Mortalitás modell	Struktúra
Poisson Lee-Carter	$\eta_{x,t}^{(j)} = \alpha_x^{(j)} + \beta_x^{(1,j)} \kappa_t^{(1,j)}$
Renshaw-Haberman	$\eta_{x,t}^{(j)} = \alpha_x^{(j)} + \beta_x^{(1,j)} \kappa_t^{(1,j)} + \gamma_{t-x}^{(j)}$
APC	$\eta_{x,t}^{(j)} = \alpha_x^{(j)} + \kappa_t^{(1,j)} + \gamma_{t-x}^{(j)}$
CBD	$\eta_{x,t}^{(j)} = \kappa_t^{(1,j)} + (x - \bar{x}) \kappa_t^{(2,j)}$
M7	$\eta_{x,t}^{(j)} = \kappa_t^{(1,j)} + (x - \bar{x}) \kappa_t^{(2,j)} + \kappa_t^{(3,j)} ((x - \bar{x})^2 + \hat{\sigma}_x^2) + \gamma_{t-x}^{(j)}$
Plat	$\eta_{x,t}^{(j)} = \alpha_x^{(j)} + \kappa_t^{(1,j)} + (x - \bar{x}) \kappa_t^{(2,j)} + \gamma_{t-x}^{(j)}$

Elevenítsük fel a korábbi jelöléseket, immár a j -edik populációra értelmezve, és továbbra is az $x = 1, 2, \dots, X$ illetve $t = 1, 2, \dots, T$ futóindexeket használva:

- $m_{x,t}^{(j)}$: a j -edik populáció központi halandósági rátája az x éves korcsoport és a t . naptári év esetén.
- $q_{x,t}^{(j)}$: a j -edik populáció x éves egyénére a t . időszakban vonatkozó egyéves halálozási valószínűség.
- $\eta_{x,t}^{(j)}$: a linkfüggvény milyenségéből adódó logaritmikus vagy logit mortalitási ráta. Poisson Lee-Carter, R-H és APC modellek esetén $\ln m_{x,t}^{(j)}$ használata a javasolt az alkotók ajánlása alapján, míg a többi modellnél $\ln \left(\frac{q_{x,t}^{(j)}}{1 - q_{x,t}^{(j)}} \right)$, de ettől eltérő alkalmazás is megengedett.
- $\alpha_{x,t}^{(j)}$: a kor szerinti hatás paramétere.
- $\kappa_{x,t}^{(i,j)}$ ($i = 1, 2, 3$): mortalitási indexek.
- $\beta_{x,t}^{(i,j)}$ ($i = 1$): a halandósági ráta érzékenysége az i -edik mortalitási indexre.

- $\gamma_{t-x}^{(j)}$: a kohorsz index.

Ahogy azt az egypopulációs modellfelírásoknál ismertettem, minden struktúránál bizonyos megszorításokkal kell élnünk, hogy az identifikációs problémát feloldjuk. Ezek természetesen továbbra is fennállnak, a 2.2. alfejezethez hasonlóan felírható mindegyik. Ezeketől, a korábbiaktól csak dimenziószámában eltérő képletektől megkímélném az Olvasót, a részletek a Li-Zhou-Hardy cikkben (2015) megtalálhatóak.

Ami viszont elengedhetetlen még a többpopulációs modellezésnél, az a hosszú távú konvergencia vizsgálata, és annak teljesülése. Az úttörő Li-Lee (2005) cikk ezt *koherenciának* nevezi, matematikai formában pedig Cairns és szerzőtársai (2011) fogalmazzák meg:

3.2.1. Definíció. *A J halmazba tartozó populációk mortalitásai pontosan akkor tekinthetők koherensnek, ha bármely $j_1, j_2 \in J$ populációpár esetén, minden x életkorra fennáll, hogy $t \rightarrow \infty$ esetén az $m_{x,t}^{j_1}/m_{x,t}^{j_2}$ arányszám nem divergál t -ben.*

Belátható, hogy a korábban vázolt többpopulációs modellek esetén az alábbi feltételek kielégítőek a koherenciára nézve:

3.2. táblázat: Koherenciát kielégítő feltételrendszerek

Mortalitás modell	Feltételrendszer
Poisson Lee-Carter	$\beta_x^{(1,j_1)} = \beta_x^{(1,j_2)}$ és $\kappa_t^{(1,j_1)} - \kappa_t^{(1,j_2)}$ mean-reverting
Renshaw-Haberman	$\beta_x^{(1,j_1)} = \beta_x^{(1,j_2)}$, illetve $\kappa_t^{(1,j_1)} - \kappa_t^{(1,j_2)}$ és $\gamma_{t-x}^{(j_1)} - \gamma_{t-x}^{(j_2)}$ m-r.
APC	$\kappa_t^{(1,j_1)} - \kappa_t^{(1,j_2)}$ és $\gamma_{t-x}^{(j_1)} - \gamma_{t-x}^{(j_2)}$ m-r.
CBD	$\kappa_t^{(i,j_1)} - \kappa_t^{(1,j_2)}$, $i = 1, 2$ m-r.
M7	$\kappa_t^{(i,j_1)} - \kappa_t^{(1,j_2)}$, $i = 1, 2, 3$ m-r.
Plat	$\kappa_t^{(1,j_1)} - \kappa_t^{(1,j_2)}$, $i = 1, 2$ és $\gamma_{t-x}^{(j_1)} - \gamma_{t-x}^{(j_2)}$ m-r.

Fontos fogalom a *mean-reverting* vagy magyarul talán átlaghoz való visszahúzásként leírható tulajdonság. Ezt olyan idősorokra szokás használni, melyek tendenciózusan visszatérnek a historikus átlagukhoz, vagyis – szemben a véletlen bolyongással – ennek van „emlékezete”.

A 3.2. táblázatban leírt feltételek egy logikus rendszert alkotnak: a hosszútávú mortalitási konvergenciához (koherenciához) elégséges feltétel, ha a korfüggő érzékenységi együtthatók megegyeznek, illetve ha a páronként vett sztochasztikus indexkülönbségek időSORA átlaghoz visszahúzó folyamatok. Ezek tesztelését konkrét példán fogom bemutatni a kö-

vetkező fejezetben.

Az itt ismertett koncepció egyik erőssége az az egyszerűség, hogy a paraméterbecslésig még kvázi függetlennek tekintjük a két populáció modelljét. Így a log-likelihoodok lényegében összeadódnak, ahogyan azt a (2.6) és (2.7) képletekből kiindulva fel is írhatjuk két populációra:

Poisson-eloszlású haláleset számokat feltételezve:

$$l(\Theta) = \sum_{j=1}^2 \sum_{x=1}^X \sum_{t=1}^T \omega_{x,t}^{(j)} \left(-E_{x,t}^{(c,j)} m_{x,t}^{(c,j)} + D_{x,t}^{(j)} (\ln E_{x,t}^{(c,j)} + \ln m_{x,t}^{(c,j)}) - \ln(D_{x,t}^{(j)}!) \right), \quad (3.1)$$

míg binomiális eloszlás esetén:

$$l(\Theta) = \sum_{j=1}^2 \sum_{x=1}^X \sum_{t=1}^T \omega_{x,t}^{(j)} \left(\ln \binom{E_{x,t}^{(0,j)}}{D_{x,t}^{(j)}} + D_{x,t}^{(j)} \ln m_{x,t}^{(0,j)} + (E_{x,t}^{(0,j)} - D_{x,t}^{(j)}) \ln(1 - m_{x,t}^{(0,j)}) \right), \quad (3.2)$$

ahol $\omega_{x,t}$ ismételen mindkét esetben az az indikátorfüggvény, mely rögzített (x, t) esetén 1-et vesz fel, ha van megfigyelésünk az x korcsoportra és t időszakra, illetve 0-t, ha nincs. A választott modellspecifikációban tehát a log-likelihood maximalizálásával található meg az optimális paraméterezés, figyelembe véve a modellre korábban tett összes megszorítást.

3.2.2. Sztochasztikus faktorok

A Li-Zhou-Hardy (2015) cikk itt egy további lépést javasol az elemzéshez: a parszimónia elvét követve tegyünk egyszerűsítő feltevéseket a sztochasztikus indexekre: κ -ra és γ -ra egy-egy populációpár esetén vizsgáljuk meg, hogy a paraméterek egyenlőségét feltételezve mennyi információt veszünk a becslendő paraméterek számának csökkentésével. Mivel erős eredményeket ezzel a kitekintő lépéssel a szerzők sem értek el a brit és a holland adatok kapcsán, illetve a modellezés numerikus megvalósítását is el tudja bonyolítani, ezért dolgozatomban ettől eltekintek, és csak a teljes modelfelírásokra fókuszálok.

A paraméterbecslést követően az egypopulációs modelleknél is a sztochasztikus faktorok vizsgálata volt a következő lépés. Erre legtöbbször ARIMA struktúrákat illesztünk, és ha ezek paraméterbecslésével készen voltunk, akkor jöhetett végső lépésként az előrejelzés. Itt is hasonló a gondolatmenet, azonban kihasználjuk a két populáció jelentette plusz információt: a sztochasztikus faktorok összefüggnek egymással. A koherenciához minden esetben feltételeztük a 3.2. táblában, hogy a páronként képzett idősorkülönbségek mean-reverting tulajdonságúak legyenek. A Li-Zhou-Hardy szerzőhármás az alábbi három struktúrát javasolja az idősorok – mind a mortalitási index, mind a kohorsz index esetén, de én most csak előbbit használom a jelöléshez – modellezéséhez:

RWAR

Angolul *Random Walk and a First Order Autoregression*, vagyis κ_t egy eltolásos véletlen bolyongással, míg a képzett differencia egy elsőrendű autoregressziós folyamatként írható fel:

$$\begin{aligned}\kappa_t^{(i,1)} - \kappa_{t-1}^{(i,1)} &= \mu^{(i,1)} + Z_t^{(i,1)}, \\ \kappa_t^{(i,1)} - \kappa_t^{(i,2)} &= \mu^{(i,2)} + \Phi^{(i,2)}(\kappa_{t-1}^{(i,1)} - \kappa_{t-1}^{(i,2)}) + Z_t^{(i,2)},\end{aligned}\tag{3.3}$$

ahol $\mu^{(i,1)}, \mu^{(i,2)}$ és $\Phi^{(i,2)}$ a modell paraméterei, $Z_t^{(i,1)}$ illetve $Z_t^{(i,2)}$ pedig véletlen értékek. Belátható, hogy amennyiben $|\Phi^{(i,2)}| < 1$ teljesülését megköveteljük, a kívánt $\kappa_t^{(i,1)} - \kappa_t^{(i,2)}$ sor átlaghoz való visszahúzása teljesülni fog.

Ez láthatóan nem egy szimmetrikus struktúra, így ezzel a továbbiakban nem foglalkozom. Viszont a gravitációs modellezésnél ezt a felírást érdemes alkalmazni.

VAR

Az elsőrendű vektor-autoregresszív folyamat (VAR) a következő módon írható fel:

$$\Delta\kappa_t^{(i,j)} = \mu^{(i,j)} + \Phi_1^{(i,j)}\Delta\kappa_{t-1}^{(i,1)} + \Phi_2^{(i,j)}\Delta\kappa_{t-1}^{(i,2)} + Z_t^{(i,j)},\tag{3.4}$$

mindkét populáció esetén, vagyis $j = 1, 2$ -re; azaz ez már egy szimmetrikus struktúra. A becsült paraméterek $\mu^{(i,j)}, \Phi_1^{(i,j)}$ és $\Phi_2^{(i,j)}$, a sztochasztikát pedig ismét a $Z_t^{(i,j)}$ változók adják. Látszik, hogy az autokorrelációt a $\Phi_1^{(i,1)}$ és a $\Phi_2^{(i,2)}$ paraméterek fejezik ki, míg a keresztkorrelációt, vagyis a számunkra oly érdekes populációk közti összefüggőséget, $\Phi_2^{(i,1)}$ és $\Phi_1^{(i,2)}$.

A mean-reverting tulajdonság teljesüléséhez az alábbi feltevéssel szokás élni:

$$\frac{\mu^{(i,1)}}{1 - \Phi_1^{(i,1)} - \Phi_2^{(i,1)}} = \frac{\mu^{(i,2)}}{1 - \Phi_1^{(i,2)} - \Phi_2^{(i,2)}}.\tag{3.5}$$

VECM

Végül pedig a vektor hibakorrekciós modellt (VECM) mutatnám be röviden, ami a VAR-t lényegében csak kiegészíti egy hibakorrekciós taggal, amivel már külön feltétel nélkül teljesül a mean-reverting tulajdonság:

$$\begin{aligned}\Delta\kappa_t^{(i,j)} &= \mu^{(i,j)} + \delta^{(i,j)}(\kappa_{t-1}^{(i,1)} - \kappa_{t-1}^{(i,2)}) + \\ &+ \Phi_1^{(i,j)}\Delta\kappa_{t-1}^{(i,1)} + \Phi_2^{(i,j)}\Delta\kappa_{t-1}^{(i,2)} + Z_t^{(i,j)}.\end{aligned}\tag{3.6}$$

Két fontos megjegyzést kell itt tennem. Egyrészt a fentiekben csak első differenciákkal írtam fel a modelleket, ahogy azt a cikk javasolja. Elemzésem során viszont látni

fogjuk, hogy bizonyos esetekben hasznos volt másod- illetve harmadrendű differenciákkal modellezni. Másrészt a VAR és a VECM közötti lényegi különbség, hogy utóbbinál a folyamatoknak kointegráltaknak kell lenniük. Ha ez nem teljesül, akkor VAR-modellt szabad csak felírni. Erre is több numerikus példát fogok majd bemutatni a következő fejezetben, ahol a becslés folyamatát és az előrejelzést is részletesebben tárgyalom.

3.3. További többpopulációs módszerek

A bemutatott, és általam majd a magyar adatokra alkalmazott fenti módszer tehát egy viszonylag egyszerű, a szimmetrikus kointegrált struktúrára építő modellezési technika. Amennyire kevés eredmény létezik még a mortalitási modellezés többpopulációs irányzatában, olyan sok fejlesztési koncepció látott napvilágot az elmúlt években. A következő néhány bekezdésben ezekről ejtenék szót mintegy kitekintésként a téma átfogóbb ismeretének érdekében. A szerteágazó irodalomból is kiviláglik, hogy ez a kockázati modellezési terület még rengeteg lehetőséget rejt magában, és elképzelhetetlenül sok elemzés vár megvalósításra.

Az egyik legkézenfekvőbb irányzat, hogy összefüggő időbeli változókról lévén szó, modellezzünk kopulákkal. Összetett pénzügyi instrumentumoknál bevett modellezési technika már bő másfél évtizede, hogy a marginális eloszlásokra kopulát alkalmazva írják fel a kívánt együttes eloszlást. Egyszerű példa erre a CDO, melyre Li (2000) fejlesztette ki az árazáshoz szükséges kopulával felírt modellt.¹ Többpopulációs mortalitási modellezéshez is kézenfekvő ez a technika, hiszen a CDO mögöttes, egymással összefüggő instrumentumai statisztikai szempontból nagyon hasonló struktúrát képeznek, mint az egymással kointegrált populációs mortalitások. A téma első cikkei Zhu-Tan-Wang (2014) illetve Chen-MacMinn-Sun (2015) írásai voltak, melyek közös vonása, hogy a komplikáltság miatt egyelőre csak a Lee-Carter modellre vezették le a konstrukciót; másrészt ahogy a CDO-modelleknél, így itt is előtérbe helyezik, hogy a reziduálisok normális eloszlástól eltérő viselkedése külön hangsúlyt kap. A párhuzam ott is felerősödik, hogy minden vonatkozó cikkben mortalitási pénzügyi termékek konstruálásához használják fel a modellt. Erről röviden még fogok szólni, lényegük, hogy a pénzügyi árkockázathoz hasonlóan ezek a longevity kockázatban rejlő kitettséget igyekeznek fedezni.

A kopulás irányzat egyik legújabb cikke már dinamikus kopulákat használ, ami érezhetően egy jobban alátámasztott struktúra mortalitások esetén. Wang, Yang és Huan (2015)

¹A legutóbbi nagy pénzügyi válság egyik kiváltójaként is szokást ezt a modellt emlegetni („a formula, mely megölte a Wall Street-et”), ami jelzi, hogy ezt a struktúrát is fejleszteni kell még, hogy igazán alkalmazható legyen.

eredményei is azt igazolják, hogy érdemes a dinamikus kopulák irányába menni mortalitási instrumentumok árazásánál, ráadásul az ő konstrukciójuk már Renshaw-Haberman modellen alapszik.

Szintén pénzügyi árazásból ismerős technika a Lévy-folyamatok használata. Qin és Jevtic (2016) modellje szimultán ugró folyamatokkal kísérli meg felírni az Egyesült Királyság férfi és női mortalitásait. Ez az irányzat még nagyon kezdetleges, az eredmények sem erősek egyelőre.

Végül pedig a statisztikai modellezés manapság legtöbbször emlegetett területéről, a bayes-i irányról szólnék röviden. A matematikai statisztika két nagy irányzata a klasszikus és a bayes-i. Előbbi hipotéziseket fogalmaz meg, majd ezeket a megfigyelések halmazán teszteli. A mögöttes feltételezésünk mindig az, hogy a keresett ismeretlenek, azaz a paraméterek fixek. Ezt láthattuk a korábban bemutatott összes modellnél. Ezzel szemben a bayes-i irányban keverjük az *a priori* feltételezésünket a paraméter(ek) eloszlására a megfigyelttel (*a posteriori*). Ez azt az elméleti hasznosságot teszi lehetővé többek közt, hogy a régi megfigyelések által implikált eloszlást a priorinak tekintjük, és így a posterioriként könnyedén vonhatjuk be a legújabb megfigyeléseinket. Mortalitások esetén egyébként ez a sokszor megfigyelhető strukturális törések esetén nagyon hasznos lehet, például később be fogom mutatni, hogy a '80-as években a férfiaknál is volt egy erős törés, ami miatt a klasszikus modellezési eszközökkel körültekintően kell bánni.

A halandósági modellezés bayes-i irányzatának úttörője Cairns és szerzőtársainak cikke (2011) volt, mely APC-modellkerettel igen jól működő struktúrát talált az angol-wales-i adatokra. Hollandiára Ouburg 2013-as, a témát is jól összefoglaló publikációját ajánlom az Olvasó figyelmébe. A legkiterjedtebb és -újabb írás a témában a Fung-Peters-Sevchenko (2016) szerzőhármastól jelent meg. Ők a GAPC modellcsalád mind a hat általam bemutatott tagjára felírják a bayes-i struktúrát, és vizsgálják azt dán mortalitásokon. Magyarország esetére is ajánlatos lenne egy hasonló elemzés.

4. fejezet

Numerikus elemzés

Az elméleti áttekintés után térjünk rá dolgozatom valós adatokon modellezett eredményeire. Ahogy említettem, szimmetrikus struktúrával szeretném a magyar országos halandósági fejlődést vizsgálni. Kézenfekvő, hogy a férfi és női szubpopulációt külön tekintsem, ahogy azt lényegében minden mai mortalitási előrejelzés esetén teszik. A magyar népességben megfigyelhető folyamatok egyik fő kérdése, hogy a Nyugathoz konvergálunk-e, legyen az most bármilyen értelemben is véve. Adódott az ötlet, hogy a korábbiakban bemutatott többpopulációs modellezéseket az osztrák női illetve férfi halandóságokkal együtt végezzem el a magyarra. Ezzel kontrasztban egy szintén a Vasfüggöny keleti oldalán létező ország, Szlovákia adatait vettem be az elemzésbe. A második világháborút követő történelmi, gazdasági hasonlóság miatt gondoltam úgy, hogy ez egy erősebb, kointegráltabb kapcsolat lehet, de a végső következtetésekhez mindkét elemzés rendkívül hasznosnak bizonyult.

Így tehát négy populációpárra írok fel kétpopulációs mortalitást: magyar női az osztrák illetve a szlovák nőivel, valamint ugyanígy a magyar férfi populációt először az osztrák, majd a szlovák férfi mortalitási fejlődéssel modellezem együtt. Végezetül egy ötödik párost is megvizsgállok: mi történik, ha a magyar férfi és női társadalmat helyezem közös keretbe. Kézenfekvő, hiszen melyik két részsokaság halandósága függne össze szorosabban, mint az országhatárokon belüli nőké és férfiaké. A fő kérdés pedig minden esetben az lesz, hogy ezek a párosítások – végigvizsgálva a 6 modellt – mely esetekben adnak jobb előrejelzést az egypopulációsnál. Végső soron érdemes-e a magyar női illetve férfi mortalitási modellezésnél használni a többpopulációs módszereket; ha igen, akkor pedig hol és hogyan?

4.1. Adatok és a modellezés menete

Az adatokat minden populációra 2016. november 4-én a *Human Mortality Database* (HMD) honlapjáról (www.mortality.org) töltöttem le. Mindhárom nemzet mindkét nemére elér-

hetők a halálozási számok ($D_{x,t}$), illetve a kezdeti kitettségek ($E_{x,t}^0$) idősorosan 1950. és 2014. között. Központi kitettségek az (1.2) egyenlet alapján könnyen számolhatóak, amennyiben azzal modelleznénk. Megjegyzendő, hogy Szlovákia esetén a csehszlovák állam kettéválását megelőzően is léteztek külön területi alapú statisztikák, így lehetséges, hogy 1950-ig visszamenően elérhetőek a kizárólag szlovák mortalitási adatok.

Az időszak (t dimenzió) után a modellek korintervallumát (x dimenzió) kell megválasztani. A HMD az adatokat 0-tól egészen 110 éves korig közli. 80 éves kor felett viszont a jóval kisebb populációméretetek miatt kisimítják ezeket az adatokat. Hasonló elemzéseknél ez egy bevett eljárás magasabb életkoroknál. A HMD Kannisto modelljét használja¹, de több eljárás is létezik, ezekkel most nem foglalkoznék. Felső korhatárnak tehát a 80 évet választom, hogy ez a simítás ne okozzon nem kívánt hatásokat elemzésemnél. Alsó korhatárként az életbiztosítások és a nyugdíj-megtakarítások piacát figyelembe véve a 30 éves kort tartom indokolhatónak. Még egyszer kiemelném, hogy ettől eltérő, akár teljes kort lefedő modellek építése is szép eredményekre vezethet.

Az adatok áttekintése után vegyük át a modellezés általános lépéseit. Az előző fejezetet követve felső indexben j -vel ($j = 1, 2$) fogom jelölni az első illetve a második választott populáció megfelelő változóit.

1. lépés: Egypopulációs modellezés

A 3. fejezetben látott elméleti struktúra szerint először a két populációt egymástól függetlenül modellezem. A választott adathalmazokon tehát becslést készíték a 3.1. táblázatban jelölt hat modell paramétereire mindkét populáció esetében. Követve Villegas és szerzőtársainak (2016) felépítését, én is mindenhol logit linkfüggvényt alkalmaztam a becslés során a könnyebb összehasonlíthatóság kedvéért. A modellek futtatását minden esetben az R nyílt forráskódú szoftver **StMoMo** csomagjának segítségével végeztem el.²

A többpopulációs modellválasztás már ebben a lépésben elkezdődik. Egyrészt a reziduálisok hőtérképe alapján kaphatunk egy képet arról, hogy az egyenkénti modell mennyire teljesít megfelelően. Erre már mutattam példát: a 2.1. ábrán a Lee-Carter reziduális hőtérképét láthattuk, és ahogy azt majd a konkrét szubpopulációs illesztéseken is bemutatom, legtöbb esetben emiatt mindig a bonyolultabb struktúrákat fogom preferálni, tehát kohorszhatást például feltétlen szerepeltetek a modellben.

A másik, mortalitási modellezésnél figyelembe veendő jelenség a *strukturális törés*. Az

¹A további részletek megtalálhatóak a *HMD Methods Protocol* dokumentumának vonatkozó részeiben, mely elérhető itt: <http://www.mortality.org/Public/Docs/MethodsProtocol.pdf>

²A kapcsolódó saját kódok, script-ek kérésre elérhetőek, illetve a dolgozathoz csatolt CD-mellékleten megtekinthetőek, futtathatóak.

alapmodellek mindegyike tartalmaz legalább egy idősoros változót. Ezek a tapasztalatok szerint sokszor törést mutatnak. Matematikailag nem egy precíz fogalom ez, de több esetben egyértelműen felfedezhető olyan trendfordulás, ami miatt nem (csak) 1950-es induló évvel érdemes elvégezni a modellillesztést. Ezeknek a strukturális töréseknek az okait nem vizsgálom az elemzésem során, csak egy szűrőként alkalmazom a modellválasztásnál.

2. lépés: Legjobb modellek kiválasztása az adott populációpárosra

Miután megvannak a jelöltek az egyes populációkra, kell a közös modellek statisztikai jellemzőit megvizsgálni. Adott modellpárosok esetén – ahogy azt a (3.1) és (3.2) egyenleteknél bemutattam – felírhatóak a log-likelihood értékek. Ezekből a maximálisat keressük. Emellett viszont még figyelembe fogom venni az Akaike és a bayes-i információs kritériumokat (AIC és BIC) a választáshoz, ami több esetben két, hasonlóan jónak tűnő modellre fog esni.

3. lépés: Összefüggőség vizsgálata

A többpopulációs modellezés legkritikusabb pontja a sztochasztikus faktorok összefüggőségének vizsgálata, majd modellezése. A páronkénti idősorokra a korábban felírtak szerint hasznos lenne, ha mean-reverting tulajdonság teljesülne. Láttuk viszont, hogy a koherenciához ez nem szükséges, de elégséges tulajdonság, így meg fogom vizsgálni ezek teljesülését *Augement Dickey-Fuller teszttel*. Azonban a nem mean-reverting tulajdonságú idősorokkal is tovább dolgozom, ugyanis számunkra a kointegrációs modell felírása lenne a cél, amihez Johansen-féle kointegrációs tesztet fogok használni. Ezt a modellezés és a tesztek könnyedsége miatt az R-ből importált adatokkal Eviews-ban végzem el, ami idősoros elemzésekhez az egyik legerősebb, pénzügyekben sokszor alkalmazott szoftver.

A kointegráció vizsgálatánál fontos kiemelni, hogy a késleltetések (angolul *lag*-ek) száma egy-egy idősornál is kardinális kérdés. Az egypopulációs modelleknél ezzel nem foglalkoztam külön, csak a szakirodalom által leginkább hatékonynak tartott modelleket írtam fel a 2.1. táblázatban. Az elemzésemnél azonban külön megvizsgálom, hogy mely idősorokra mely késleltetéspár bizonyul számos mutató alapján a leghatékonyabbnak, és ennek együtthatóit fogom aztán megbecsülni.

Ahogy már említettem, minden populációpárnál a szimmetrikus modellkeretnél szeretnék maradni. Ezért kointegrációs kapcsolat esetén VECM modellt, ellenkező esetben pedig VAR modellt fogok alkalmazni, amiket korábban a 3.2.2. szakaszban mutattam be.

4. lépés: Kétpopulációs modell

A modell részletes kiválasztási folyamatának végén tehát elkészítem az egyes faktorok VECM vagy VAR modellegyütthathatóinak becsléseit. Ezekkel előrejelzés adható az adott két populáció sztochasztikus indexeire, amelyeket visszahelyettesítve a használt alapmodellekbe megkaphatjuk a részsokaságokra becsült halandóságok előrejelzését.

Emellett a becslések bizonytalanságát felhasználva az előrejelzés bizonytalanságát is érdemes lehet modellezni, használva a korábban már említett Monte Carlo szimulációs illetve félparaméteres bootstrap módszereket. A dolgozat elemzése erre a vizsgálatra már nem terjed ki, a később bemutatott alkalmazásoknál ugyanis nem tartottam szükségesnek. Más, kiterjedtebb alkalmazásoknál – főleg S2 vonatkozásában – viszont ez egy hasznos kiegészítése lehet a hasonló elemzéseknek.

5. lépés: Visszatesztelés

Az előrejelzést különböző területeken alkalmazhatjuk, erre fogok példákat hozni a magyar mortalitások esetén. Emellett historikus statisztikai modellezésnél egy nagyon fontos lépés a visszatesztelés. Lényege, hogy a vizsgált időszakot két részre osztjuk: a korábbi, tanuló időszakra – itt építjük a modellt –, illetve az azt követő tesztelőre – itt mérjük össze az előrejelzést a valós adatokkal. Dolgozatom egyik fő kérdése, hogy hogyan teljesítenek a többpopulációs modellek az egypopulációs társaikhoz képest. Ennek megválaszolására ezt a tanuló-tesztelő metódust fogom használni: a tesztelő időszaknak a 2005 és 2014 közti 10 éves időintervallumot választottam minden esetben az összehasonlíthatóság kedvéért. Ez bővíthető lenne, de több esetben a '80-as évektől látható egy erős strukturális törés, ezért – hogy a modell építése viszonylag bő tanuló időszakon legyen elvégezhető – a 2005-2014 időszakot tartom a megfelelő tesztelési periódusnak.

A teljesítmények összehasonlíthatóságához egy jól értelmezhető és mortalitási ráta esetén is érvényes képet mutató statisztika szükséges. A halandóságok kortól való nagyságrendi függése miatt mindenképp százalékos mértékegységű mérőszám kívánatos. Az *átlagos abszolút százalékos eltérés* (angolul *mean absolute percentage error* vagy MAPE) erre a célra kiválóan megfelel:

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|, \quad (4.1)$$

ahol $n = 10$ lesz most minden esetben a megfigyelések, vagyis a tesztelő időszak éveinek száma. A_t jelöli a t . időszak ténylegesen megfigyelt mortalitási rátáit, F_t pedig az előrejelzett értéket. Ezeket a statisztikákat fogom modellenként felírni, illetve korosztályokra bontva is vizsgálni, hogy eldönthető legyen, mely modell milyen hatékonyan teljesít.

4.2. Kétpopulációs modellezés és eredményei

Legelőször is megjegyezném, hogy a Renshaw-Haberman modell a már említett konvergálási probléma miatt kizárólag a magyar férfi adatokra futott le sikeresen. A többpopulációs modellezésnél így egyik párnál sem alkalmaztam ezt a struktúrát.

A következő táblázatokban tekintsük át az öt populációpárosra a teljes időhorizonton illesztett modelleket és azok statisztikáit:

4.1. táblázat: A kétpopulációs modellek log-likelihood értékei

Populációpár	LC	CBD	APC	M7	Plat
magyar ffi - osztrák ffi	-130928	-125391	-45020	-35948	-37617
magyar női - osztrák női	-90787	-141848	-37086	-32508	-34004
magyar ffi - szlovák ffi	-146644	-136526	-50492	-38303	-40462
magyar női - szlovák női	-109055	-141320	-42143	-35547	-37044
magyar ffi - magyar női	-123053	-140472	-45464	-33496	-36403

4.2. táblázat: A kétpopulációs modellek AIC értékei

Populációpár	LC	CBD	APC	M7	Plat
magyar ffi - osztrák ffi	262515	251303	90927	73100	76374
magyar női - osztrák női	182234	284217	75060	66219	69148
magyar ffi - szlovák ffi	293949	273571	101872	77810	82064
magyar női - szlovák női	218770	283161	85175	72299	75229
magyar ffi - magyar női	246766	281463	91817	68195	73946

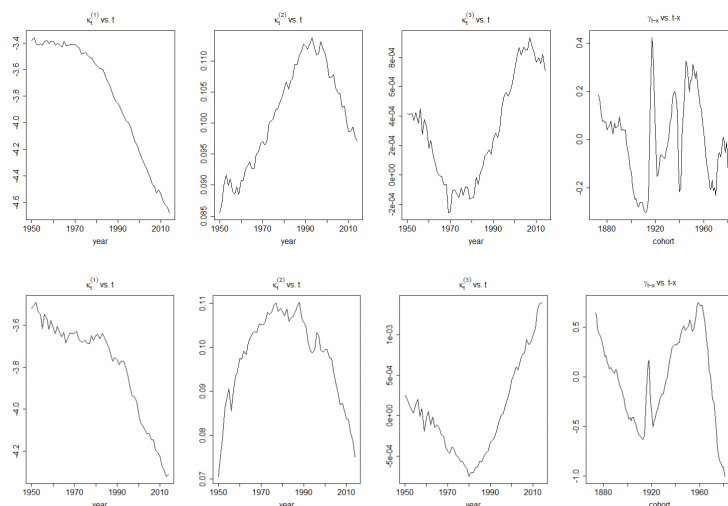
4.3. táblázat: A kétpopulációs modellek BIC értékei

Populációpár	LC	CBD	APC	M7	Plat
magyar ffi - osztrák ffi	264529	252890	93637	76774	79853
magyar női - osztrák női	184248	285804	77769	69893	72626
magyar ffi - szlovák ffi	295963	275158	104582	81484	85543
magyar női - szlovák női	220784	284748	87884	75972	78707
magyar ffi - magyar női	248779	283050	94526	71869	77425

Fontos, hogy ezek a számok azonos típusú modellt feltételeznek mindkét populációra, tehát például nem kever M7-et Plat modellel. A legtöbb struktúrájánál felírható lenne ilyen vegyes pár is, viszont ez már annyi lehetőséget tesz elérhetővé az elemzésnél a vizsgált öt populációra, mely meghaladja ezen dolgozat kereteit. A fenti táblákból leolvasható, hogy általában az APC, M7 és Plat modellek teljesítenek a legjobban. Ráadásul a reziduális hőtérkép is számos esetben kizárhatóvá tette az LC és CBD struktúrákat, így utóbbiakkal egyik populációpárnál sem foglalkozom a továbbiakban, csak az egy populációs modellezésnél térek vissza rájuk.

4.2.1. Magyar-osztrák férfi

Az elemzést a magyar és osztrák férfi populációk együttes modellezésével kezdem. A log-likelihood értékek miatt, illetve a hőtérképek alapján az APC modellt elvettem. Az M7 modellnél az alábbi ábrán láthatjuk a sztochasztikus faktorok alakulását 1950-től:



4.1. ábra. Fent az osztrák, lent a magyar férfi M7-es modell sztochasztikus faktorainak alakulása 1950 és 2014 között. (saját szerkesztés)

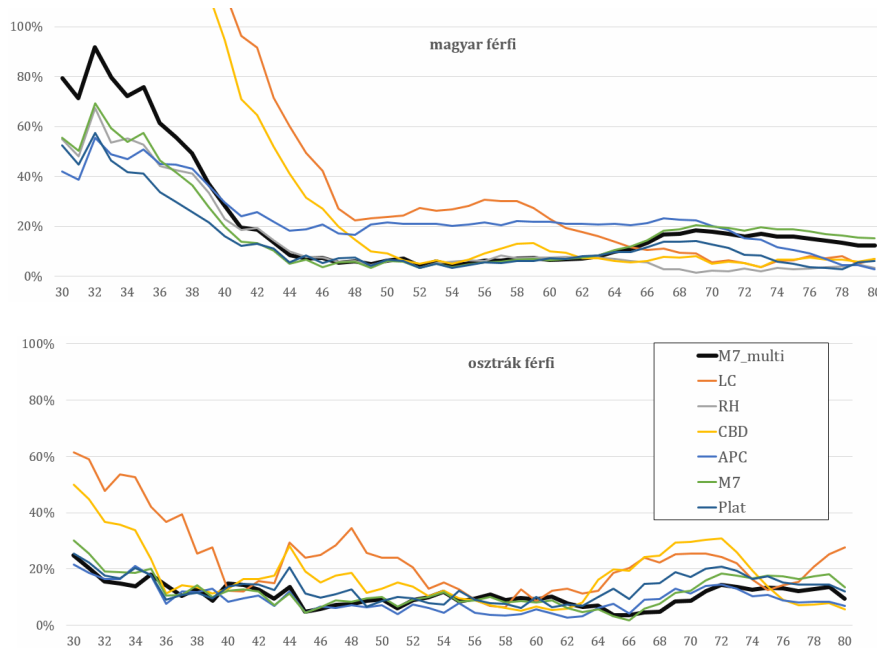
Mindkét populációra éles törés figyelhető meg az 1980-as évek elején a mortalitási indexeknél. Így az M7 modell esetén az 1985-2004 közötti éveket választottam tanuló időszaknak. A Plat modellnél nem figyelhető meg egyik populáció esetén sem ilyen éles törés az idősoroknál, így ott a teljes, 1950-től tartó időhorizonton végeztem el a becslést.

Az M7 modell γ_{t-x}^j idősorai és differenciája is átment az Augement Dickey-Fuller (ADF) teszten, vagyis a mean-reverting tulajdonság fennáll. A késleltetések számának vizsgálatakor, hogy Johansen-tesztet végezhetünk, olyan lag-struktúrát kell választanunk,

melyre az idősorok nem stacionáriusak, de ugyanazon rendű differenciájuk már igen. Ebben az esetben a Johansen-teszt 1 késleltetés esetén még $p = 0,0026$ értéket adott, ami azt jelenti, hogy elvethetjük azt a nullhipotézist, miszerint nincs kointegrációs kapcsolat a két populáció γ_{t-x}^j kohorsz indexe között. Így itt egy-késleltetésű VECM modell írható fel. A konkrét egyenletek megtalálhatóak a CD-melléklet vonatkozó Excel-fájljaiban, a továbbiakban ezek fontosabb tulajdonságait közlöm csak.

A hibatag együttthatója mindkét esetben szignifikánsan negatív, ami azt jelenti, hogy hosszú távú kointegrációs kapcsolat (angolul *long-run causality*) van a két populáció kohorsz indexe között. A három mortalitási index közül $\kappa_t^{1,j}$ esetén már egészen a 3-lages VECM bizonyult a legjobb választásnak: Johansen-teszten átment (0,0003), az AIC érték pedig erre a legalacsonyabb. $\kappa_t^{2,j}$ esetén a mean-reverting tulajdonság sem teljesül az ADF teszt szerint, de minden késleltetéses lehetőséggel a Johansen-teszt is azt mutatta, hogy nincs kointegrációs kapcsolat. A 4.1. ábrára pillantva valóban felfedezhető, hogy bár ezen indexek alakulása hasonló csúcsos alakot ír le, azonban más a meredekség, a csúcs időpontja és az utolsó évtized lecsengése is – erre tehát VAR(1) modell illeszthető. Végül $\kappa_t^{3,j}$ esetén szintén azt tapasztaltam, hogy a VAR(1) modell használható kizárólag.

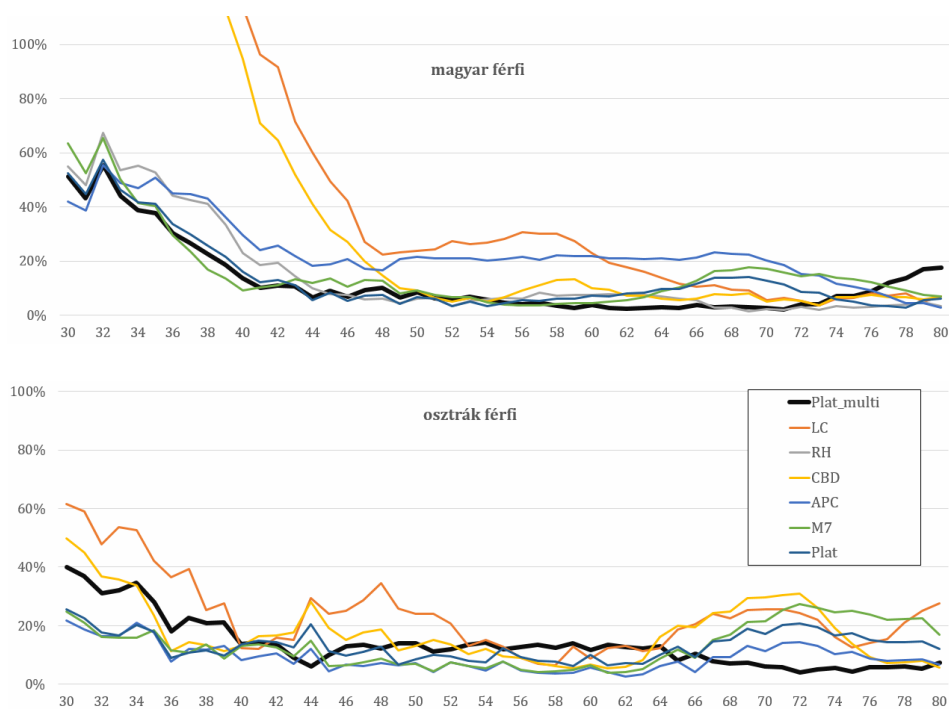
A modell teljesítményének megállapításához tekintsük korosztályonként (vízszintes tengely) ábrázolva az 1950-2004 időszakon épített egy- és kétpopulációs modellek 2005-2014 évekre becsült előrejelzéseiből számolt MAPE értékeket (függőleges tengely) külön a magyar és osztrák valós mortalitásokra:



4.2. ábra. Magyar (fent) és osztrák (lent) férfi modellek MAPE értékei. (saját szerkesztés)

Az ábráról egyrészt leolvasható, hogy az egypopulációs LC és CBD modellek valóban gyengén teljesítenek, főként a 40 alatti korosztályra, ami már az ábrára sem fért rá a magyar esetben. A legfontosabb megállapítás számunkra, hogy a multipopulációs M7 igazán egyik nemzetre sem tudta javítani az egypopulációs párját, kivéve az idősebb korosztályokban, de nem jelentősen, ráadásul ott eleve gyengébben teljesít az M7. A hazai férfiakra 45-65 éves korban viszont az egyik legjobb modell, tehát a felső középkorosztályban érdemes figyelembe venni az osztrákokkal közös előrejelzést.

Nézzük az 1950-es kezdetű Plat modellt. Itt mindegyik sztochasztikus faktor mean-revertingnek bizonyult, és átment a Johansen-teszten. γ_{t-x}^j és $\kappa_t^{1,j}$ esetén VECM(1) modell írható fel, míg $\kappa_t^{2,j}$ -re AIC alapján is ajánlatosabb választás volt a 2-időszakos használata. A mortalitási indexek közül egyébként csak a másodikra áll, hogy hosszútávú kapcsolat van a két idősor között. Az előzőhöz hasonló MAPE ábra most így alakul az összes egypopulációs modellt és a kétpopulációs Plat modellt tekintve:



4.3. ábra. Magyar (fent) és osztrák (lent) férfi modellek MAPE értékei. (saját szerkesztés)

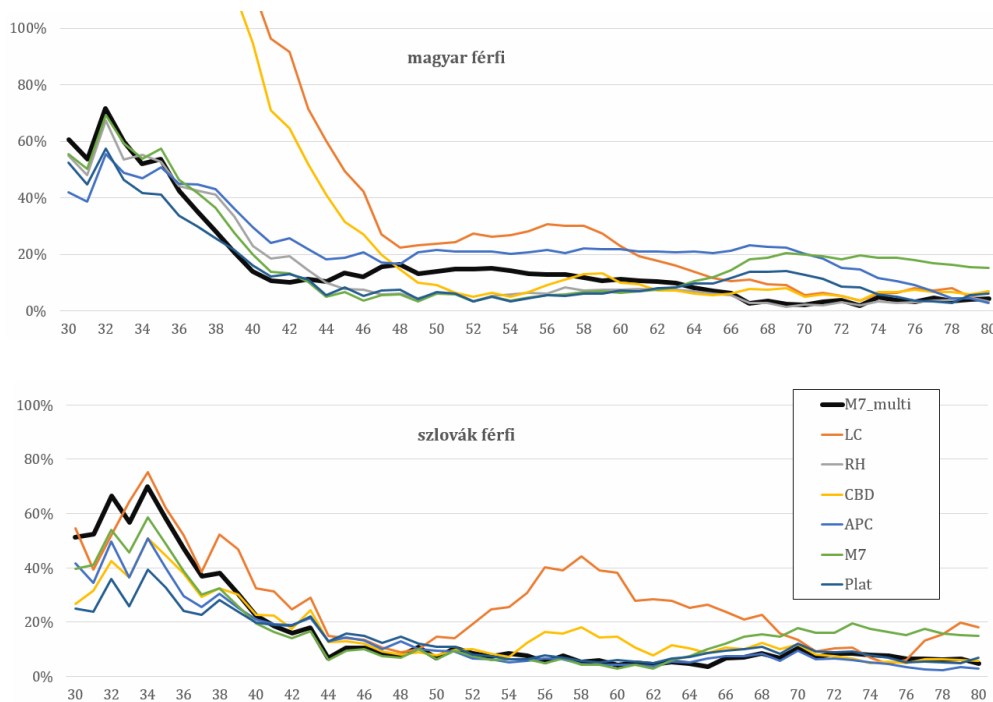
Ez már jelentős javulást mutat az egypopulációsokhoz képest: az átlagos MAPE érték a legalacsonyabb a magyar populációra (12,55%). Kiemelkedően az 55-73 éves korosztályra teljesít, a legidősebbekre viszont jóval gyengébben. Az osztrák adatokon pont az ellenkezője figyelhető meg: 65 és 80 éves kor között erős. Ebből megfogalmazható egy olyan hipotézis, hogy a gazdaságilag fejlettebb ország férfi nyugdíjas mortalitásait érdemes egy szegényebb

szomszéd hasonló populációjával együtt modellezni, míg fordítva, az alacsonyabb gazdasági erejű társadalomban a középkorú populációra érdemes használni ezt az eljárást. A hipotézis alátámasztáshoz persze sokkal több közép-európai elemzésre lenne szükség, mindenesetre érdekes lesz ennek fényében vizsgálni a szlovákokkal vett közös modelleket, illetve a női megfelelőket.

4.2.2. Magyar-szlovák férfi

Maradjunk egyelőre a férfiaknál, és nézzük meg a szlovákokkal közös modelleket. A magyar félről ugyanaz a kezdeti okfejtés mondható el strukturális törés tekintetében, a szlovákoknál pedig semmi jelentős eltérést nem tapasztaltam a magyar vagy osztrák folyamatokhoz képest. A log-likelihood, AIC és BIC értékek alapján is egyértelmű volt a döntés, miszerint erre a populációpárra ugyanazt a két modellt futtatom le, mint a magyar-osztrákra: 1985-től indított M7 és 1950-es Plat.

M7 esetén ismét VAR modell illeszthető csak $\kappa_t^{2,j}$ és $\kappa_t^{3,j}$ folyamatokra, míg $\kappa_t^{1,j}$ és γ_{t-x}^j esetén is a kétidőszakos VECM tűnt az optimális választásnak. A MAPE értékek:

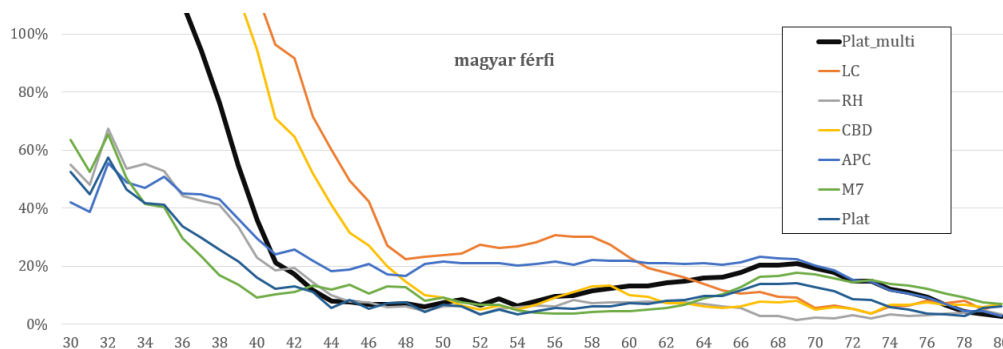


4.4. ábra. Magyar (fent) és szlovák (lent) férfi modellek MAPE értékei. (saját szerkesztés)

Itt valóban az osztrákkal közös modell ellenkezőjét figyelhetjük meg: a magyar férfi mortalitások előrejelzése jelentősen javult a nyugdíjas korosztályra, de rosszabb a középkorúak esetében. A szlovákoknál még ennél is jobban teljesít a magyarokkal képzett modell,

szinte végig vezet, átlagos MAPE értéke ennek a modellnek a legjobb. Ráadásul az M7 kétpopulációs változata mindkét országban erősen felülteljesíti szimpla változatát.

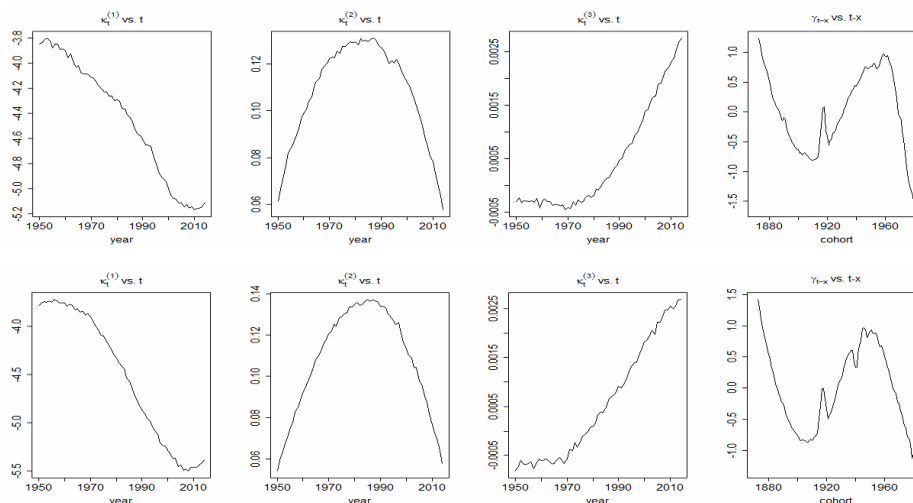
Ez, az osztrákokkal ellentétes javulás, a Plat modellnél is fennmarad, a szlovák-magyar szisztéma ugyanis jóval gyengébben teljesít a hazai férfi halandósági megfigyeléseken:



4.5. ábra. Magyar férfi modellekre mért MAPE értékek. (saját szerkesztés)

4.2.3. Magyar-osztrák női

A nők esetében is kezdjük az osztrák összevetéssel. A férfiakhoz képest itt az M7 és a Plat modell is egy kicsit eltérő strukturális törést mutat:



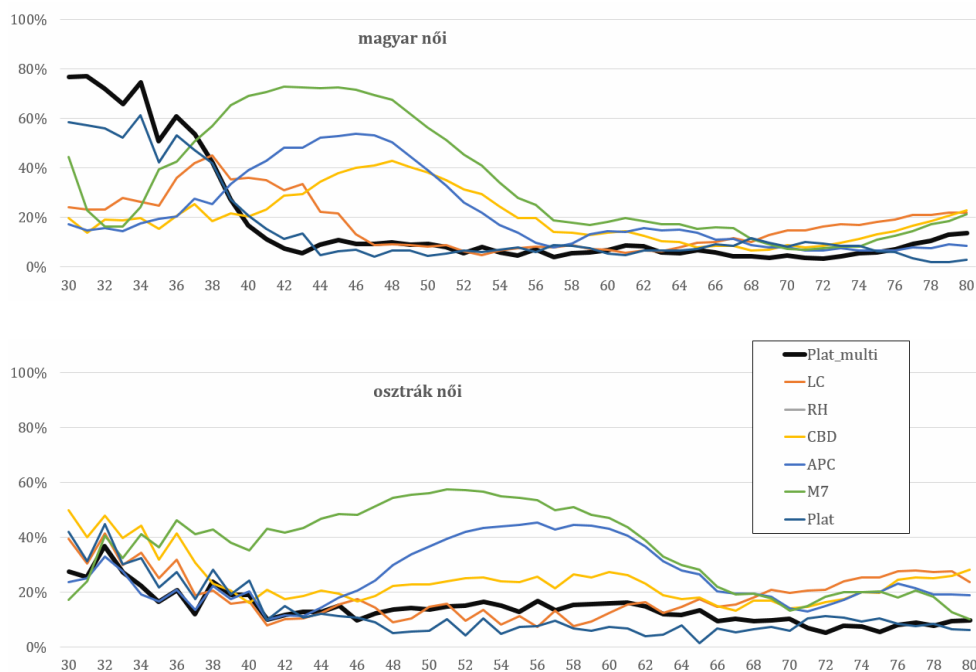
4.6. ábra. Magyar (fent) és osztrák (lent) női M7-es modell sztochasztikus faktorainak alakulása 1950 és 2014 között. (Adatok forrása: saját szerkesztés.)

Az M7 esetében a 4.6. ábrán látható az idősorok alakulása. Három fontos megállapítást tehetünk ez alapján. Egyrészt az idősorok itt is nyilvánvaló együttmozgást mutatnak,

másrészt a $\kappa_t^{1,j}$ sorok 2000-es években látható trendfordulója miatt ez a modell biztosan gyengébb előrejelzést adna, ha a tanuló időszak nem terjedne ki ezen évekre. Így az M7 modellt ebben a vizsgálatban elvettem. Végül szembeötlő, hogy strukturális törés 1970. környékén látható mindkét nemzet esetén. Hasonló megállapítás tehető a Plat modellnél, ezért a nőknél a végső választás az 1970-2004 tanuló időszakos Plat modellre esett.

A kohorsz index erősen mean-revertingnek bizonyult az ADF teszt alapján: -4,57 a tesztstatisztika a külön vett idősorokra, míg -3,41 a két populáció között képzett differenciára, miközben az 1%-os kritikus érték -3,51, az 5%-os pedig -2,89 – ezeknél alacsonyabb értékek mellett nem vehetjük el a mean-reverting tulajdonságot mint nullhipotézist. Késletetések számában a Johansen kointegrációs teszt egészen 5-ig alátámasztja nullhipotézisünket, az AIC alapján végül a 2-lages struktúrát választottam. A hibakorrekciós együtthatók itt is szignifikánsan negatívak lettek, vagyis hosszú távú kapcsolat áll fent a két populáció kohorsz idősora között – érdemes tehát a többpopulációs módszert használni.

A $\kappa_t^{1,j}$ mortalitási index sem mean-reverting tulajdonságot (differenciában -0,75-ös tesztstatisztika), sem Johansen-teszten való megfelelést nem mutatott (0,08 a kívánt maximális 0,05 helyett), így itt VAR(1) modellt alkalmaztam. $\kappa_t^{2,j}$ esetén viszont már kointegrációs kapcsolat fedezhető fel a tesztek alapján, és itt is a 2-lages struktúra tűnt a leghatékonyabbnak. A modell becslött együtthatói ismét megtalálhatóak a csatolmányban.

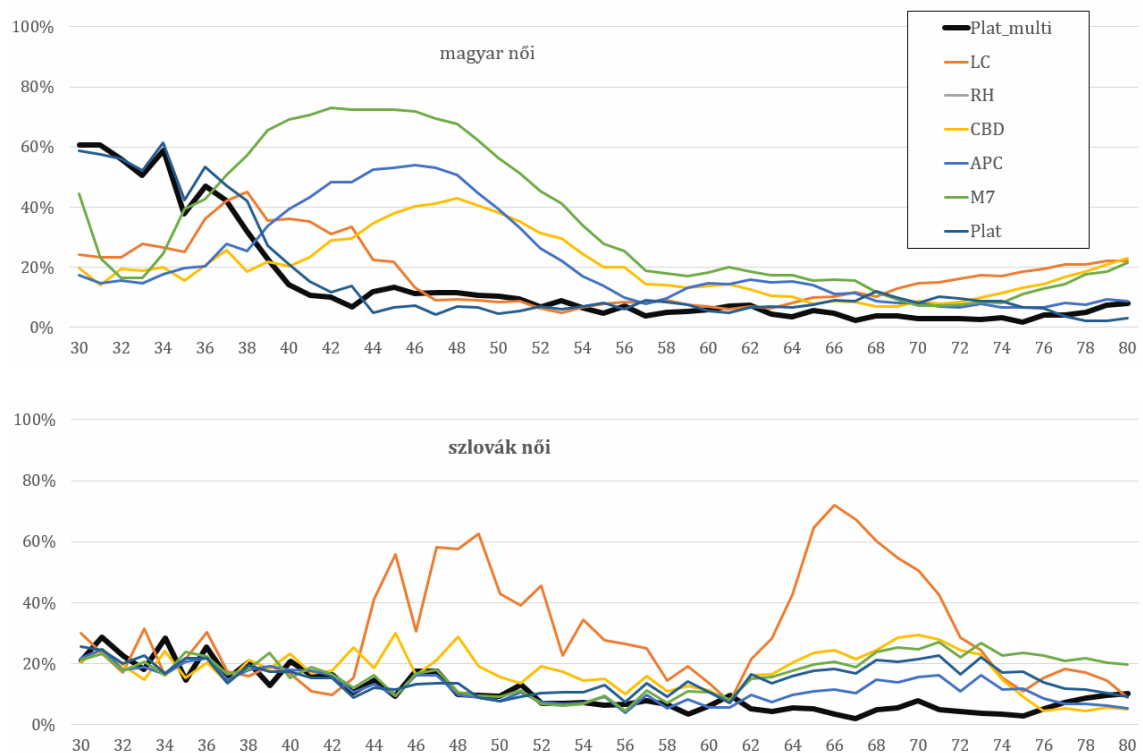


4.7. ábra. Magyar (fent) és osztrák (lent) női modellek MAPE értékei. (saját szerkesztés)

A 4.7. ábrán jól látható egyrészt, hogy az M7 modell tényleg nagyon gyengén teljesített a tesztidőszakban. Másrészt a férfiakkal ellentétben itt azt a megállapítást tehetjük, hogy a magyar női társadalom mortalitását az idősebb korosztályokra is érdemes az osztrákokkal együtt modellezni. 40 éves kor alatt ez nem igaz, ott rendkívül gyenge a modell teljesítménye MAPE alapján, ezzel szemben 40 és 77 között egyértelműen a legjobb. A kritikus 65-75 éves korban viszont messze jobb az összes egy populációsnál. Az osztrák teszt eredményekre már nem ilyen egyértelmű a helyzet, de hasonlóan jó eredményeket kaptunk, mint az eredeti egy populációs modellel.

4.2.4. Magyar-szlovák női

Tekintsük ezek után a nők esetén is a szlovákokkal közös struktúrákat. Követve a férfi modellezésénél alkalmazott logikát, itt is ugyanazt írom fel, mint az osztrák modellnél, ugyanis merőben eltérő idősoros viselkedés nem tapasztalható. A lényegi különbség, hogy most $\kappa_t^{1,j}$ 1 helyett 2 késleltetést tartalmazó VAR-modellként bizonyult a legalkalmasabb struktúrának, míg $\kappa_t^{2,j}$ -re nem lehet VECM-et alkalmazni a Johansen-teszt szerint, így itt maradt a VAR(1).



4.8. ábra. Magyar (fent) és szlovák (lent) női modellek MAPE értékei. (saját szerkesztés)

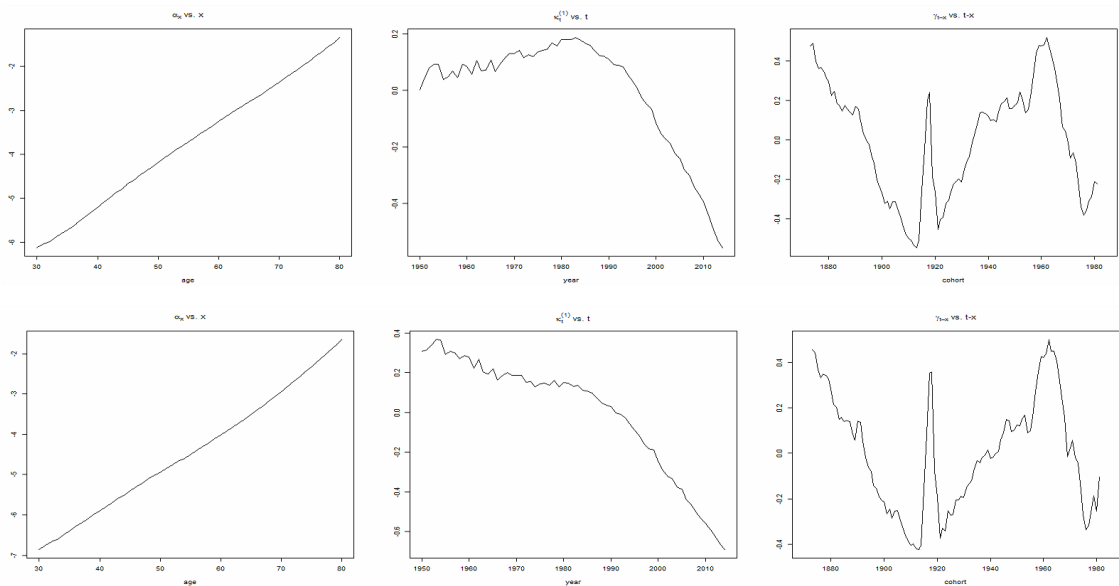
Merőben hasonló képet láthatunk a magyar előrejelzésnél, mint az osztrákkal közös

modellnél. Nőkre tehát mindkét többpopulációs struktúra hasonló korfüggő visszatesztelési értékeket produkált, míg a férfiaknál ellentétes viselkedést tapasztaltunk a szlovák illetve osztrák esetben. Erre a jelenségre igazán jó magyarázatot nem találtam. A halálózáshoz vezető életvitel tekintetében úgy tűnik, a kelet-közép-európai országok férfi társadalma sokkal inkább különül el a nyugatitól, szemben a nőkkel.

A 4.8.-as ábrán a szlovák esetre tekintve még megállapítható, hogy itt a magyarokkal együttes modell messze a legjobb átlagos MAPE értéket adta, vagyis Szlovákia női társadalmára is kimondottan ajánlatos ez a modellezési technika.

4.2.5. Magyar férfi-magyar női

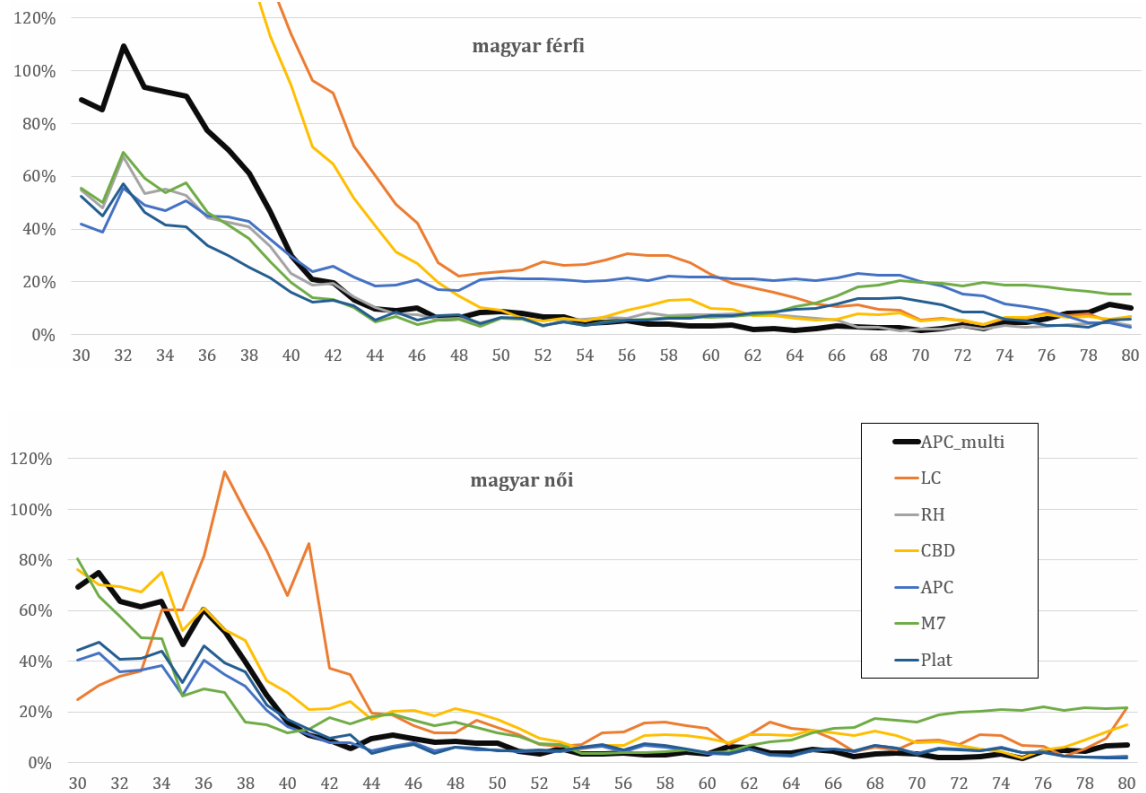
Végezetül modellezzük hazánk két szubpopulációját a közös modellkeretben! Két struktúrát írtam fel végül: a magyar férfinál már bizonyított 1950-es kezdetű Plat modellt, illetve az eddig még nem alkalmazott APC-t. Törés tekintetében a Plat konstrukcióban a nőknél 1970-nél látható egy dinamikaváltás, de ez nem olyan erős, hogy emiatt 20 évnnyi megfigyelést kihagyjunk a férfiakkal közös modellépítésnél. Az APC modellnél ettől eltérő képet láthatunk:



4.9. ábra. Magyar férfi (fent) és női (lent) APC modell sztochasztikus faktorainak alakulása 1950 és 2014 között. (saját szerkesztés)

A kohorszhatás magas volatilitása miatt nem befolyásolja a strukturális törés vizsgálatát, viszont a mortalitási indexnél egyértelmű trendváltás figyelhető meg 1985 környékén, így választásom is erre a kezdőévre esett az APC modellnél. A kohorsz indexre mindkét

modellnél 2-időszakos VECM-et írhattam fel, míg a mortalitási indexek esetében APC-nél VAR(1), míg Plat modellnél VAR(2) bizonyult alkalmazhatónak. Az eredmények:



4.10. ábra. Magyar férfi (fent) és női (lent) MAPE értékek. (saját szerkesztés)

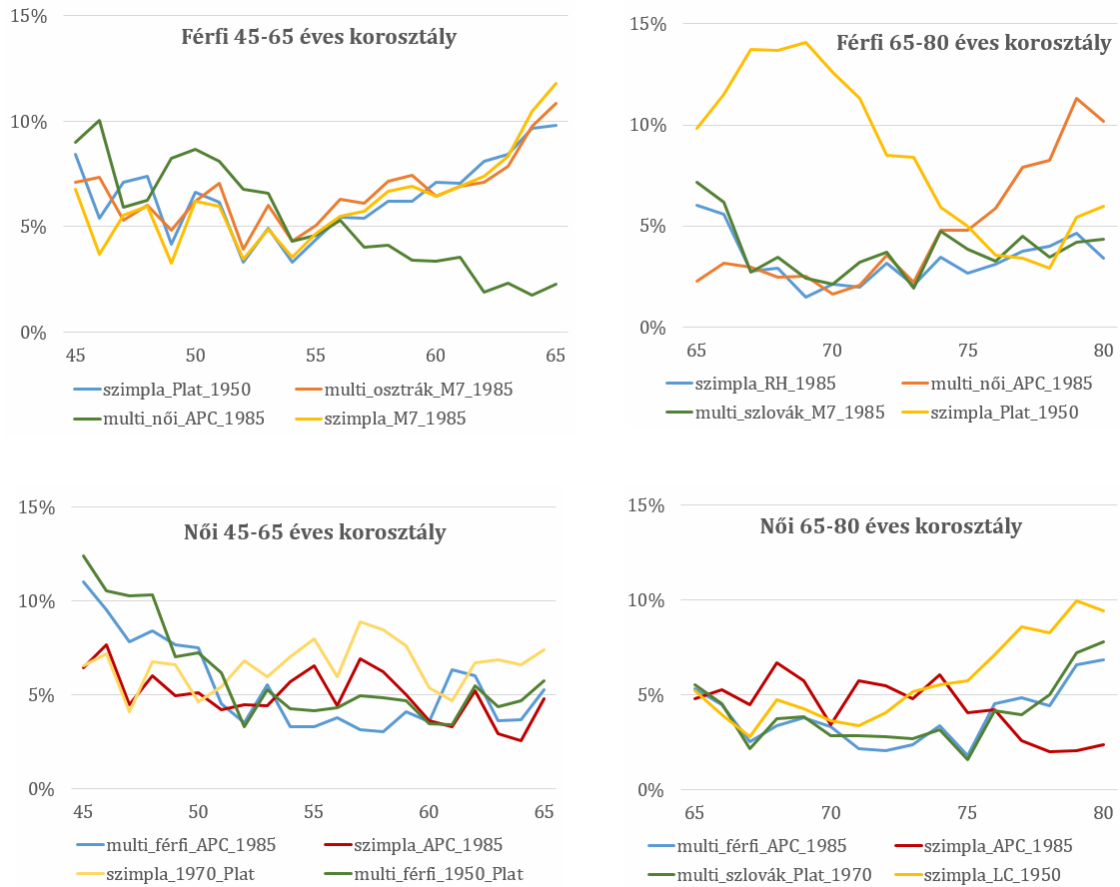
Az együttes APC modell – ismét leszámítva a 40 év alatti korosztályokat – kiválóan teljesít mindkét nem esetében. A férfiaknál ráadásul mindezt úgy, hogy az egypopulációs APC az egyik leggyengébb. Plat multipopulációs esetben hasonló, de gyengébb képet láthatunk, így ennek bemutatásától eltekintek, de az eredmények megtekinthetőek a csatolt CD-mellékleten erre az elemzésre is.

4.3. Összehasonlítás az egypopulációs modellek eredményeivel, alkalmazások

4.3.1. A legjobb magyar mortalitási modellek

A rengeteg bemutatott modell után fókuszáljunk most csak a magyar férfi és női mortalitásokra, és nézzük meg, melyek a legjobban teljesítő struktúrák ezekre a populációkra! Észrevehető, hogy a fiatalabb középkorúakra (45 év alattiak) minden modell bőven túlbecsülte a tesztidőszak mortalitásait hazánkra, ezzel ezért most nem foglalkozom. Jól pél-

dázza azonban ez a sokszerű mortalitáscsökkenés³ is, hogy mennyire váratlan javulások történnek a halandóságban. Ezen felül a nyugdíjas korosztályt is érdemes volt külön tekinteni. Tekintsük tehát külön a 45-65 és 65-80 korcsoportokra a legjobban teljesítő magyar egy- és kétpopulációs modelleket nemek szerint:



4.11. ábra. A legjobban teljesítő magyar modellek. A modell nevének első tagja jelöli, hogy egy- (szimpla) vagy többpopulációs (multi) a struktúra, ezt követi multipopulációs esetén a párpopuláció megnevezése, majd a modell alaptípusa, végül a tanulói időszak első éve. (Adatok forrás: saját szerkesztés.)

Férfiak esetén a 45-65 éves korosztályra a kétpopulációs, magyar nőekkel együttes APC modell (1985-től indítva) adta a legalacsonyabb átlagos MAPE értéket: 5,26%. Nagyságrendileg is ez egy nagyon erős modellt mutat, ekkora tévedés mortalitási modellezésnél már

³Valójában a MAPE által vett abszolút értékű eltérések miatt a becslés és a valós adat eltéréseinek iránya nem állapítható meg ezekből az ábrákból. Ahogy viszont a mellékletben is látszik, legtöbb esetben a modellek felülbecsülték a magyar mortalitásokat a 2005-2014-es időszakra.

közel kiválónak nevezhető. Az M7-es szimpla illetve osztrákkal együttes modell, továbbá az 1950-es indítású Plat struktúra szintén nagyon jó előrejelzést adtak, de a nyugdíjhoz közelebbi években az első modell messze a legjobbnak bizonyult. Ezen a szubpopuláción tehát egyértelműen sikerült javítani az előrejelzésen a kétpopulációs eljárással.

A nyugdíjas férfiak esetén 75 éves korig szintén a nőikkel vett kétpopulációs APC a legjobb MAPE alapján, de az utolsó évekre ez már nem olyan hatékony. Szemben a szlovák kétpopulációs M7-essel, mely ezen a korcsoporton 3,82% átlagos MAPE értéket adott. Még ennél is jobban teljesített az egypopulációs RH-modell, mely – mint azt korábban említettem – csak erre a populációra futott le sikeresen.

A nőknél a 45-65 éves korcsoportban végül a szimpla APC modell bizonyult a legjobbnak az átlagos 4,99%-kal. Kiemelendő, hogy a modell magyar férfiakkal együttes változata szintén hasonlóan jól teljesít. Összességében tehát a magyar késő-középkorú társadalom mortalitását legjobban az APC modellek írták le.

Végül a női nyugdíjasoknál ismét a nemzeten belüli kétpopulációs APC bizonyult a legjobb előrejelzőnek (3,82%), illetve a szlovák nőikkel közös Plat modell emelkedik még ki. Megjegyzendő, hogy ebben a csoportban meglepően jól teljesített a Poisson Lee-Carter modell is.

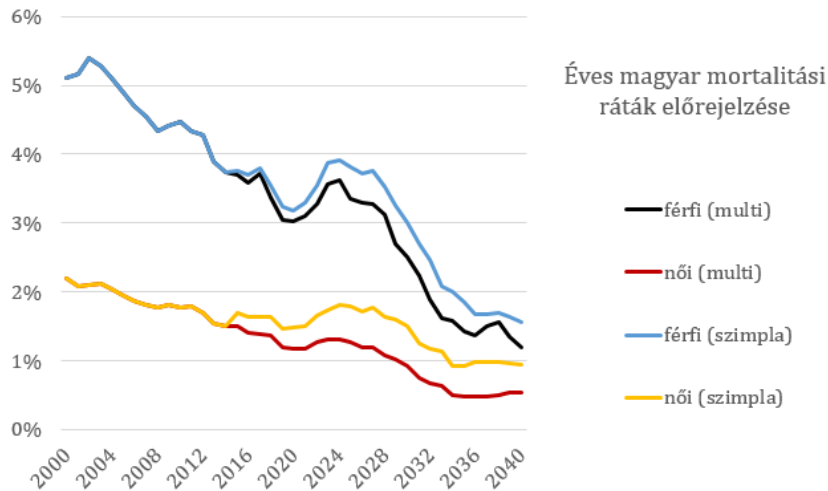
4.3.2. Magyar, osztrák és szlovák mortalitás előrejelzések 2040-ig

A legjobb modellek bemutatása után térjünk rá néhány alkalmazásra. Először is egy egyszerű előrejelzést szeretnék adni 2040-ig a magyar férfi és női társadalomra. Ezt most a 65 éves korosztályra teszem meg, összehasonlítva néhány modellt, hogy képet kapjunk ilyen perspektívából is arról, milyen fejlődés várható a következő évtizedekben a jelenlegi nyugdíjkorhatárt elérők halandóságait illetően.

Elsőként a magyar férfi és női mortalitási fejlődést szeretném bemutatni. Ehhez a legjobb közös modellt, vagyis a kétpopulációs magyar női-férfi APC modellt 1985-től indítva most egészen 2014-ig futtattam újra, hogy a legfrissebb megfigyelések információtartalma is bekerüljön az előrejelzésbe.

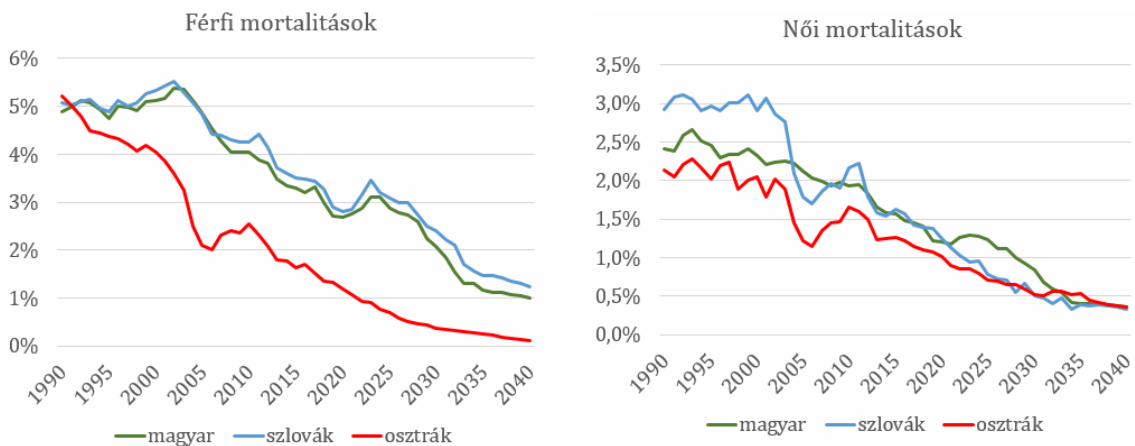
A 4.12. ábrán az első szembetűnő jelenség, hogy az egypopulációs APC-nél is egymáshoz közeledő görbéket látunk a nemek között. Tehát a többpopulációs modellezés igazán nem is azt vetíti előre, hogy gyorsabban konvergál egymáshoz két összefüggő populáció mortalitása, hanem hogy az azokban lejátszódó folyamatok kölcsönösen erősítik egymást. Ezek a folyamatok pedig a javuló trendet testesítik meg, így még annál is kedvezőbben alakul a halandóság a 65 évesek körében, mint ahogyan azt az egypopulációs modell indokolná. Az egyik legjobban működő modell, a kétpopulációs magyar férfi-női modell tehát azt mutatja, hogy a longevity kockázat nagyobb, mint azt korábban, az egypopulációs

modellkeretben gondoltuk. A következő szakaszban konkrét példán mutatom meg, ennek mekkora pénzügyi hatása lehet.



4.12. ábra. Magyar női és férfi mortalitások előrejelzése egy- és kétpopulációs APC modellekkel. (Adatok forrás: saját szerkesztés.)

A fenti ábrán még érdekesség, hogy a 2020-as évek közepén mindkét nemre emelkedés figyelhető meg. Ennek oka, hogy az 1957-1965 között születettekre nagyon magas kohorszhatás mutatkozott. Ez talán a szabadságharc utóhatásaival függ össze, de ennek megállapítása külön elemzést igényelne. Emellett a nagyságrendekről még megjegyezném, hogy a mai női mortalitási szintet 15 év múlva érik el a férfiak. Ez a fejlődés egy körülbelüli feleződést jelent a 65 évesek halandósági rátájában.



4.13. ábra. Magyar, szlovák és osztrák mortalitás előrejelzések. (saját szerkesztés)

A másik előrejelzés-vizsgálatnál a három országot nemenként szemléltetem a 4.13. ábrán. A férfiaknál a 65 éves korban kiemelkedően teljesítő magyar-szlovák M7 modellt használtam, az osztrákokra pedig a magyarokkal közös kétpopulációs Plat modellt. A nők esetében szintén a magyar-szlovák közös struktúrát (1970-es Plat modell), az osztrákokra pedig az egypopulációs Plat modellt, mely ebben a korban legjobbként teljesített.

Ahogy azt a korábbi eredmények alapján várhattuk, a férfiaknál az osztrákok nagyságrendileg más utat járnak be az előrejelzés szerint, mint Magyarország és Szlovákia. Ennek ellenére sokkal jobban megközelítjük őket abszolút mortalitásban, mint ahogy napjainkban állunk, a szlovákokkal pedig rendkívül szoros az együttmozgás. Nőknél nagyon erősen egybekonvergál a három görbe, és nem kétpopulációs technikákkal ez minden bizonnyal nem ilyen képet mutatna.

4.3.3. Életjáradék árazása és várható élettartam

A mortalitás előrejelzés számos alkalmazási területe közül még a járadékárazást, és a nyugdíjba vonuláskor – azaz 65 évesen – várható élettartamot szeretném bemutatni. Ezek kalkulálásához egészen $\omega = 110$ -ig terjedő mortalitásokra van szükség, így a modelleket újrafuttattam a bővített korhorizonton, és úgy végeztem el a becslést. Minden esetben dinamikus halandóság-előrejelzéssel dolgozom, tehát figyelembe veszem, hogy a halandósági tábla évről évre változik. Ez ellentétes, de sokkal életszerűbb megközelítés, mint a legtöbbször közölt, statikus halandósági tábla, melyre például az 1.1. ábrán láthattunk.

Az évente egy forintot fizető életjáradék egyszeri nettó díjképlete az alábbiak szerint írható fel Banyár (2003) alapján:

$$\ddot{a}_x = \sum_{i=0}^{\omega-x} \left(v^i \prod_{j=0}^{i-1} (1 - q_{x+j, T+j}) \right) \quad (x = 0, 1, \dots, \omega), \quad (4.2)$$

ahol $q_{x,t}$ az x éves egyén t . naptári évhez tartozó halálozási valószínűsége, T az aktuális naptári év, v pedig a diszkonttényező. Utóbbihoz szükséges a technikai kamatláb, ezt végig 2,3%-nak fogom feltételezni, 2016. július 1-jétől ugyanis ennyi a technikai kamatláb hazai maximális mértéke az aktuális MNB rendelet (MNB, 2015) szerint.

A férfiak esetén a kétpopulációs, szlovákokkal vett M7 modellt, illetve ennek egypopulációs megfelelőjét vizsgáltam. Nőknél az APC modell multi és szimpla változatával áraztam. Az EU gender irányelve miatt uniszex díjat számoltam külön az egy- és kétpopulációs technikával: előbbi esetben a nettó díj 12 millió 988 ezer forint lett, míg utóbbinál jóval magasabb, 14 millió 334 ezer forint. Ez a körülbelül 10%-os eltérés óriási a valóságban, ennyivel alulárzni egy terméket nagyon kockázatos.

Ha a 65 éves egyén 2014-ben várható hátralévő életkorát vizsgáljuk, ismét hasonló eredményre jutunk. A hátralévő várható életkor képlete dinamikus halandóságot feltételezve:

$$e_x = \sum_{i=1}^{\omega-x} \prod_{j=0}^{i-1} (1 - q_{x+j, T+j}) + \frac{1}{2} \quad (x = 0, 1, \dots, \omega). \quad (4.3)$$

Férfiak esetén az egypopulációs modell 13,87 évet, míg a kétpopulációs 15,12-t ad. Vagyis, ha utóbbi modellt feltételezzük a helyesnek, akkor a férfiak nyugdíjba vonuláskor várható élettartamát több, mint egy évvel alulbecsültük. Ez a nyugdíj kalkulációjakor rendkívüli hatással bíró eltérés. A nők esetén az egypopulációs APC 17 évet eredményez, míg a kétpopulációs már 19,66-ot. Egyrészt örömteli hír, hogy a mai 65 éves nők már várhatóan majdnem 85 éves korukig élnek, másrészt a nyugdíjrendszer fenntarthatósága szempontjából ez egy aggasztó eredmény.

4.4. Kitekintés és konklúzió

Összegezzük a kétpopulációs modellezés eredményeit! Férfiaknál azt kaptuk, hogy a 45-65 év közöttiek esetén a magyar nőkkel vett APC, illetve az osztrák férfiakkal képzett M7 modell javított az egypopulációs legjobb MAPE értékű modellekhez képest a 2005-2014-es tesztidőszakban. A nőkkel közös modell hatékonysága a nyugdíjas korosztályra is megmaradt, emellett itt a szlovákokkal közös M7 modell, illetve az egypopulációs RH-modell bizonyult még kiválónak.

A nőknél legjobb modell a kétpopulációs, magyar férfi populációval képzett APC lett. A szomszédos országokkal együttes modelleknél viszont eltérő képet láttunk, mint a férfiaknál: a hazai női mortalitások ezekkel a struktúrákkal a teljes korhorizonton jól szerepeltek.

Ausztria és Szlovákia esetén azt láttuk, hogy szintén érdemes alkalmazni ezeket a technikákat a magyar adatokkal együtt. A férfiaknál egy fordított logika érvényesül, vagyis a 65-80 évesekre inkább Ausztria esetén érdemes magyar kétpopulációs modellezést végezni, míg Szlovákiára a fiatalabb korosztálynál.

Az alkalmazásoknál kiderült, hogy a multipopulációs struktúra a mortalitásjavulást felerősíti, a konvergencia viszont inkább csak a nőkre áll fent a következő évtizedekben. Az egypopulációs modellek által alábecsült javuló trend azt eredményezi, hogy a nyugdíjba vonuláskor várható élettartam és az egyösszegű járadéktermékek nettó díjai is nagyobbak, mint azt korábban vártuk. Ez komoly longevity kockázat-emelkedést foglal magában, ezért mindenképpen érdemes megfontolni hasonló pénzügyi tevékenységnél a többpopulációs mortalitás modellezést.

További alkalmazásként megemlíteném még, hogy a pénzügyi piacokon léteznek már longevity risk hedge-elésére létrehozott instrumentumok. Az első ilyen volt a frappásan

q-forwardnak nevezett értékpapír, melyet az amerikai J.P. Morgan és a svájci Lucida csoport kötött 2008 januárjában. Az első longevity swap ügyletet ugyanezen év júliusában kötötte a Canada Life. Az első többpopulációs élettartam-kockázathoz kapcsolódó értékpapírt pedig a világ egyik vezető biztosítója, a Swiss Re bocsátotta ki 2010 decemberében *Kortis notes* néven. Ehhez hasonló instrumentumokból még nagyon kevés létezik, viszont a jövőben indokolható lehet ezen papírok piacának ugrásszerű növekedése, köszönhetően a multipopulációs keret jelentette összefüggő kockázatmodellezés lehetőségének.

Összefoglalás

Dolgozatomban a mortalitási modellezés alapjaitól kezdve mutattam be a multipopulációs keret elméleti háttérét. A hat standard modell kétpopulációs kiterjesztését szimmetrikus struktúrába ültetve alkalmaztam, melyhez az intuitív alapot minden esetben a kointegrációs kapcsolat biztosította: ha bizonyos populációk közti idősorok összefüggenek, akkor az ebben rejlő többletinformációt VAR vagy VECM segítségével beépíthetjük a halandósági modellbe.

Magyar, szlovák és osztrák adatokon vizsgáltam meg, hogy a valóságban ezek a struktúrák mennyivel teljesítenek jobban az egypopulációsaknál. Számos esetben jutottam arra az eredményre, hogy érdemes a többpopulációs keret nyújtotta összefüggőségi struktúrát belefoglalni a mortalitási elemzésekbe. Enélkül ugyanis jellemzően jobban felülbecsüljük a halandósági rátákat, ami fokozott longevity kockázathoz vezet. Ennek magyar vonatkozású mértékét több példán is szemléltettem.

Úgy vélem, hogy dolgozatom alapjául szolgálhat jövőbeli magyar és nemzetközi mortalitási modellezéseknek – a többpopulációs módszertannal ugyanis nemzetközi életbiztosítók, illetve hazai nyugdíjpénztárok vagy maga az állam, mint nyugdíjfolyosító is hatékonyabb előrejelzéseket készíthet, mellyel csökkentheti a javuló várható élettartamból fakadó kockázatokat. Emellett olyan pénzügyi instrumentumok létrehozásához és árazásához jelenthet alapot, melyek a magyar piacon még nem léteznek, de a hazai életbiztosítás számos szereplőjének hasznos kockázatfedezeti eszközeül szolgálhatnának.

Irodalomjegyzék

- [1] Ágoston, K. Cs. & Kovács, E. (2000). Halandósági modellek. *Aktuáriusi jegyzetek*, 3. kötet, Budapest.
- [2] Antonio, K. (2016). EAA Seminar 3. modul. *University of Amsterdam, szemináriumi jegyzetek*. Előadva: Madrid, 2016. március.
- [3] Bajkó, A., Maknics, A., Tóth, K., & Vékás, P. (2015). A magyar nyugdíjrendszer fenntarthatóságáról. *Közgazdasági Szemle*, 62(12):1229-1257.
- [4] Banyár, J. (2003). Életbiztosítás. *Aula Kiadó, Budapest*.
- [5] Baran, S., Gáll, J., Ispány, M., & Pap, G. (2007). Forecasting Hungarian mortality rates using the Lee-Carter method. *Acta Oeconomica*, 57:21-34.
- [6] Blake, D., Dowd, K. & Cairns, A. J. G. (2008). Longevity risk and the Grim Reaper's toxic tail: The survivor fan charts. *Insurance: Mathematics and Economics*, 42 (2008): 1062–1066.
- [7] Brouhns, N., Deniut, M. & Van Keilegom, I. (2005). Bootstrapping the Poisson log-bilinear model for mortality forecasting. *Scandinavian Actuarial Journal*, 2005(3):212-224.
- [8] Brouhns, N., Deniut, M. & Vermunt, J. K. (2002a). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, 31:373-393.
- [9] Brouhns, N., Deniut, M. & Vermunt, J. K. (2002b). Measuring the Longevity Risk in Mortality Projections. *Bulletin of the Swiss Association of Actuaries*, 2:105-130.
- [10] Cairns, A. J. G., Blake, D. & Dowd, K. (2006). A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *Journal of Risk and Insurance*, 73(4):687-718.

- [11] Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A. & Balevich, I. (2009). A Quantitative Comparison of Stochastic Mortality Models Using Data From England and Wales and the United States. *North American Actuarial Journal*, 13(1):1-35.
- [12] Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D. & Khalaf-Allah, M. (2011). Bayesian Stochastic Mortality Modelling for Two Populations. *AUSTIN Bulletin*, 41(1):29-59.
- [13] Cairns, A. J. G., Kallestrup-Lamb, M., Rosenskjold, C. P. T., Blake, D. & Dowd, K. (2016). Modelling Socio-Economic Differences in the Mortality of Danish Males Using a New Affluence Index. *Department of Economics and Business Economics, Aarhus University*.
- [14] Carstensen, B. (2007). Age-period-cohort models for the Lexis diagram. *Statistics in Medicine*, 26:3018-3045.
- [15] Currie, I. (2016). On fitting generalized linear and non-linear models of mortality. *Scandinavian Actuarial Journal*, 2016(4):356-383.
- [16] Dickson, D. C. M., Hardy, M. R. & Waters, H. R. (2009). *Cambridge University Press*.
- [17] Dowd, K., Blake, D., Cairns, A. J. G. & Dawson, P. (2006). Survivor Swas. *Journal of Risk and Insurance*, 73:1-17.
- [18] Fung, M. C., Peters, G. W. & Shevchenko, P. V. (2016). A unified approach to mortality modelling using state-space framework: characterisation, identification, estimation and forecasting. *Identification, Estimation and Forecasting*.
- [19] Hobcraft, J., Menken, J. & Preston, S. (1982). Age, Period and Cohort Effects in Demography: A Review. *Population Index*, 48(1):4-43.
- [20] Hunt, A. & Blake, D. (2014). A general procedure for constructing mortality models. *North American Actuarial Journal*, 18(1):116-138.
- [21] Hunt, A. & Villegas, A. (2015). Robustness and convergence in the Lee–Carter model with cohort effects. *Insurance: Mathematics and Economics*, 64:186–202.
- [22] Jarner, S. F. & Kryger, E. M. (2011). Modelling adult mortality in small populations: The SAINT Model. *ASTIN Bulletin*, 41, 377-418.
- [23] Kleinow, T. (2015). A common age effect model for the mortality of multiple populations. *Insurance: Mathematics and Economics*, 63, 147-152.

- [24] Lee, R. D. & Carter, L. R. (1992). Modelling and forecasting U. S. mortality. *Journal of the American Statistical Association*, 87:659-671.
- [25] Lee, R. (2000). The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications. *North American Actuarial Journal*, 4(1):80-93.
- [26] Lee, R. & Miller, T. (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography*, 38(4):537-549.
- [27] Li, D. X. (2000): On Default Correlation: A Copula Function Approach. *The Journal of Fixed Income*, 6, 43-54.
- [28] Li, N. & Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, 42, 575-594.
- [29] Li, J. S. H., Zhou, R. & Hardy, M. R. (2015). A step-by-step guide to building two-population stochastic mortality models. *Insurance: Mathematics and Economics*, 63, 121-134.
- [30] Lovász, E. (2011). Analysis of Finnish and Swedish mortality data with stochastic mortality models. *European Actuarial Journal*, 1(2):259-289.
- [31] Májer, I. & Kovács, E. (2011). Élettartam-kockázat – a nyugdíjrendszerre nehezedő egyik teher. *Statisztikai Szemle*, 7-8:790-812.
- [32] MNB (2015). 54/2015. (XII. 21.) MNB rendelet a technikai kamatláb maximális értékéről.
- [33] Ouburg, W. (2013). Single and Multi-Population Mortality Models for Dutch Data. *University of Amsterdam*.
- [34] Plat, R. (2009). On stochastic mortality modeling. *Insurance: Mathematics and Economics*, 45(3):393-404.
- [35] Qin, C. & Jevtic, P. (2016). Multi-Population Mortality Modelling with Lévy Processes.
- [36] 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.

- [37] Vékás, P. (2016). Az élettartam-kockázat modellezése (doktori értekezéstervezet, 178 p.). *Budapesti Corvinus Egyetem, Általános és Kvantitatív Közgazdaságtan Doktori Iskola*.
- [38] Villegas, A. M., Millossovich, P. & Kaishev, V. (2016). *StMoMo: An R Package for Stochastic Mortality Modelling*.
- [39] Wang, C.-W., Yang, S. S. & Huang, H.-C. (2015). Modeling multi-country mortality dependence and its application in pricing survivor index swaps – A dynamic copula approach. *Insurance: Mathematics and Economics*, 63:30-39.
- [40] Wilmoth, J. (1993). Computational Methods for Fitting and Extrapolating the Lee-Carter Model of Mortality Change. *Technical report. Department of Demography, University of California, Berkeley, California*.
- [41] Wilson, C., (2001). On the scale of global demographic convergence 1950–2000. *Population and Development Review* 27, 155–172.
- [42] Yang, S. S. & Wang, C. W. (2012). Pricing and securitization of multi-country longevity risk with mortality dependence. *Insurance: Mathematics and Economics*, 52, 157-169.
- [43] Zhou, R., Wang, Y., Kaufhold, K., Li, J. S. H. & Tan, K. S. (2012). Modeling period effects on multi-population mortality models: Applications to Solvency II. *North American Actuarial Journal*, 18, 150-167.

NYILATKOZAT

Név: Horváth Roland

ELTE Természettudományi Kar, szak:

Biztosítási és pénzügyi matematika MSc

Neptun azonosító: RR518B

Szakdolgozat címe:

Magyar halandósági táblák előrejelzése multipopulációs modellekkel

A **szakdolgozat** 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, 2016. 12. 31.

a hallgató aláírása