

SZAKDOLGOZAT

Halandósági előrejelzések hibái

BÉNYI GÁBOR

Témavezető: Arató Miklós
egyetemi docens
ELTE Matematikai Intézet
Valószínűségelméleti és Statisztika Tanszék



BCE-ELTE

2018

Tartalomjegyzék

Előszó	2
1. Bevezetés	3
1.1. A téma rövid ismertetése	4
1.2. Alapfogalmak	5
1.2.1. Néhány valószínűségszámítási fogalom és jelölés	5
1.2.2. A halandóság statisztikai mérőszámai	6
1.3. Lee–Carter-modell	7
2. A várható élettartam explicit meghatározása és közelítései	9
3. Numerikus elemzések	18
3.1. Magyarországi halandósági adatokon vett paraméterbecslések	18
3.2. Bázisidőszakok kiválasztása az előrevetítési torzítások figyelembe vétele mellett	18
3.3. Illeszkedésvizsgálat a teljes lakosság halálozási valószínűségeire női és férfi bontásban is	23
3.4. A módosított Lee–Carter-modell mortalitási valószínűségek becslésére	26
3.5. Az $e_{n,t}^{(0)}$, $e_{n,t}^{(1)}$ és $e_{n,t}^{(2)}$ előrevetítések megrajzolása	31
3.6. Egy- és kétfázisú élettartamok generálása	36
4. Az eredmények összegzése	41
5. Függelék	43

Előszó

Jelen dolgozat az Eötvös Loránd Tudományegyetem és a Budapesti Corvinus Egyetem által közösen indított Biztosítási és Pénzügyi Matematika mesterképzésen a Diplomamunka szeminárium 1. és a Diplomamunka szeminárium 2. nevű tárgyak keretein belül készült. Halandósági előrejelzésekkel kapcsolatos torzításokat, hibákat dolgoz fel.

Halandósági előrejelzéseket egyebek közt az emberi élettartam-kockázattal összefüggésben álló eseményeknél szükséges górcső alá venni. Például állami nyugdíjrendszerek reformintézkedéseinek a tervezéséhez hasznos, ha a szakemberek az ún. longevity-kockázatot is figyelembe veszik. Vagyis azt, hogy a később születettek élettartama a korábbi születésűekhez képest magasabb is lehet, mivel a halandósági mérőszámok tekintetében a mortalitási ráták több országban is csökkenő tendenciát mutatnak az elmúlt 50 évben. Hiszen nyilvánvaló, hogy aki többet él, annak az élete végéig számított nyugdíjjáradékok jelenértékösszege nagyobb lesz mint annak, aki kevesebbet¹, így az állami deficit kellő odafigyelés hiányában súlyosbodhat. Természetesen ehhez még hozzá kell venni a nyugdíjrendszer sok más sajátosságát is, amelyektől az meghatározódik (például tőkésített, részben tőkésített vagy folyófinanszírozású e, szolgáltatással vagy járulékbefizetéssel meghatározott e, stb.) és amelyekre a dolgozat már nem épít. Élet- és bizonyos nem-életbiztosítási termékek (például felelősségbiztosítás) művelésénél is elengedhetetlen, hogy az aktuáriusok számításba vegyék a mortalitási valószínűségek alakulását, hogy a jövőbeli eseményeket precízebben legyenek képesek modellezni és egyensúlyban tartsák a tartalékokat, ezáltal a kockázatközösségek vagyoniát biztonságosan kezeljék, nem utolsó sorban pedig a részvényesek hozamelvárásainak is eleget tegyenek.

A tét tehát nagy, ennél fogva érdemes ügyelni arra, hogy a halandósági előrejelzések kellő mértékben helyesek legyenek illetve érdemes figyelemmel kísérni azt, hogy hibás paraméterezéssel, téves előfeltételezéssel valamint a valószínűségi modell megalkotottságából adódóan az elméleti minta szóródása által mennyire fog torzítani a becslés a tényleges mortalitási adatokhoz képest. Dolgozatomban pontosan ezekre keresem a választ, röviden a halandósági előrejelzések hibáira.

Köszönettel tartozom témavezetőmnek, Arató Miklós tanár úrnak, aki szakértelmével mindig önzetlenül segítette a munkámat és mindazoknak, akik mellettem álltak és szeretetükkel a nehezebb időkben is jobb kezet nyújtottak.

¹azonos nyugdíjak esetében

1. Bevezetés

A matematikát alkalmazó tudományok az előrejelzésekhez leggyakrabban valószínűségi modelleket állítanak fel. Azt is mondhatjuk, hogy meghatározzák, melyek azok az események, amelyek együttesen eseményteret alkotnak, majd az ahhoz tartozó valószínűségi mezőt is. Az utóbbi megkonstruálása múltbeli adatok és feltételezések alapján valósul meg. Esetünkben a mennyiségek – amelyeket előre kívánunk jelezni – halálzási valószínűségek.

Adódik a kérdés, hogyha adott a modell, akkor milyen hibák merülhetnek fel a modellel kapcsolatban? Első megközelítésben a modell lehet helyes vagy sem. Ez azt jelenti, hogy a valószínűségi mezőt helyesen konstruáltuk-e meg vagy sem. Ha az előbbi eset áll fenn, akkor félrevezető lehet azt hibának tekinteni, hogy a minták önmaguk természeténél fogva szóródnak, azaz nem elfajuló eloszlásoknál a szórásnégyzet pozitív. Ugyanakkor abból a szempontból, hogy nem látunk a jövőbe, hibázási eshetőséggel is kell számolni, mert a jövő előtti döntés meghozatalánál nem lehetünk teljesen biztosak abban, hogy a várható érték szerinti becslés fog bekövetkezni. Másfelől pedig soha nem lehetünk abban bizonyosak, hogy a modell valóban helyes, még akkor sem, ha hipotézisvizsgálatok azt alátámasztják. Ezért az utóbbi megfontolás miatt a valószínűségi modellből vett szórásnégyzeteket is hibának fogjuk tekinteni. Ha pedig a modell nem helyes, adódhat például abból, hogy az eloszláscsalád paramétereit rosszul adtuk meg vagy alkalmasabb eloszláscsaládot kellett volna megválasztani.

A fenti megfontolások mellett ha körültekintően járunk el, akkor egy, a mennyiségek természetét jól leíró többváltozós, elméleti valószínűségi modellt kapunk. Azonban a gyakorlatban feltett kérdésekre a válaszok általában nem triviálisak, hiszen a modellezni kívánt mennyiségek bonyolult transzformációinak eloszlásait, várható értékeit meghatározni roppant nehéz feladat lehet. Esetünkben halálzási rátákat szeretnénk előrejelezni és azokból várható élettartamokat számolni minden lehetséges korra. A nehézséget a várható élettartam meghatározása jelenti, hiszen azok a mortalitási valószínűségek bizonyos szorzatösszegeiként kaphatók meg, a mortalitási valószínűségek pedig a Lee–Carter-modellből adódnak. A transzformáció bonyolultságának ellenére természetesen olyan alakhoz is juthatunk a meghatározandó várható élettartamoknál, amelyeket az ésszerűség határain belül pontosan ki lehet számolni. Például egy nyugdíjbiztosítási szerződés díjszabásánál fontos, hogy a szükséges formulát ne tartson napokig kiszámolni. Ellenben ha az explicit alak túlságosan bonyolult, akkor közelítő módszereket szoktak alkalmazni. Például legyen adott T a valós számok halmazán értelmezett valószínűségi változó és g pedig Borel-mérhető függvény. Ekkor a $g(T)$ változónak a várható értékét ($\mathbb{E}(g(T))$ -t) a T várható értékének a g függvénybeli értékével becslik, azaz $g(\mathbb{E}(T))$ -vel. S minthogy a szóban forgó két kifejezés általában nem azonos, ezért ez egy újabb hibázási eshetőségre adhat okot. Így a halandósági modellekből az élettartam-előrejelzés gyakorlati megvalósításainál is hibába eshetünk, ha a becslés jelentősen torzít. Érdemes tehát az ilyen

esetekben a valós és a becsült eredmények közötti különbségek érzékenységét is vizsgálni.

1.1. A téma rövid ismertetése

A dolgozatban arra keressük a választ, hogy egy általunk helyesnek tekintett Lee–Carter-moddal hogyan kaphatjuk meg magyarországi halandósági adatokon a várható élettartamokat nemenként és a teljes lakosságra nézve attól függően, hogy mennyi idős és melyik évre vonatkoztatjuk. Azokat számítógépes program segítségével számítjuk ki valamennyi korra és évre figyelembe véve azt, hogy az élettartamok explicit alakjait mennyire nehéz megadni. Hamar problémába ütközünk, mivel a várható élettartamok explicit alakjainak a kiszámítására belátható időn belül nincs lehetőség. Ugyanis születés kori esetekben legalább 10^{36} db. mennyiséget kellene összeadni, ami egy mai személyi számítógépnek végeláthatatlan összeg. Ezért használjuk a bevezetésben már említett, gyakorlatban is alkalmazott közelítő módszert, valamint a bonyolult struktúrát megpróbáljuk egyszerűbb alakra hozni. Az utóbbira saját közelítésként is hivatkozunk. Nem evidens azonban, hogy a numerikus elemzésekhez nőknél, férfiaknál és a teljes lakosságnál mely időszakot használjuk fel a paraméterbecslésekhez. A BAJKÓ–MAKNICS–TÓTH–VÉKÁS [4] cikket alapul véve nőknél az [1950, 2012]-es, a férfiaknál pedig az [1980, 2012]-es intervallumot vesszük. A teljesség igénye mellett χ^2 -próbával unisex táblára is meghatározzuk a bázisidőszakot. Legalkalmasabbnak sok szempont mellett az [1980, 2012]-es időszak tűnik. Majd a paraméterbecslések helyességét hipotézisvizsgálatokkal teszteljük, hogy a reziduálisok valóban a kívánt eloszlásokból származnak-e. Észrevesszük, hogy az individuális hibák szórásai az 50 év alattiak és az 51 év felettiak esetében elütnek, ezért bővítjük az eredeti modellt. A kibővített modellben teljes lakosságra nézve ugyancsak a legalkalmasabb bázisidőszakot keressük. A kapott χ^2 -próbát összehasonlítva az előbbivel láthatóvá válik, hogy 95%-os megbízhatósági intervallumon az elfogadási és elutasítási tartományok azonosak, ezért arra következtetünk, hogy nem szükséges bővíteni a modellt. Ezek után Monte-Carle-szimulációval kiszámítjuk a születés kori és a 65 évesek explicit várható élettartamait a 2018-as évre vonatkoztatva, amelyeket összehasonlítjuk a saját és a gyakorlatban alkalmazott becslésekkel. Az utóbbiak felül, de pontosabban becslik a tényleges várható élettartamokat, mint a saját becslések. Majd hisztogramok bemutatásával szemléltetjük, hogy 90%-os szignifikancia szinten milyen széles intervallumokban mozoghatnak a várható életek. Végül azt nézzük meg, hogy a Lee–Carter-modell mennyire torzít a paraméterbecsléseknél feltéve, hogy a modellt meghatározó paraméterek pontosan azok, mint amiket a becslésekből kaptunk. Az elvárásainknak megfelelően számos forgatókönyv legenerálásával a hisztogramokról leolvasható, hogy az átlagos torzítás a 2018-as évre nézve nem jelentős.

Összességében így két hibalehetőséggel foglalkozunk a dolgozatban. Az egyik a közelítő formulák hibája, a másik pedig az a hiba, amely a paraméterbecslésekből eredeztethető. Azt tesszük fel, hogy

egy speciális Lee–Carter-modell helyesen írja le a valóságot, tehát a modellhibák nem tárgyai a jelen munkának. Hasonlóan nem foglalkozunk a ténylegesen bekövetkező halálok számából eredő bizonytalanságokkal, csak a hátralévő élettartam becslési (előrejelzési) torzítását vizsgáljuk.

1.2. Alapfogalmak

1.2.1. Néhány valószínűségszámítási fogalom és jelölés

Jelöljön $(\Omega, \mathcal{A}, \mathbb{P})$ egy valószínűségi teret az Ω halmazon, ahol \mathcal{A} egy Ω feletti σ -algebra, \mathbb{P} pedig az \mathcal{A} -n értelmezett valószínűségi mérték. Az $\mathbf{X} : (\Omega, \mathcal{A}) \rightarrow \mathbb{R}^n$, $\mathcal{B}(\mathbb{R}^n)$ mérhető függvényt n -dimenziós valószínűségi változónak nevezzük.

Jelentse a már elemi valószínűségszámításnál megszokott jelölésekkel $F_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})$ és $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})$ az \mathbf{X} változó feltételes eloszlás- és sűrűségfüggvényét az $\mathbf{Y} = \mathbf{y}$ feltétel mellett. Abszolút folytonos eloszlásoknál a feltételes sűrűségfüggvényt a következő módon definiálják:

$$f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) = \begin{cases} \frac{f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})} \left(= \frac{\partial^2 F(\mathbf{x},\mathbf{y})/[\partial \mathbf{x} \partial \mathbf{y}]}{f_{\mathbf{Y}}(\mathbf{y})} \right) & \text{ha } f_{\mathbf{Y}}(\mathbf{y}) \neq 0 \\ 0 & \text{egyébként} \end{cases} \quad (1)$$

ahol $F_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})$ és $f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})$ az \mathbf{X} és \mathbf{Y} szerinti valószínűségi változók együttes eloszlás- és sűrűségfüggvénye. Átszorzással $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) \cdot f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})$ alakhoz jutunk. Ha mindkét oldalt \mathbf{y} -szerint integráljuk, akkor a jobb oldalon megkapjuk \mathbf{X} marginális sűrűségfüggvényét:

$$\int f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) \cdot f_{\mathbf{Y}}(\mathbf{y}) \, d\mathbf{y} = f_{\mathbf{X}}(\mathbf{x}) \quad (2)$$

Ez különösen nagy segítséget nyújt akkor, ha szeretnénk megtudni egy valószínűségi vektorváltozó (\mathbf{X}) eloszlását, amikor ismerjük valamely más változó (\mathbf{Y}) feltétele mellett annak sűrűségfüggvényét és a másik változóéna a sűrűségfüggvényét is. Számunkra a fenti összefüggés jelentős, mert a szakdolgozatban arra keressük a választ, hogy ha az egyes életévekre és évjáratokra vonatkozó halálozási valószínűségekre vannak ex post meghatározott (és számunkra igaznak vélt) eloszlások, akkor minden szükséges évjáratban megadhatjuk a születéskor, egy-, két-, hároméves stb. korban várható élettartam eloszlását (vagy sűrűségfüggvényét).

Jelöljön $\mathcal{G} \subseteq \mathcal{A}$ egy σ -algebrát. Ekkor az $\mathbb{E}(\mathbf{X}|\mathcal{G})$ -n azt a \mathcal{G} -mérhető, n -dimenziós vektorváltozót értjük, amelyre teljesül a parciális átlagolási tulajdonság, azaz

$$\mathbb{E}(\mathbb{E}(\mathbf{X}|\mathcal{G}) \mathbb{1}_{\{A\}}) = \mathbb{E}(\mathbf{X} \mathbb{1}_{\{A\}}) \quad (\forall A \in \mathcal{A}), \quad (3)$$

ahol $\mathbb{1}_{\{A\}}(\omega) = 1$, ha $\omega \in A$, egyébként 0. Megmutatható, hogy ilyen mindig létezik és 1 valószínűséggel az egyértelmű. Speciálisan

$$\mathbb{E}(h(\mathbf{X}, \mathbf{Y}) | \sigma(\mathbf{Y})) = \int h(\mathbf{x}, \mathbf{y}) f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) \, d\mathbf{x}, \quad (4)$$

ha az egyenlőségbeli feltételes sűrűségfüggvény létezik és $\mathbb{E}(|h(\mathbf{X}, \mathbf{Y})|) < \infty$.

Legyen továbbá $A \in \mathcal{A}$. Ekkor az $\mathbf{X}|A$ -en azt az n -dimenziós eloszlást értjük, amelynek eloszlásfüggvénye a következő:

$$F_{\mathbf{X}|A}(\mathbf{x}) := \frac{\mathbb{P}(\{\mathbf{X} < \mathbf{x}\} \cap A)}{\mathbb{P}(A)}. \quad (5)$$

1.2.2. A halandóság statisztikai mérőszámai

VÉKÁS [2] doktori értekezésében pontosan és világosan kitér a halandóság statisztikai mérőszámaina és matematikai modellezésére, majd a Lee–Carter-modellt alaposan részletezi. Néhányat onnan kiemelek, hogy megmutassam, a dolgozatomban mely további fogalmakat veszi vizsgálat tárgyává, valamint a jelölések egységes formában mutatkozzanak.

Tekintsünk egy N fős csoportot, amelynek minden tagja ugyanolyan idős. Feltesszük, hogy egy év múlva közülük D egyén már nem él. Ezekből az adatokból szeretnénk egy mérőszámot, ami megmutatja, hogy ugyanarra a kezdeti csoportra nézve milyen nagy a halálozás aránya. A válasz attól függ, hogy mire szeretnénk vetíteni. VÉKÁS a kitettséget (jelölje E) tekinti alapnak, azaz a csoport megélt éveinek a számát. Ha az elhunytak év végén haltak meg, akkor minden egyén ugyanabban az évben N évet élt. Ezt hívja kezdeti kitettségnek (jelölje E^0). Ekkor a halálozási valószínűség becslését az alábbi alakban írhatjuk fel:

$$m^0 = \frac{D}{E^0} \quad (6)$$

A valóságban persze kicsi az esélye annak, hogy az egyének pontosan egy év múlva, ugyanazon a napon halnak meg, ezért a megélt évek számát $E^0 - (1 - A)D$ -vel közelítik, ahol $A \in [0, 1]$ egy korrekciós paraméter. Ha $A = 1$, úgy visszkapjuk az előbbi esetet, $A = 0$ -nál pedig azt feltételezzük, hogy az egyének év elején haltak meg. A gyakorlatban inkább valószínűbbnek tűnik, hogy egyesek közelebb az év elejéhez, mások az év vége felé halnak meg, így $A = \frac{1}{2}$ egyszerűsítő feltevésről szoktak élni. Innen kapjuk a központi kitettséget (E^c -t), a központi halálozási valószínűséget (m^c -t) és az utóbbinak a kezdeti halálozási valószínűséggel való kapcsolatát:

$$m^c = \frac{D}{E^c} = \frac{D}{E^0 - \frac{1}{2}D} = \frac{\frac{D}{E^0}}{1 - \frac{1}{2}\frac{D}{E^0}} = \frac{m^0}{1 - \frac{1}{2}m^0}. \quad (7)$$

Dolgozatomban a kezdeti kitettségre vonatkozó mortalitási rátára² fogom illeszteni a Lee–Carter-modellt. Ez a módszer az élettartam-előrejelzéseket bemutató tanulmányokkal, sőt az eredeti Lee–Carter szerzőpáros által publikált cikkel szemben nem a központi halálozási rátát veszi alapul. Ennek az az oka, hogy a dolgozatomban jelentősen összetettebbé válna, a várható élettartam kiszámításánál további approximáció szükségeltetne.

²Szokták nyers halálozási valószínűségnek is nevezni.

1.3. Lee–Carter-modell

A nagyközönség számára a Központi Statisztikai Hivatal és a Human Mortality Database online felületéről halálozási ráták is elérhetőek. Bár magyarországi vonatkozásban 1950-től 2014-ig állnak rendelkezésre nyers halálozási ráták, mi csak a 2012-es évig bezárólag fogjuk venni azokat, hogy a BAJKÓ–MAKNICS–TÓTH–VÉKÁS [4] cikkel összhangban végezhesünk elemzéseket. Természetesen több cikk is foglalkozott korábban magyarországi alkalmazásokkal. A teljesség igénye nélkül megemlíthetjük BARAN–GÁLL–ISPÁNY–PAP [7] cikket, ahol a 2004 és 2040 közötti időszakon végeztek előrejelzéseket a halálozási valószínűségekre az 1949-től 2003-ig rendelkezésre álló mortalitási adatok birtokában.

$q_{x,t}$ -vel fogjuk jelölni a nyers halálozási valószínűségeket, ahol t és x azt mutatja meg, hogy az mely évhez és korcsoporthoz tartozik ($t \in \{1950, 1951, \dots, 2012\}$, $x \in \{1, 2, \dots, \omega\}$) és ω az elérhető legnagyobb életkor.³ A modell a következő alakban írható fel (a dolgozatomban kezdeti kitettségi halálozási rátára):

$$\ln(q_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t}, \quad (8)$$

ahol a k_t folyamatról feltesszük, hogy driftes, véletlenbolyongásos folyamatot követ:

$$k_t = k_{t-1} + c + Z_t. \quad (9)$$

$\varepsilon_{x,t}$ és Z_t független, normális eloszlásból származó, 0 várható értékű valószínűségi változók σ_ε^2 és σ_Z^2 szórásnégyzettel $\forall x, t$ -re, c pedig egy konstans.⁴ Az identifikációs probléma kiküszöböléséhez további két megkötést is alkalmaznak:

$$\sum_{x=0}^{\omega} b_x = 1, \quad \text{és} \quad \sum_{t=1950}^{2012} k_t = 0. \quad (10)$$

Ugyanis az $\mathbf{a} = (a_x)_{x \in \{0,1,\dots,\omega\}}$, $\mathbf{b} = (b_x)_{x \in \{0,1,\dots,\omega\}}$, $\mathbf{k} = (k_t)_{t \in \{1950,\dots,2012\}}$ vektorparaméterek esetén az $\tilde{\mathbf{a}} = \mathbf{a} + \alpha \mathbf{b}$, $\tilde{\mathbf{b}} = \frac{1}{\beta} \mathbf{b}$, $\tilde{\mathbf{k}} = \beta(\mathbf{k} - \alpha \mathbf{1})$ paraméterek is kielégítik a (8) egyenletet, ahol $\mathbf{1}$ a csupa 1-ből álló, $(2012 - 1950 + 1)$ -dimenziós vektor, $\alpha, \beta \in \mathbb{R}$, $\beta \neq 0$. Legyen $T = 2012 - 1950 + 1$ és $\mathbf{M} = \left[\ln(q_{x,t}) - \frac{1}{T} \sum_{t=1950}^{2012} \ln(q_{x,t}) \right]_{(\omega+1) \times (T)}$ -dimenziós mátrix. Ekkor bizonyítás nélkül az

³VÉKÁS [2] a modell felépítésének a margójára megjegyzi, hogy a modell általánosabb keretek között is alkalmazható (például ha x és t nem éves, hanem attól eltérő felosztású, azonos hosszúságú, egymást követő, diszjunkt időintervallumok).

⁴Valójában az eredeti Lee–Carter-modell feltételei között nem szerepel a k_t folyamat ilyen alakja és az $\varepsilon_{x,t}$ normalitása sem. A dolgozathoz elvégzett számításokat azonban nagyban megkönnyítik ezek – a szakirodalomban gyakran alkalmazott – feltételek. A modell idősoros bővítéseiről kapcsolatosan rövid leírást a LEE [6] cikkben is olvashatunk.

egyres paraméterek maximum likelihood becllései az alábbiak:^{5,6}

$$(\hat{\mathbf{a}})_x = \frac{1}{T} \sum_{t=1950}^{2012} \ln(q_{x,t}), \quad \hat{\mathbf{b}} = \mathbf{M}\mathbf{M}^\top \text{ normált domináns sajátvektora}, \quad \hat{\mathbf{k}} = \mathbf{M}^\top \hat{\mathbf{b}} \quad (11)$$

$$\hat{c} = \frac{\hat{\mathbf{k}}_T - \hat{\mathbf{k}}_1}{T-1}, \quad \hat{\sigma}_Z^2 = \frac{1}{T-1} \sum_{t=1950}^{2011} \left(\hat{\mathbf{k}}_{t+1} - (\hat{\mathbf{k}})_t - \hat{c} \right)^2 \quad (12)$$

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{T(\omega+1)} \sum_{x=0}^{\omega} \sum_{t=1950}^{2012} \left(\ln(q_{x,t}) - (\hat{\mathbf{a}})_x - (\hat{\mathbf{b}})_x (\hat{\mathbf{k}})_t \right)^2 \quad (13)$$

⁵Valamely mátrix domináns sajátvektora alatt a legnagyobb sajátértékhez tartozó sajátvektort értjük.

⁶Euklideszi térben valamely \mathbf{a} vektor normáltján az $\mathbf{a}/\|\mathbf{a}\|$ vektort értjük.

2. A várható élettartam explicit meghatározása és közelítései

Mi 2012 utáni életbenmaradási eloszlásokat kívánunk megadni és az ebben rejlő „hibákat”. Jelenleg – mint ahogyan arról már korábban írtunk – 2014-ig visszamenőleg tudunk aggregált halálozási adatokkal foglalkozni, ezért számunkra az a releváns kérdés, hogy 2014 után milyen halálozásokat tudunk megbecsülni. Jelölje $\xi_{(2012+t)}^{(n)}$ azt a valószínűségi változót, amelynek értéke azt mutatja meg, hogy $(2012+t)$ -ben az n éves egyén még hány évet fog élni. Nyilván $\xi_{(2012+t)}^{(n)} \in \{0, 1, 2, \dots, \omega - n\}$. Ennek speciális esete, amikor $n = 0$, azaz a születéskor várható élettévek eloszlásaira vagyunk kíváncsiak.

Ha az egyén elérte az $(2012+t+k)$ évet, akkor tegyük fel, hogy az $(n+k)$ éves egyén $p_{(n+k, 2012+t+k)}$ valószínűséggel éli túl az $(2012+t+k)$ évet és lép a $(2012+t+k+1)$ évbe, illetve $q_{(n+k, 2012+t+k)} := 1 - p_{(n+k, 2012+t+k)}$ valószínűséggel hal meg $\forall k \in \{0, 1, \dots, \omega - n\}$ -ra. Ebből könnyen megállapíthatjuk $\xi_{(2012+t)}^{(n)}$ eloszlását:

$$\begin{aligned}
 & \bullet \left[\begin{array}{l} \mathbb{P}(\text{Feltéve, hogy az egyén } n \text{ éves és nem éli túl a következőt}) = \\ \mathbb{P}(\xi_{(2012+t)}^{(n)} = 0) = q_{(n+0, 2012+t+0)} \end{array} \right. \\
 & \bullet \left[\begin{array}{l} \mathbb{P}(\text{Feltéve, hogy az egyén } n \text{ éves és megéri az } n+1 \text{ éves kort, de abban az évben meghal}) = \\ \mathbb{P}(\xi_{(2012+t)}^{(n)} = 1) = q_{(n+1, 2012+t+1)} \cdot p_{(n+0, 2012+t+0)} \end{array} \right. \\
 & \bullet \left[\begin{array}{l} \mathbb{P}(\text{Feltéve, hogy az egyén } n \text{ éves és megéri az } n+2 \text{ éves kort, de abban az évben meghal}) = \\ \mathbb{P}(\xi_{(2012+t)}^{(n)} = 2) = q_{(n+2, 2012+t+2)} \cdot p_{(n+1, 2012+t+1)} \cdot p_{(n+0, 2012+t+0)} \end{array} \right. \\
 & \vdots \\
 & \bullet \left[\begin{array}{l} \mathbb{P}(\text{Feltéve, hogy az egyén } n \text{ éves és megéri az } n+k \text{ éves kort, de abban az évben meghal}) = \\ \mathbb{P}(\xi_{(2012+t)}^{(n)} = k) = q_{(n+k, 2012+t+k)} \cdot \prod_{j=0}^{k-1} p_{(n+j, 2012+t+j)} \end{array} \right. \\
 & \vdots \\
 & \bullet \left[\begin{array}{l} \mathbb{P}(\text{Feltéve, hogy az egyén } n \text{ éves és megéri az } \omega \text{ éves kort, de abban az évben meghal}) = \\ \mathbb{P}(\xi_{(2012+t)}^{(n)} = \omega - n) = q_{(\omega, 2012+t+\omega-n)} \cdot \prod_{j=0}^{\omega-n-1} p_{(n+j, 2012+t+j)} \end{array} \right.
 \end{aligned}$$

Természetesen $q_{(\omega, 2012+t+\omega-n)} = 1$, hiszen ω -ra úgy tekintünk, mint a lehetséges legnagyobb évre, amit már senki nem él túl. Láthatóan az utóbbi megkötéssel egy jól definiált valószínűségi változót kapunk, mert

$$q_{(n, 2012+t)} + \sum_{i=1}^{\omega-n} q_{(n+i, 2012+t+i)} \cdot \left(\prod_{j=0}^{i-1} p_{(n+j, 2012+t+j)} \right) = 1. \quad (14)$$

Sajnálatos módon a 2014 utáni évekből már nem állnak rendelkezésre a megfelelő ún. kohorsz valószínűségi adatok, hiszen a jelenleg Magyarországon működő, hivatalos statisztikai adatokat publikáló szerv, a Központi Statisztikai Hivatal (röv. KSH) oldalán is csak 2014-ig visszamenőleg tudunk mortalitási adatokat letölteni. Ennek hiányában felvetődik a kérdés, hogy akkor mégis milyen adatokkal dolgozzunk. A kérdés tulajdonképpen magában foglalja a választ: modellekkel jelezzük azokat előre és adjunk azokra becsléseket. További megfontolást igényel az, hogy vajon a konkrét becslést adjuk meg a halandósági valószínűségekre, azaz „konkrétakká” tegyük-e a $q_{(n+i,2012+t+j)}$ paraméterértékeket és így $\xi_{(2012+t)}^{(n)}$ eloszlása is egyszerű számolással adódna vagy igyekezzünk a kissé rögzösebb, ámde pontosabb utat választani, ahol egy összetett problémára keressük a választ: mi $\xi_{(2012+t)}^{(n)}$ eloszlása, ha a mortalitási ráták is valószínűségi változók. Megjegyezzük, hogy az előbbi módszer egy (természetes) tipikus becslési hibához vezet, amely áldozatot a feladat egyszerűsítése érdekében alkalmaznak. Ezek után már világos, hogy tulajdonképpen $q_{(n+i,2012+t+j)}$ -k maguk is valószínűségi változók, amelyeknél az összetévesztést elkerülendő a továbbiakban $q_{(n+i,2012+t+j)}$ egy konkrét (0-1 közötti) értéket fog jelölni, míg $Q_{(n+i,2012+t+j)}$ magát a valószínűségi változót, amelyet a halandósági előrejelzések modelljeivel kapunk. Természetesen más-más modellek eltérő eredményekhez vezethetnek, ezért a szóban forgó halálózási valószínűségek is eltérő valószínűségi változók lehetnek.

Visszatérve az eredeti problémához várható élettartamok eloszlására keressük a választ. Ha $q_{(n+i,2012+t+j)}$ -k minden $i, j \in \mathbb{N}$ -re ismert, akkor

$$\begin{aligned} \mathbb{E} \left(\xi_{(2012+t)}^{(n)} \middle| Q_{(n,2012+t)} = q_{(n,2012+t)}, \dots, Q_{(\omega,2012+t+\omega-n)} = q_{(\omega,2012+t+\omega-n)} \right) = \\ = \underbrace{(0 \cdot q_{(\omega,2012+t+\omega-n)})}_{=0} + \sum_{i=1}^{\omega-n} i \cdot q_{(n+i,2012+t+i)} \cdot \left(\prod_{j=0}^{i-1} P_{(n+j,2012+t+j)} \right), \end{aligned} \quad (15)$$

ha pedig nem, akkor a következőt kapjuk:

$$\begin{aligned} \mathbb{E} \left(\xi_{(2012+t)}^{(n)} \middle| Q_{(n,2012+t)}, Q_{(n+1,2012+t+1)}, \dots, Q_{(\omega,2012+t+\omega-n)} \right) = \\ = \underbrace{(0 \cdot Q_{(\omega,2012+t+\omega-n)})}_{=0} + \sum_{i=1}^{\omega-n} i \cdot Q_{(n+i,2012+t+i)} \cdot \left(\prod_{j=0}^{i-1} P_{(n+j,2012+t+j)} \right), \end{aligned} \quad (16)$$

ahol $P_{(n+j,2012+t+j)} = 1 - Q_{(n+j,2012+t+j)}$. Mindezek után természetes az az igény, ha a fenti kifejezés várható értékét (nevezetesen $\xi_{(2012+t)}^{(n)}$ -nek a várható értékét a valószínűségi számításból már jól ismert várhatóérték-tétel miatt) keressük azon megkötés mellett, hogy az 1950-től 2012-ig terjedő időszakban minden életkorra ismertek a (tapasztalati) mortalitási valószínűségek ($q_{x,u}$) és a mortalitási indexek (k_u) valamint elfogadjuk azt, hogy nemenként egy, az 1950-es és 2012-es évek között alkalmasnak választott bázisidőszakon végzett Lee–Carter-modell paraméterbecslései fogják a halálózási valószínűségeket meghatározni. Mielőtt a megfelelő bázisidőszakot kiválasztanám,

egyelőre az egyszerűség kedvéért azon az $[1950, 2012]$ -es intervallum által határolt évek halmazát fogom érteni. Formálisabban jelentse \mathcal{F}_{2012} azt az információmennyiséget, amellyel a rendelkezésünkre álló mortalitási adatok birtokában az 1950 és 2012 közötti halálozási valószínűségeket és a becsléssel kapott mortalitási indexeket rögzítjük:

$$Q_{(x,u)} = q_{(x,u)} \quad \text{és} \quad K_u = k_u \quad (\forall x \in \{0, 1, 2, \dots, \omega\} \forall u \in \{1950, 1951, \dots, 2012\}) \quad (17)$$

A fenti feltétel nem elhanyagolható, minekután azok egymástól nem független mennyiségek. Külön hangsúlyozom, hogy \mathcal{F}_{2012} feltételrendszer, megkötések halmaza. A szakdolgozat egyik lényegi része, hogy ezt a várható értéket $\left(\mathbb{E}\left(\xi_{(2012+t)}^{(n)} \mid \mathcal{F}_{2012}\right)\right)$ -t meghatározzam a Lee–Carter-modell segítségével. Lássuk, végül aritmetikai átalakításokkal milyen, számunkra releváns alakhoz jutunk.

Felhasználjuk, hogy

$$\mathbb{E}\left(\xi_{(2012+t)}^{(n)} \mid \mathcal{F}_{2012}\right) = \mathbb{E}\left(\mathbb{E}\left(\xi_{(2012+t)}^{(n)} \mid Q_{(n,2012+t)}, Q_{(n+1,2012+t+1)}, \dots, Q_{(\omega,2012+t+\omega-n)}, \mathcal{F}_{2012}\right) \mid \mathcal{F}_{2012}\right), \quad (18)$$

ami a várhatóérték-tételből követlenül adódik. Most már minden szükséges eszköz rendelkezésünkre áll $\xi_{(2012+t)}^{(n)}$ -élettartamváltozó várható értékének a kiszámításához az \mathcal{F}_{2012} -információt felhasználva:

$$\begin{aligned} \mathbb{E}\left(\xi_{(2012+t)}^{(n)} \mid \mathcal{F}_{2012}\right) &= \mathbb{E}\left(\mathbb{E}\left(\xi_{(2012+t)}^{(n)} \mid Q_{(n,2012+t)}, \dots, Q_{(\omega,2012+t+\omega-n)}, \mathcal{F}_{2012}\right) \mid \mathcal{F}_{2012}\right) \\ &= \mathbb{E}\left(\sum_{i=1}^{\omega-n} i \cdot Q_{(n+i,2012+t+i)} \cdot \left(\prod_{j=0}^{i-1} P_{(n+j,2012+t+j)}\right) \mid \mathcal{F}_{2012}\right) \\ &= \sum_{i=1}^{\omega-n} i \cdot \mathbb{E}\left(Q_{(n+i,2012+t+i)} \cdot \left(\prod_{j=0}^{i-1} P_{(n+j,2012+t+j)}\right) \mid \mathcal{F}_{2012}\right) \\ &= \sum_{i=1}^{\omega-n} i \cdot \mathbb{E}\left(\exp(a_{n+i} + b_{n+i}K_{2012+t+i} + \varepsilon_{(n+i,2012+t+i)}) \cdot \left(\prod_{j=0}^{i-1} (1 - \exp(a_{n+j} + b_{n+j}K_{2012+t+j} + \varepsilon_{(n+j,2012+t+j)}))\right) \mid \mathcal{F}_{2012}\right), \end{aligned} \quad (19)$$

ahol a_{n+j} és b_{n+j} a Lee–Carter-modellből kapott, rögzített paraméterek, $K_{2012+t+j}$ és $\varepsilon_{(n+j,2012+t+j)}$ pedig valószínűségi változók, amelyekből csak az utóbbi független az \mathcal{F}_{2012} -információtól. Az egyszerűség kedvéért legyen $g(i) := a_{n+i} + b_{n+i}K_{2012+t+i} + \varepsilon_{(n+i,2012+t+i)}$. Ekkor – mivel K_u véletlenbolyongásos driftelt folyamatot követ, – $K_{2012+t+j}$ kifejezhető K_{2012} és attól független, azonos eloszlású valószínűségi változók összegeként:

$$K_{2012+t+j} = K_{2012+t+j-1} + Z_{2012+t+j-1} + c = \dots = K_{2012} + \sum_{s=0}^{t+j-1} Z_{2012+s} + (t+j)c, \quad (20)$$

ahol c a (9) egyenletből származó paraméter. Így

$$g(i) = a_{n+i} + b_{n+i} \left(K_{2012} + \sum_{s=0}^{t+i-1} Z_{2012+s} + (t+i)c \right) + \varepsilon_{(n+i, 2012+t+i)}. \quad (21)$$

Az egyszerűsített alakot felhasználva tehát tovább írhatjuk az egyenlőséget:

$$\begin{aligned} & \mathbb{E} \left(\xi_{(2012+t)}^{(n)} \middle| \mathcal{F}_{2012} \right) \\ &= \sum_{i=1}^{\omega-n} i \cdot \mathbb{E} \left(\exp(g(i)) \cdot \prod_{j=0}^{i-1} (1 - \exp(g(j))) \middle| \mathcal{F}_{2012} \right) \\ &= \sum_{i=1}^{\omega-n} i \cdot \mathbb{E} \left(\exp(g(i)) \right. \\ & \quad \cdot \left[1 - \left(\sum_{0 \leq j_1 \leq i-1} \exp(g(j_1)) \right) + \dots + (-1)^i \left(\sum_{0 \leq j_1 < \dots < j_i \leq i-1} \exp(g(j_1) + \dots + g(j_i)) \right) \right] \middle| \mathcal{F}_{2012} \right) \\ &= \sum_{i=1}^{\omega-n} i \cdot \mathbb{E}(\exp(g(i)) | \mathcal{F}_{2012}) + i \sum_{k=1}^i (-1)^k \left[\sum_{0 \leq j_1 < \dots < j_k \leq i-1} \mathbb{E} \left(\exp \left(g(i) + \sum_{l=1}^k g(j_l) \right) \middle| \mathcal{F}_{2012} \right) \right] \end{aligned} \quad (22)$$

Hogy nélkülözsem a terjengősséget, célszerűnek tartom most csak az exponenciális kitevőjében lévő kifejezéseket kibontani:

$$\begin{aligned} g(i) + \sum_{l=1}^k g(j_l) &= \left[a_{n+i} + b_{n+i} \left(K_{2012} + \sum_{s=0}^{t+i-1} Z_{2012+s} + (t+i)c \right) + \varepsilon_{(n+i, 2012+t+i)} \right] \\ & \quad + \left[\sum_{l=1}^k a_{n+j_l} + b_{n+j_l} \left(K_{2012} + \sum_{s=0}^{t+j_l-1} Z_{2012+s} + (t+j_l)c \right) + \varepsilon_{(n+j_l, 2012+t+j_l)} \right] \\ &= \left[a_{n+i} + \sum_{l=1}^k a_{n+j_l} \right] + \left[\left(b_{n+i} + \sum_{l=1}^k b_{n+j_l} \right) K_{2012} \right] \\ & \quad + \left[\left((t+i)b_{n+i} + \sum_{l=1}^k (t+j_l)b_{n+j_l} \right) c \right] \\ & \quad + \left[\varepsilon_{(n+i, 2012+t+i)} + \sum_{l=1}^k \varepsilon_{(n+j_l, 2012+t+j_l)} \right] \\ & \quad + \left[\sum_{s=0}^{t+j_1-1} (b_{n+j_1} + \dots + b_{n+j_k} + b_{n+i}) Z_{2012+s} + \dots \right. \\ & \quad \quad \left. + \sum_{s=t+j_k-1}^{t+j_k-1} (b_{n+j_k} + b_{n+i}) Z_{2012+s} + \sum_{s=t+j_k}^{t+i-1} b_{n+i} Z_{2012+s} \right] \end{aligned} \quad (23)$$

Nem nehéz látni, hogy a Z_{2012+s} változók függetlenek az F_{2012} -es feltételtől (mivel véletlen bolyongást követ K_u ($u \in \mathbb{Z}$) és K_{2012} -t pedig helyettesíthetjük a feltételbeli k_{2012} -vel). Vagyis

ellegendő az egymástól is független Z_{2012+s} ($\sim \mathcal{N}(0, \sigma_Z^2)$) és $\varepsilon_{(n+j_i, 2012+t+j_j)}$ ($\sim \mathcal{N}(0, \sigma_\varepsilon^2)$) valószínűségi változókból képzett bonyolult kifejezés várható értékét kiszámolni, ahol σ_Z^2 és σ_ε^2 a modellből adóan ismertek. Mielőtt annak nekifutnánk, érdemesnek tartom kimeleni, hogy egy $\mathcal{N}(\mu, \sigma^2)$ eloszlásból származó Y valószínűségi változó esetében $\exp(Y)$ -nak mi a várható értéke:

$$\mathbb{E}(\exp(Y)) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \quad (24)$$

Nem kis erőfeszítés árán látható tehát $\mathbb{E}\left(\xi_{(2012+t)}^{(n)} \middle| \mathcal{F}_{2012}\right)$ explicit alakja, amelyre a következő fejezetekben $e_{n,t}^{(0)}$ -ként is hivatkozunk:

$$\begin{aligned} & \mathbb{E}\left(\xi_{(2012+t)}^{(n)} \middle| \mathcal{F}_{2012}\right) \\ &= \sum_{i=1}^{\omega-n} i \cdot \left(\exp\left(a_{n+i} + b_{n+i}(i+t)c + b_{n+i}k_{2012} + \frac{\sigma_\varepsilon^2}{2} + (i+t)b_{n+i}^2 \frac{\sigma_Z^2}{2}\right) \right. \\ & \quad + \sum_{k=1}^i (-1)^k \sum_{0 \leq j_1 < \dots < j_k \leq i-1} \exp\left(\left[a_{n+i} + \sum_{l=1}^k a_{n+j_l}\right] + \left[\left(b_{n+i} + \sum_{l=1}^k b_{n+j_l}\right)k_{2012}\right] \right. \\ & \quad \quad \quad \left. + \left[\left((t+i)b_{n+i} + \sum_{l=1}^k (t+j_l)b_{n+j_l}\right)c\right] + (k+1)\frac{\sigma_\varepsilon^2}{2} \right. \\ & \quad \left. + \frac{\sigma_Z^2}{2} \left[(j_1+t) \left(b_{n+i} + \sum_{l=1}^k b_{n+j_l}\right)^2 + (j_2-j_1) \left(b_{n+i} + \sum_{l=2}^k b_{n+j_l}\right)^2 + \dots + (i-j_k)b_{n+i}^2 \right] \right) \\ & = : e_{n,t}^{(0)} \end{aligned} \quad (25)$$

Rendkívül bonyolult, sőt számításigényes, hiszen a fenti egyenletben található k index $n = 0$ esetben az $\omega/2$ értéket többször is eléri, ezért legalább $\binom{\omega}{\lfloor \omega/2 \rfloor}$ (ha $\omega = 100$, akkor kb. 10^{36}) db. tagot kell összeadni, ami még egy felhasználói szintű személyi számítógépnek is végeláthatatlan összeg. Természetesen a célunk az, hogy használhatóbb alakra hozzuk a kapott kifejezést. Jelöljük a $\frac{\sigma_Z^2}{2}$ -hez tartozó együtthatót $\beta_{\sigma_Z^2/2}$ -vel és alakítsuk át azt:

$$\begin{aligned}
\beta_{\sigma^2/2} &= (j_1 + t) \left(b_{n+i} + \sum_{l=1}^k b_{n+j_l} \right)^2 \\
&+ (j_2 - j_1) \left(b_{n+i} + \sum_{l=2}^k b_{n+j_l} \right)^2 + \cdots + (j_k - j_{k-1}) (b_{n+i} + b_{n+k})^2 \\
&+ (i - j_k) b_{n+i}^2 = (j_1 + t) \left[\left(\sum_{l=1}^k b_{n+j_l} \right)^2 + 2b_{n+i} \sum_{l=1}^k b_{n+j_l} + b_{n+i}^2 \right] \\
&+ (j_2 - j_1) \left[\left(\sum_{l=2}^k b_{n+j_l} \right)^2 + 2b_{n+i} \sum_{l=2}^k b_{n+j_l} + b_{n+i}^2 \right] + \cdots \\
&+ (j_k - j_{k-1}) [b_{n+j_k}^2 + 2b_{n+i} b_{n+j_k} + b_{n+i}^2] + (i - j_k) b_{n+i}^2 = (i + t) b_{n+i}^2 \\
&+ 2b_{n+i} \sum_{l=1}^k (j_l + t) b_{n+j_l} + \left[(j_1 + t) \left(\sum_{l=1}^k b_{n+j_l} \right)^2 + (j_2 - j_1) \left(\sum_{l=2}^k b_{n+j_l} \right)^2 \right. \\
&\left. + \cdots + (j_k - j_{k-1}) b_{n+j_k}^2 \right] = (i + t) b_{n+i}^2 + 2b_{n+i} \sum_{l=1}^k (j_l + t) b_{n+j_l} \\
&+ \left[(j_1 + t) \left(\sum_{l=1}^k b_{n+j_l}^2 + \sum_{1 \leq l_1 < l_2 \leq k} 2b_{n+j_{l_1}} b_{n+j_{l_2}} \right) \right. \\
&+ (j_2 - j_1) \left(\sum_{l=2}^k b_{n+j_l}^2 + \sum_{2 \leq l_1 < l_2 \leq k} 2b_{n+j_{l_1}} b_{n+j_{l_2}} \right) + \cdots \\
&\left. + (j_k - j_{k-1}) b_{n+k}^2 \right] = (i + t) b_{n+i}^2 + 2b_{n+i} \sum_{l=1}^k (j_l + t) b_{n+j_l} \\
&+ \left[\sum_{l=1}^k (j_l + t) b_{n+j_l}^2 + (j_1 + t) \left(\sum_{1 \leq l_1 < l_2 \leq k} 2b_{n+j_{l_1}} b_{n+j_{l_2}} \right) \right. \\
&+ (j_2 - j_1) \left(\sum_{2 \leq l_1 < l_2 \leq k} 2b_{n+j_{l_1}} b_{n+j_{l_2}} \right) + \cdots \\
&\left. + (j_{k-1} - j_{k-2}) b_{n+k} b_{n+k-1} \right] = (i + t) b_{n+i}^2 + 2b_{n+i} \sum_{l=1}^k (j_l + t) b_{n+j_l} \\
&+ \sum_{l=1}^k (j_l + t) b_{n+j_l}^2 + \sum_{l=1}^{k-1} 2(j_l + t) b_{n+j_l} \left(\sum_{l < l_1 \leq k} b_{n+j_{l_1}} \right) = (i + t) b_{n+i}^2 \\
&+ \underbrace{\sum_{l=1}^k (j_l + t) b_{n+j_l} (2b_{n+i} + b_{n+j_l}) + \sum_{l=1}^{k-1} 2(j_l + t) b_{n+j_l} \left(\sum_{l < l_1 \leq k} b_{n+j_{l_1}} \right)}_{=: \tau}
\end{aligned} \tag{26}$$

Az átrendezés után kapott összeg utolsó tagját (ami ugyancsak egy összeg) jelölje τ , amelyben látható, hogy a futóindex (l) csak $k - 1$ -ig tart. Ezt külön kezeljük. Végül $\beta_{\sigma^2/2}$ átalakítás utáni

alakját beleillesztjük az $\mathbb{E} \left(\xi_{(2012+t)}^{(n)} \middle| \mathcal{F}_{2012} \right)$ explicit alakjába és egy újabb lépésben egy futóindex alá rendezzük azokat (kivéve τ -t):

$$\begin{aligned}
& \mathbb{E} \left(\xi_{(2012+t)}^{(n)} \middle| \mathcal{F}_{2012} \right) \\
&= \sum_{i=1}^{\omega-n} i \cdot \left(\exp \left(a_{n+i} + b_{n+i}(i+t)c + b_{n+i}k_{2012} + \frac{\sigma_{\varepsilon}^2}{2} + (i+t)b_{n+i}^2 \frac{\sigma_Z^2}{2} \right) \right. \\
&\quad + \sum_{k=1}^i (-1)^k \sum_{0 \leq j_1 < \dots < j_k \leq i-1} \exp \left(\left[a_{n+i} + \sum_{l=1}^k a_{n+j_l} \right] + \left[\left(b_{n+i} + \sum_{l=1}^k b_{n+j_l} \right) k_{2012} \right] \right. \\
&\quad \quad \quad \left. + \left[\left((t+i)b_{n+i} + \sum_{l=1}^k (t+j_l)b_{n+j_l} \right) c \right] + (k+1) \frac{\sigma_{\varepsilon}^2}{2} \right. \\
&\quad \quad \quad \left. \left. + \frac{\sigma_Z^2}{2} \left[(i+t)b_{n+i}^2 + \sum_{l=1}^k (j_l+t)b_{n+j_l} (2b_{n+i} + b_{n+j_l}) + \tau \right] \right) \right) \\
&= \sum_{i=1}^{\omega-n} i \cdot \left(\exp \left(a_{n+i} + b_{n+i}(i+t)c + b_{n+i}k_{2012} + \frac{\sigma_{\varepsilon}^2}{2} + (i+t)b_{n+i}^2 \frac{\sigma_Z^2}{2} \right) \right. \\
&\quad + \sum_{k=1}^i (-1)^k \sum_{0 \leq j_1 < \dots < j_k \leq i-1} \exp \left(\left[a_{n+i} + b_{n+i}(i+t)c + b_{n+i}k_{2012} + \frac{\sigma_{\varepsilon}^2}{2} + (i+t)b_{n+i}^2 \frac{\sigma_Z^2}{2} \right] \right. \\
&\quad \left. + \sum_{l=1}^k \left[a_{n+j_l} + b_{n+j_l}k_{2012} + (t+j_l)b_{n+j_l}c + \frac{\sigma_{\varepsilon}^2}{2} + \frac{\sigma_Z^2}{2} ((j_l+t)b_{n+j_l} (2b_{n+i} + b_{n+j_l})) \right] + \tau \right) \Bigg) \\
&\hspace{20em} (27)
\end{aligned}$$

Ha τ -t figyelmen kívül hagyjuk, akkor meglehetősen szerencsések vagyunk, hiszen ha a korábbi $g(i)$ és $g(j)$ helyett a $g^*(i) := a_{n+i} + b_{n+i}(i+t)c + b_{n+i}k_{2012} + \frac{\sigma_{\varepsilon}^2}{2} + (i+t)b_{n+i}^2 \frac{\sigma_Z^2}{2}$ és $g^*(j) := a_{n+j} + b_{n+j}k_{2012} + (t+j)b_{n+j}c + \frac{\sigma_{\varepsilon}^2}{2} + \frac{\sigma_Z^2}{2} ((j+t)b_{n+j} (2b_{n+i} + b_{n+j}))$ szereposztásokat tekintjük, akkor hasonló kifejezéshez jutunk, mint amit a legelején vehettünk szemügyre, de azzal nem azonos:

$$\begin{aligned}
\mathbb{E} \left(\xi_{(2012+t)}^{(n)} \middle| \mathcal{F}_{2012} \right) &\approx \sum_{i=1}^{\omega-n} i \cdot \exp(g^*(i)) \cdot \prod_{j=0}^{i-1} (1 - \exp(g^*(j))) \\
&= \sum_{i=1}^{\omega-n} i \cdot \exp \left(a_{n+i} + b_{n+i}(i+t)c + b_{n+i}k_{2012} + \frac{\sigma_{\varepsilon}^2}{2} + (i+t)b_{n+i}^2 \frac{\sigma_Z^2}{2} \right) \\
&\quad \cdot \prod_{j=0}^{i-1} \left(1 - \exp \left(a_{n+j} + b_{n+j}k_{2012} + (t+j)b_{n+j}c + \frac{\sigma_{\varepsilon}^2}{2} \right. \right. \\
&\quad \quad \quad \left. \left. + \frac{\sigma_Z^2}{2} ((j+t)b_{n+j} (2b_{n+i} + b_{n+j})) \right) \right) \\
&=: e_{n,t}^{(1)}
\end{aligned} \tag{28}$$

A főnti becslésre a továbbiakban úgy fogok hivatkozni, mint saját közelítésre. A gyakorlatban a következő módszert alkalmazzák. Ha valamely $\mathbf{T} \in \mathbb{R}^n$ valószínűségi vektorváltozó eloszlása és

várható értéke ismert, akkor \mathbf{T} valamely $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ Borel-mérhető függvény általi képének a várható értékét \mathbf{T} várható értékének h -beli értékével, azaz:

$$\mathbb{E}(h(\mathbf{T})) \approx h(\mathbb{E}(\mathbf{T}))-vel$$

becslik feltéve, hogy a h általi torzítás nem jelentős. Tekintsük a következő $\mathbf{g}_n = (g(0), g(1), \dots, g(n))$ vektorváltozót és a $h_n(x_0, x_1, \dots, x_n) := \exp(x_n) \cdot \prod_{j=0}^{n-1} (1 - \exp(x_j))$ függvényt majd használjuk az előbb részletezett módszert az általunk keresett várható élettartam becslésére:

$$\begin{aligned} \mathbb{E} \left(\xi_{(2012+t)}^{(n)} \middle| \mathcal{F}_{2012} \right) &= \sum_{i=1}^{\omega-n} i \cdot \mathbb{E} \left(\exp(g(i)) \cdot \prod_{j=0}^{i-1} (1 - \exp(g(j))) \middle| \mathcal{F}_{2012} \right) \\ &= \sum_{i=1}^{\omega-n} i \cdot \mathbb{E} (h_i(\mathbf{g}_i) | \mathcal{F}_{2012}) \\ &\approx \sum_{i=1}^{\omega-n} i \cdot h_i(\mathbb{E}(\mathbf{g}_i | \mathcal{F}_{2012})) \\ &= \sum_{i=1}^{\omega-n} i \cdot \exp(\mathbb{E}(g(i) | \mathcal{F}_{2012})) \cdot \prod_{j=0}^{i-1} (1 - \exp(\mathbb{E}(g(j) | \mathcal{F}_{2012}))) \\ &= \sum_{i=1}^{\omega-n} i \cdot \exp \left(\mathbb{E} \left(a_{n+i} + b_{n+i} \left(K_{2012} + \sum_{s=0}^{t+i-1} Z_{2012+s} + (t+i)c \right) + \varepsilon_{(n+i, 2012+t+i)} \middle| \mathcal{F}_{2012} \right) \right) \\ &\quad \cdot \prod_{j=0}^{i-1} \left(1 - \exp \left(\mathbb{E} \left(a_{n+j} + b_{n+j} \left(K_{2012} + \sum_{s=0}^{t+j-1} Z_{2012+s} + (t+j)c \right) + \varepsilon_{(n+j, 2012+t+j)} \middle| \mathcal{F}_{2012} \right) \right) \right) \\ &= \sum_{i=1}^{\omega-n} i \cdot \exp(a_{n+i} + b_{n+i}(k_{2012} + (t+i)c)) \cdot \prod_{j=0}^{i-1} (1 - \exp(a_{n+j} + b_{n+j}(k_{2012} + (t+j)c))) \\ &=: e_{n,t}^{(2)} \end{aligned} \tag{29}$$

Adott tehát egy explicit alak $(e_{n,t}^{(0)})$ és annak két közelítése $(e_{n,t}^{(1)}$ és $e_{n,t}^{(2)})$. Ránézésre nem egyszerű megállapítani, hogy a három kifejezés milyen relációban áll egymással, ezért érdemes segítségül hívni a számítógépes programozást, hogy valamennyi évre (t -re) és korra (n -re) kiszámoljuk azokat. Az utóbbi két közelítés legfeljebb párszáz összeadást és szorzást tartalmaz, ezért azokat könnyen megkaphatjuk. Mint ahogyan azt már korábban is említettem, a pontos $e_{n,t}^{(0)}$ kifejezést egy közönséges számítógépen ma még lehetetlen kiszámolni, főleg ha születéskori várható élettartamot szeretnénk. Járható útnak a Monte-Carlo-szimuláció mutatkozik. A (8) és a (9) egyenletekből származó, $(2(\omega+1)+3)$ -dimenziós $(\hat{\mathbf{a}}^\top, \hat{\mathbf{b}}^\top, \hat{c}, \hat{\sigma}_Z^2, \hat{\sigma}_\varepsilon^2)$ vektorparaméterrel tehát generáljunk egymástól független, kellően nagy, N darabszámú mintát a Lee-Carter-modellben részletezett valószínűségi változókra. Az r . replikációnál realizáljuk az egyes változókat:

$$Z_{2012}^{(r)} = z_{2012}^{(r)}, Z_{2013}^{(r)} = z_{2013}^{(r)}, \dots, Z_{2012+t+\omega-n-1}^{(r)} = z_{2012+t+\omega-n-1}^{(r)} \quad \left(Z_{2012+i}^{(r)} \sim \mathcal{N}(0, \hat{\sigma}_Z^2) \forall r \in \mathbb{N} \right) \tag{30}$$

$$\varepsilon_{(n,2012+t)}^{(r)} = \varepsilon_{(n,2012+t)}^{(r)}, \dots, \varepsilon_{(\omega,2012+t+\omega-n)}^{(r)} = \varepsilon_{(\omega,2012+t+\omega-n)}^{(r)} \quad \left(\varepsilon_{(n+i,2012+t+i)}^{(r)} \sim \mathcal{N}(0, \hat{\sigma}_\varepsilon^2) \forall r \in \mathbb{N} \right) \quad (31)$$

Ekkor a halálozási valószínűségek pontos értékei meghatározhatók a (8) és a (21) egyenletekből:

$$q_{(n+i,2012+t+i)}^{(r)} = \exp \left(\hat{\mathbf{a}}_{n+i} + \hat{\mathbf{b}}_{n+i} \left(k_{2012} + \sum_{s=0}^{t+i-1} z_{2012+s}^{(r)} + (t+i)\hat{c} \right) + \varepsilon_{(n+i,2012+t+i)} \right) \quad (32)$$

A keresett várható élettartamot az r . replikációnál a (15) formulából kapjuk meg:

$$\begin{aligned} \mathbb{E} \left(\xi_{(2012+t)}^{(n,r)} \middle| Q_{(n,2012+t)}^{(r)}, \dots, Q_{(\omega,2012+t+\omega-n)}^{(r)} = q_{(\omega,2012+t+\omega-n)}^{(r)}, \mathcal{F}_{2012} \right) &= \\ &= \sum_{i=1}^{\omega-n} i \cdot q_{(n+i,2012+t+i)}^{(r)} \cdot \left(\prod_{j=1}^{i-1} \left(1 - q_{(n+j,2012+t+j)}^{(r)} \right) \right) \end{aligned} \quad (33)$$

Az egyszerűség kedvéért a (33) egyenlet bal oldalát röviden $\mathbb{E} \left(\xi_{(2012+t)}^{(n,r)} \middle| \mathbf{Q}^{(r)} = \mathbf{q}^{(r)}, \mathcal{F}_{2012} \right)$ -val jelöljük. Minthogy $\left\{ \mathbb{E} \left(\xi_{(2012+t)}^{(n,r)} \middle| \mathbf{Q}^{(r)}, \mathcal{F}_{2012} \right), r \in \{1, 2, \dots, N\} \right\}$ egymástól független, azonos eloszlású valószínűségi változókból képzett minta $\mathbb{E} \left(\mathbb{E} \left(\xi_{(2012+t)}^{(n,r)} \middle| \mathbf{Q}^{(r)}, \mathcal{F}_{2012} \right) \right) = \mathbb{E} \left(\xi_{(2012+t)}^{(n)} \middle| \mathcal{F}_{2012} \right)$ véges várható értékkel, ezért a mintából képzett átlag a nagy számok erős törvény miatt 1 valószínűséggel konvergál a várható értékhez, vagyis

$$\mathbb{P} \left(\frac{\sum_{r=1}^N \mathbb{E} \left(\xi_{(2012+t)}^{(n,r)} \middle| \mathbf{Q}^{(r)}, \mathcal{F}_{2012} \right)}{N} \rightarrow \mathbb{E} \left(\xi_{(2012+t)}^{(n)} \middle| \mathcal{F}_{2012} \right) \right) = 1. \quad (34)$$

Így az $\left\{ \mathbb{E} \left(\xi_{(2012+t)}^{(n,r)} \middle| \mathbf{Q}^{(r)} = \mathbf{q}^{(r)}, \mathcal{F}_{2012} \right), r \in \{1, 2, \dots, N\} \right\}$ halmaznak az átlaga jó közelítéssel megadja a keresett várható élettartamot.

3. Numerikus elemzések

3.1. Magyarországi halandósági adatokon vett paraméterbecslések

Az élettartam előrejelzések matematikai módszertanának ismertetése után bemutatjuk az előbbi fejezetben részletezett halálozási ráták várható időbeli alakulását magyarországi vonatkozásban. A Human Mortality Database online felületén sikeres regisztráció után lehetőség van arra, hogy nem szerinti bontásban külön nőkre és férfiakra, valamint együtt az egész lakosságra nézve nyers halálozási valószínűségeket letöltsünk. Ezen adatoknak a numerikus elemzését az R-Studio integrált fejlesztői környezetben hajtjuk végre részint saját készítésű, másfelől R-beli függvények segítségével. Az utóbbak egy része a 2017.10.01-én letöltött programban már eleve megtalálható, a többit a demography-programcsomagból használjuk. A szóban forgó csomag telepítése után lehetővé válik, hogy az lca-függvény meghívásával az egyes mortalitási valószínűségekre vonatkozó paraméterbecsléseket megkapjuk anélkül, hogy erre saját függvényt kellene készíteni. Az lca-függvény „output”-ja egy ún. lca-objektum, amely tartalmazza egyebek közt a (11), (12) és a (13) formulákkal kapható $\hat{\mathbf{a}}, \hat{\mathbf{b}} \in \mathbb{R}^{\omega+1}, \hat{\mathbf{k}} \in \mathbb{R}^T$ paramétereket és az $\hat{\varepsilon}_{(x,t)} = \ln(q_{x,t}) - (\hat{\mathbf{a}})_x - (\hat{\mathbf{b}})_x(\hat{\mathbf{k}})_t$ reziduálisokat is. Természetesen lehetőség van a $\hat{\mathbf{k}}$ mortalitási index kiigazítására, ha az lca-függvény argumentumában a Lee–Carter-nél alkalmazott kiigazítási módszert⁷ adjuk meg (alap esetben is így számol) vagy ha mégsem kérjük, akkor az `adjust="none"` opciót kell megadni. Mi az utóbbit fogjuk használni, hogy a (1.3) fejezetben foglaltak szerint járjunk el.

KOVÁCS–MÁJER [3] tanulmányában a paraméterek szemléletes jelentőségéről is olvashatunk. $\hat{\mathbf{a}}$ egy, a korcsoportokat jellemző tag a (8) egyenletben. Annálfogva, hogy a (1.3) fejezetben tárgyalt megkötések a becslésekre is teljesülnek, ezért $\sum_{x=0}^{\omega} (\hat{\mathbf{b}})_x = 1$, és $\sum_{t=1950}^{2012} (\hat{\mathbf{k}})_t = 0$. Így a (8) egyenletben a $(\hat{\mathbf{b}})_x(\hat{\mathbf{k}})_t$ tag korrigálja az elsőt. Azt mutatja meg, hogy az időtől függő változót milyen súllyal vegyük bele a halálozási valószínűség természetes alapú logaritmusának a becslésénél.

3.2. Bázisidőszakok kiválasztása az előrejelzési torzítások figyelembe vétele mellett

Sajnos nem triviális, hogy a birtokunkban lévő adathalmazt hogyan használjuk fel a Lee–Carter-modell felállításához. Ugyanis nem szükségszerű kizárólag években történt változásokban gondolkodni, vehetünk az éves tartamú rendszerektől eltérő felosztású, azonos tartamú, diszjunkt, egymást

⁷ KOVÁCS–MÁJER [3] tanulmányban kifejtik, hogy érdemes a becsült és a modellezett halálozások számát egyenlővé tenni, mivel az eredeti modellben a fiatalok és az időskori halandósági valószínűségeket ugyanolyan hangsúlyt kapnak, holott az előbbiek lényegesen kisebb mértékben járulnak az összes halálozás számához. Vagyis ha $\tilde{\mathbf{k}}$ jelöli a kiigazított mortalitási indexeket, akkor azok egyértelműen meghatározhatók a $\sum_{x=0}^{\omega} D_{x,t} = \sum_{x=0}^{\omega} E_{x,t} \exp\left((\hat{\mathbf{a}})_x + (\hat{\mathbf{b}})_x(\tilde{\mathbf{k}})_t\right)$ ($t \in \{1950, \dots, 2012\}$) egyenletekből.

követő időszakokat is. Mivel 1950-től 2012-ig állnak rendelkezésre mortalitási adatok, ezért az előbbi feltételnek a $A_1 = \{t_1, t_1+1, t_1+2, \dots, \tilde{t}_1\}, \dots, A_k = \{t_k, t_k+1, t_k+2, \dots, \tilde{t}_k\}$ struktúrájú időhalmazok tesznek eleget, ahol $t_i, \tilde{t}_i \in \{1950, 1951, \dots, 2012\}$, $t_i < \tilde{t}_i < \tilde{t}_i + 1 = t_{i+1} < \tilde{t}_{i+1}$ ($\forall 1 \leq i \leq k$) és $|A_1| = |A_2| = \dots = |A_k|$. A BAJKÓ–MAKNICS–TÓTH–VÉKÁS [4] cikkben ugyanezt a problémát járták körbe a szerzők. A teljes adatbázist időhorizont alapján partíciókra bontották 10 éves lépésközökkel, így 1960–2000, 1970–2000, 1980–2000 és 1989–2000 intervallumba eső évek szerinti bázisidőszakokat vettek nő–férfi bontásban. Az utolsónál a rendszerváltás utáni időszak a gazdasági transzformáció okozta hatásokat már kisebb súllyal tartalmazza. Következő lépésben meghatározták a Lee–Carter-paramétereket, majd a kapott valószínűségi modellel várható értékben a 2001-es, 2002-es, ..., 2012-es évekre vonatkozó alternatív halálozási valószínűségeket és azokat összevették a tényleges realizációval. Az összehasonlítás alapja az a feltételezés, hogy a t . időszakban az x éves egyének egymástól függetlenül $q_{x,t}$ valószínűséggel halnak meg. $E_{x,t}^0$ kezdeti kitettséggű csoportnál ez azt jelenti, hogy a halálozások száma binomiális eloszlást követ $E_{x,t}^0$ és $q_{x,t}$ paraméterekkel, azaz $D_{x,t} \sim \text{Binom}(E_{x,t}^0, q_{x,t})$ ($\forall x \in \{0, 1, \dots, \omega\}$, $\forall t \in \{2001, 2002, \dots, 2012\}$). Ha feltesszük, hogy az egyes életkorokhoz tartozó kezdeti kitettségek kellően nagyok, akkor alkalmazható a De-Moivre-Laplace-tétel, miszerint $X \sim \text{Binom}(N, p)$ esetén

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(a \leq \frac{X - Np}{\sqrt{Np(1-p)}} < b \right) = \Phi(b) - \Phi(a) \quad (\forall a, b \in \mathbb{R}, a < b) \quad (35)$$

ahol Φ a standard normális eloszlásfüggvény, $p \in (0, 1)$. Esetünkben tehát a

$$\frac{D_{x,t} - E_{x,t}^0 q_{x,t}}{\sqrt{E_{x,t}^0 q_{x,t} (1 - q_{x,t})}} \sim \mathcal{N}(0, 1) \quad (36)$$

feltételezéssel élünk. Ebből minden évre χ^2 -próbát konstruálhatunk. $H_0 : q_{x,t} = \hat{q}_{x,t}$ nullhipotézis fennállásánál így

$$\chi^2 = \sum_{x=0}^{\omega} \frac{(D_{x,t} - E_{x,t}^0 \hat{q}_{x,t})^2}{E_{x,t}^0 \hat{q}_{x,t} (1 - \hat{q}_{x,t})} \sim \chi^2(\omega + 1) \quad (\forall t \in \{2001, 2002, \dots, 2012\}). \quad (37)$$

A képletben szereplő $D_{x,t}$ és $E_{x,t}^0$ a ténylegesen realizált halálozási és nyers kitettségi mutatók, amelyeket összevették az alternatív halálozási valószínűségekkkel, azokat pedig a Lee–Carter-modellből várható értékben lehet kinyerni. Pontosabban

$$\hat{q}_{x,t} := \mathbb{E}(Q_{x,t} | \mathcal{F}) \quad (38)$$

attól függően, hogy mely bázisidőszakra (\mathcal{F}) alkalmazták azokat. Bár a cikkben arról nem olvashatunk, hogy milyen széles kritikus tartományt jelöltek ki a modell validálásához, de úgy talál-ták, hogy a nőknél az összes bázisidőszakból képzett Lee–Carter-modell alkalmasnak bizonyult a 2001–2012-es évek előrejelzéseihez, míg a férfiaknál csak a 1980–2000 és az 1989–2000 időszak lett

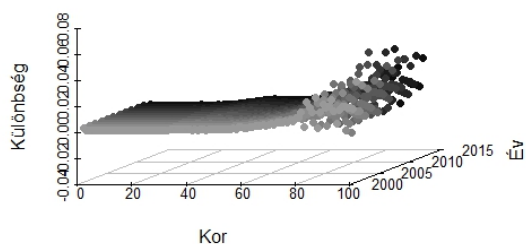
elfogadható. Ennélőva a tanulmány arra az álláspontra jutott, hogy a nőknél az 1950–2012-es, a férfiaknál pedig az 1980–2012-es időszakot érdemes a paraméterbecsléshez felhasználni.

A teljes lakosság halálozási valószínűségeire nézve viszont nem végeztek előrevetítéseket. A teljesség igényének figyelembevétele mellett elvégezzük azt. Szerencsére az alternatív várható valószínűségek kiszámítása kevesebb technikai akadályt jelent, hiszen a korábbi felálláshoz képest nem kell a mortalitási rátákat bonyolult módon transzformálni. Az előző módszerhez hasonlóan vesszük a korábban részletezett bázisidőszakokat és az azokhoz tartozó információkat, amelyeket rendre $F_{2000}^{50}, F_{2000}^{60}, F_{2000}^{70}, F_{2000}^{80}, F_{2000}^{89}$ fog jelölni, majd minden $F_{2000}^{(i)}$ -hez ($i \in \{50, 60, 70, 80, 89\}$) meghatározzuk a Lee–Carter-paramétereket $(\hat{\mathbf{a}}^\top, \hat{\mathbf{b}}^\top, \hat{c}, \hat{\sigma}_Z^2, \hat{\sigma}_\varepsilon^2)$. Az egyszerűség kedvéért az i -indexet a paramétereknél elhagyjuk. Ha $t > 2000$, akkor

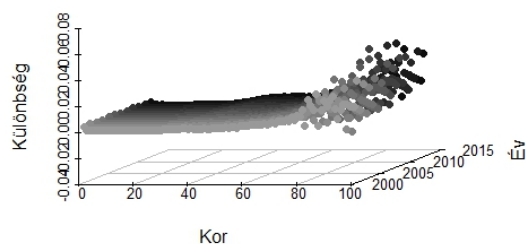
$$\begin{aligned}
\hat{q}_{x,t}^{(i)} &= \mathbb{E} \left(Q_{x,t} \mid F_{2000}^{(i)} \right) \\
&= \mathbb{E} \left(\exp \left((\hat{\mathbf{a}})_x + (\hat{\mathbf{b}})_x K_t + \varepsilon_{(x,t)} \right) \mid F_{2000}^{(i)} \right) \\
&= \mathbb{E} \left(\exp \left((\hat{\mathbf{a}})_x + (\hat{\mathbf{b}})_x \left(K_{2000} + \sum_{s=0}^{(t-1)-2000} Z_{2000+s} + (t-2000)\hat{c} \right) + \varepsilon_{(x,t)} \right) \mid F_{2000}^{(i)} \right) \\
&= \mathbb{E} \left(\exp \left((\hat{\mathbf{a}})_x + (\hat{\mathbf{b}})_x \left((\hat{\mathbf{k}})_{2000} + \sum_{s=0}^{(t-1)-2000} Z_{2000+s} + (t-2000)\hat{c} \right) + \varepsilon_{(x,t)} \right) \right) \\
&= \exp \left((\hat{\mathbf{a}})_x + (\hat{\mathbf{b}})_x \left((\hat{\mathbf{k}})_{2000} + (t-2000)\hat{c} \right) \right) \mathbb{E} \left(\exp \left(\sum_{s=0}^{(t-1)-2000} (\hat{\mathbf{b}})_x Z_{2000+s} + \varepsilon_{(x,t)} \right) \right) \\
&= \exp \left((\hat{\mathbf{a}})_x + (\hat{\mathbf{b}})_x \left((\hat{\mathbf{k}})_{2000} + (t-2000)\hat{c} \right) \right) \prod_{s=0}^{(t-1)-2000} \mathbb{E} \left(\exp \left((\hat{\mathbf{b}})_x Z_{2000+s} \right) \right) \mathbb{E} \left(\exp \left(\varepsilon_{(x,t)} \right) \right) \\
&= \exp \left((\hat{\mathbf{a}})_x + (\hat{\mathbf{b}})_x \left((\hat{\mathbf{k}})_{2000} + (t-2000)\hat{c} \right) + (t-2000) \frac{(\hat{\mathbf{b}})_x^2 \hat{\sigma}_Z^2}{2} + \frac{\hat{\sigma}_\varepsilon^2}{2} \right),
\end{aligned} \tag{39}$$

ahol felhasználtuk a 2000-ig rendelkezésre álló információkat, a Z_{2000+s} valószínűségi változók egymástól és $\varepsilon_{x,t}$ -től való függetlenségét és a (24) egyenletet. Mielőtt alkalmaznánk a χ^2 -statisztikát, érdemes megnézni, hogy a halálozási valószínűségek előrevetítései az egyes bázisidőszakokban mennyire robusztusak. Ezt az alábbi öt ábrán szemléltetjük.

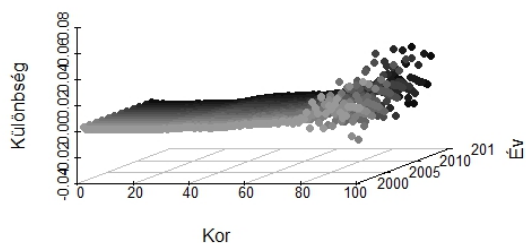
1950-2000-es bázisidőszakra vonatkozó eltérések



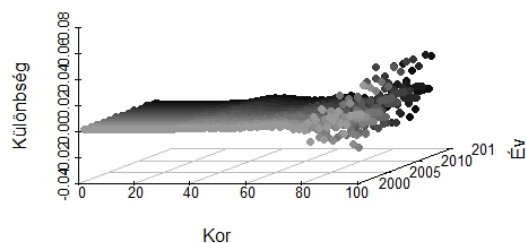
1960-2000-es bázisidőszakra vonatkozó eltérések



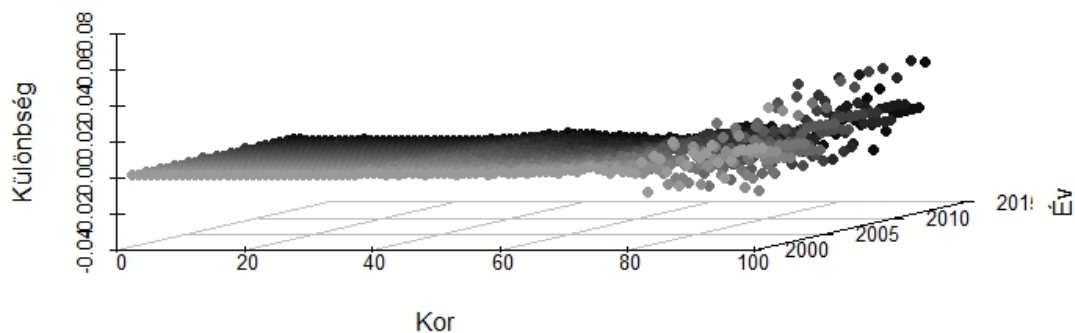
1970-2000-es bázisidőszakra vonatkozó eltérések



1980-2000-es bázisidőszakra vonatkozó eltérések

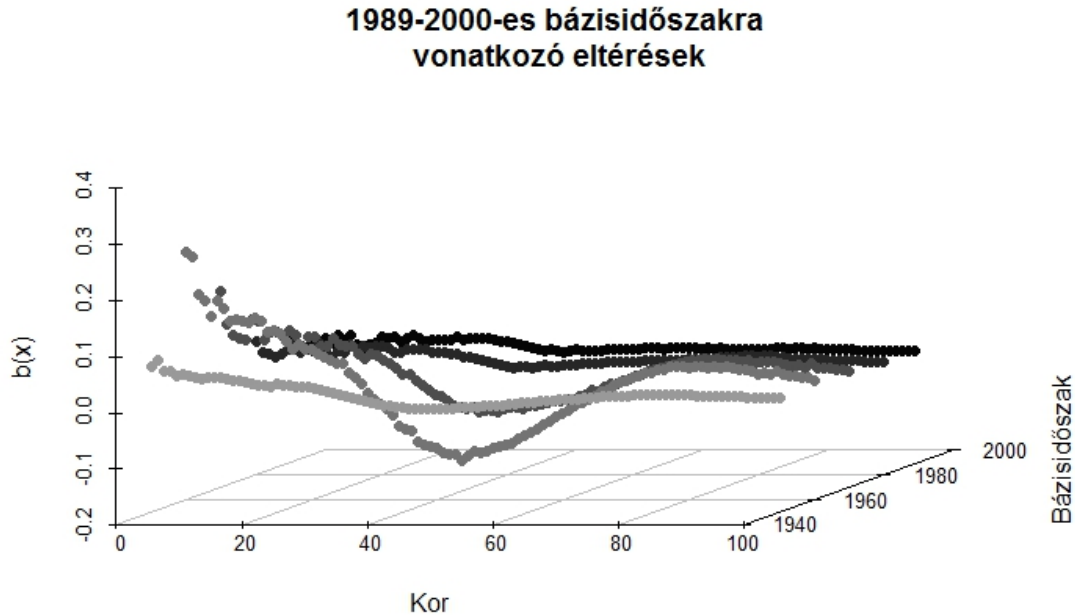


1989-2000-es bázisidőszakra vonatkozó eltérések



Mindegyik ábrán egy-egy bázisidőszakra vonatkozóan láthatjuk korhoz és a 2001, 2002, ..., 2012-es évekhez rendeltén a becslt és a tényleges halálozási valószínűségek különbségét, azaz $\hat{q}_{x,t}^{(i)} - \tilde{q}_{x,t}^{(i)}$ -ket. Az ábrákról leolvasható, hogy túlnyomórészt az előrejelzett értékek meghaladják a ténylegeseket. Ezt részben az is okozza, hogy a (39) egyenlet jobb oldalán az exponenciális kifejezés értékét növelik a $(t - 2000) \frac{(\hat{\mathbf{b}}_x)^2 \hat{\sigma}_Z^2}{2}$ és $\frac{\hat{\sigma}_\varepsilon^2}{2}$ tagok. Azok akkor csak akkor lennének 0-k, ha a Z_{2000+s} és $\varepsilon_{(x,t)}$ változók szórásai zérusak lennének, azaz 1 valószínűséggel determinisztikus értéket vennének fel,

ami éppen a várható értékük, vagyis 0. Mindebből az következik, hogy az előrejelzett mortalitási ráták a ténylegesekhez képest magasabbak, másképpen az életbenmaradási valószínűségeket alulbecsülém, így várható hátralévő élettartamokat is. Továbbá mindegyik bázisidőszakra egyöntetűen igaz, hogy a fiatalabb korosztály halálozási rátáit nagyobb sikerrel találtuk el, mint az idősebbekét. Ez például arra vezethető vissza, hogy a fiatalabb korban élők populációja nagyobb, mint az idősebbeké, ezért várható értékhez való konvergencia feltétele kevésbé fog sérülni a fiatalabbaknál. Természetesen arra is gondolhatunk, hogy az idősebbek között nagyobb súllyal szerepelnek a (39) egyenletben a $(\hat{\mathbf{b}})_x$ paraméterek, amelyek kiemelnék $(\hat{\mathbf{k}})_t$ hozzájárulását korrigálva az $(\hat{\mathbf{a}})_x$ -t az exponensen belül. Ezt a lehetőséget elvethetjük, ugyanis a következő ábrán meggyőződhetünk arról, hogy leginkább a 25-éves kor alatt veszi fel $(\hat{\mathbf{b}})_x$ a legmagasabb értékeket, 25 és 75 éves korok között pedig a legalacsonyabbakat. A 75 éves kor feletti értékek bár magasak, de általában nem érik el a fiatakkoriakét.



Az előbb említett esetek miatt a χ^2 statisztikát a magasabb korosztály fogja jelentősen növelni. Átalakítjuk ugyanis a (37) kifejezést alkalmazva a (6) egyenlőséget:

$$\begin{aligned}
 \chi^2 &= \sum_{x=0}^{\omega} \frac{(D_{x,t} - E_{x,t}^0 \hat{q}_{x,t})^2}{E_{x,t}^0 \hat{q}_{x,t} (1 - \hat{q}_{x,t})} \\
 &= \sum_{x=0}^{\omega} \frac{(E_{x,t}^0 \tilde{q}_{x,t} - E_{x,t}^0 \hat{q}_{x,t})^2}{E_{x,t}^0 \hat{q}_{x,t} (1 - \hat{q}_{x,t})} \\
 &= \sum_{x=0}^{\omega} \frac{E_{x,t}^0 (\tilde{q}_{x,t} - \hat{q}_{x,t})^2}{\hat{q}_{x,t} (1 - \hat{q}_{x,t})},
 \end{aligned} \tag{40}$$

ahol $\tilde{q}_{x,t}$ a tényleges halálozási ráta. χ^2 annál nagyobb, minél nagyobb a számlálójában a négyzetes szorzó, vagyis az előrevetített és a tényleges valószínűség különbségének a négyzete. Ezt elkerülendő a szummát 25 éves korban elvágjuk, tehát az összegzés 0-tól tart ω helyett 25-ig. Így a χ^2 statisztikák 2-tizedes jegyre kerekítve a következők:

χ^2	'50–2000	'60–2000	'70–2000	'80–2000	'89–2000
2001	58,59	144,34	92,85	34,00	32,58
2002	67,08	143,41	81,47	18,76	24,07
2003	55,95	145,76	83,75	23,97	34,69
2004	53,00	159,70	89,76	23,38	38,07
2005	68,25	173,03	98,79	39,41	66,61
2006	90,49	197,74	115,14	42,24	55,18
2007	59,72	167,24	93,17	41,10	73,42
2008	51,45	174,34	99,90	51,42	89,04
2009	68,00	206,53	119,04	52,03	89,41
2010	71,98	244,20	151,47	60,00	76,69
2011	52,90	211,86	117,22	42,22	85,81
2012	59,28	213,13	123,22	63,86	112,07
szumma	757,69	2181,27	1265,79	492,38	777,64

A 26 és a $12 \times 26 = 312$ szabadságfokokhoz tartozó χ^2 -eloszlás 95%-os kvantilisei 38,88 és 354,19, amelyek 5%-os szignifikanciaszinten nézve éppen a kritikus értékeket adják meg. Az utóbbi a „szumma”-val jelölt sorban található értékekkel összehasonlítható. Pirossal a kritikus értékeket meghaladó statisztikákat jelöltük. A táblázat alapján elmondható, hogy csak az 1980–2000-es és az 1989–2000-es bázisidőszakok esetén tudjuk elfogadni a nullhipotézist, és azokat sem az összes előrevetített évekre, hanem csak 2001-től 2004-ig.

A megfigyelések után végül úgy határozunk, hogy a BAJKÓ–MAKNICS–TÓTH–VÉKÁS [4] tanulmányban javasoltaknak megfelelően a nőknél az 1950–2012-es, a férfiaknál pedig az 1980–2012-es időszakot, míg a teljes lakosságra vett előrejelzésekhez azt a leghosszabb időszakot veszem, amelyben a legtöbb elfogadási tartományban lévő χ^2 statisztika szerepel, ami éppen az 1980–2012-es intervallum.

3.3. Illeszkedésvizsgálat a teljes lakosság halálozási valószínűségeire női és férfi bontásban is

Egy modell jóságát nem csak az adja, hogy az előrejelet értékek a realizált értékektől mennyire térnek el. Az előbbi fejezet pontosan erről szólt. A mintát kétfelé osztottuk. Az első feléből

kiszámoltuk a modell paramétereit maximum likelihood elvvel, majd előrevetítéseket végezve a másik felét a várttal összehasonlítottuk. Ez a stratégia azért volt számunkra jelentős, mert sikerült elkészíteni az optimális intervallumokat, amelyekre később a modellt építeni szeretnénk. A modellben szereplő valószínűségi változókra vonatkozó illeszkedésvizsgálat is a modell megfelelőségét mutatja. Visszatekintve a (8) és a (9) egyenletekre látható, miként konstruálhatók meg az $\varepsilon(x, t)$ és a Z_t egymástól független, valószínűségi változók, mint reziduálisok. Szeretnénk megtudni, hogy azok valóban normális eloszlásból származnak-e 0 várható értékkel és $\hat{\sigma}_\varepsilon^2$ illetve $\hat{\sigma}_Z^2$ varianciával. A fentiekre vonatkozó hipotézisvizsgálatnál körültekintően kell eljárni, ugyanis ha vesszük például a

$$\left\{ \hat{Z}_t = z_t = \hat{\mathbf{k}}_{t+1} - \hat{\mathbf{k}}_t - \hat{c} \right\}_{t \in M} \quad M = \begin{cases} \{1950, \dots, 2011\} & \text{nőkre} \\ \{1980, \dots, 2011\} & \text{férfiakra és unisex táblára} \end{cases} \quad (41)$$

(9) autoregressziós egyenletből származó mintát, akkor feltesszük, hogy azok normális eloszlásból származnak és egymástól függetlenek. Ez utóbbi megállapítás önkényes, hiszen az elején csak azt a megköttést vettük, hogy k_t véletlenbolyongásos, driftes folyamatot követ és azt pedig nem állapíthatjuk meg teljes bizonyossággal, hogy a \hat{c} és $\hat{\sigma}_\varepsilon^2$ a kapott paraméterek valóban a folyamatot leíró megfelelő paraméterek, ezért a függetlenség nem tisztázott. Ha ettől eltekintünk, akkor vehetjük azokat a nullhipotéziseket, miszerint a $\left\{ \hat{Z}_t = z_t = \hat{\mathbf{k}}_{t+1} - \hat{\mathbf{k}}_t - \hat{c} \right\}_{t \in M}$ és az $\left\{ \hat{\varepsilon}_{(x,t)} = \varepsilon_{(x,t)} = \ln(q_{x,t}) - \hat{\mathbf{a}}_x - \hat{\mathbf{b}}_x \hat{\mathbf{k}}_t \right\}_{t \in L}$ $L = \begin{cases} \{1950, \dots, 2012\} & \text{nőkre} \\ \{1980, \dots, 2012\} & \text{férfiakra és unisex táblára} \end{cases}$ a megfelelő normális eloszlásokból származnak. Készítünk egy táblát, amelyben szerepeltetjük a szórásokat nőkre, férfiakra és a teljes lakosságra egyaránt, és azokból megkapjuk a két mintaátlagra vonatkozó statisztikai mutatónak a kétoldali, 95%-os konfidenciaintervallumait:

	$\hat{\sigma}_\varepsilon$	$\hat{\sigma}_Z$
Nő	0,1605	3,6676
Férfi	0,1397	3,1025
Unisex	0,1131	2,7483

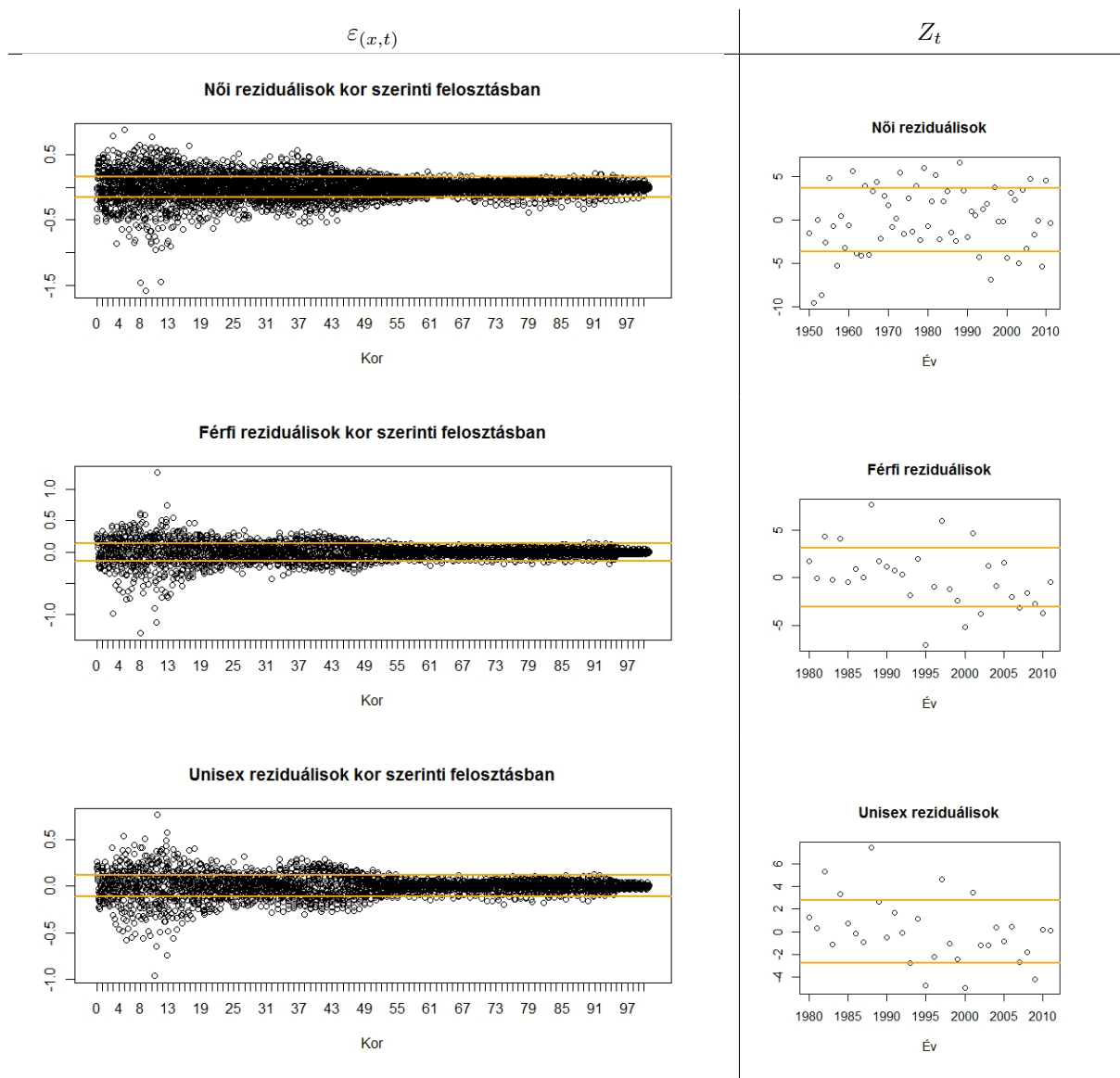
$$\varepsilon_{(x,t)} \text{ konfidenciaintervalluma: } \left(\frac{\sum_{x=0}^{\omega} \sum_{t \in L} \hat{\varepsilon}_{(x,t)}}{(\omega + 1)|L|} \pm u_{5\%/2} \frac{\hat{\sigma}_\varepsilon}{\sqrt{(\omega + 1)|L|}} \right) = \quad (42)$$

$$= \begin{cases} (-0,00063; 0,00063) & \text{nőkre} \\ (-0,00066; 0,00066) & \text{férfiakra} \\ (-0,00043; 0,00043) & \text{unisex táblára} \end{cases} \quad (43)$$

$$Z_t \text{ konfidenciaintervalluma: } \left(\frac{\sum_{t \in M} \hat{Z}_t}{|M|} \pm u_{5\%/2} \frac{\hat{\sigma}_Z}{\sqrt{|M|}} \right) = \quad (44)$$

$$= \begin{cases} (-5, 21908; 1, 47755) & \text{nőkre} \\ (-5, 29959; 1, 37053) & \text{férfiakra} \\ (-4, 47355; 0, 76022) & \text{unisex táblára} \end{cases} \quad (45)$$

Látható, hogy a 0 a fenti hat intervallumnak a része, ezért elfogadjuk a nullhipotéziseket mind a kétféle reziduálisnál nő–férfi bontásban is. Nem evidens azonban, hogy a nullhipotézis ellenére van-e szemmel látható struktúrája a reziduális változóknak az előbbi megfontolások miatt, – ha nem tartjuk nem elhanyagolhatónak a függetlenségi feltételt, – ezért ábrázoljuk azokat. $\varepsilon_{(x,t)}$ esetében is 2-dimenziós ábrát veszünk, hogy a minta szóródását jobban szemügyre vehessük, ahol koronként és a korokon belül évenként csoportosítjuk az értékeket.



A fenti hat ábrán a sárga színű egyenesekkel a horizontális tengelyeknek megfelelő szórásokkal

vett távolságát jelöltük. Látható, hogy a Z_t alakulásában nincsen minta és trend, és az értékek nagyobb hányada a sárga színűs egyenesek által határolt területen található. Ez azt jelenti, hogy a zajon kívül már nem maradt információ, amit a rendszerből ki lehetne szűrni. $\varepsilon_{(x,t)}$ -nél már észrevehetjük, hogy a minta szóródása nem homogén. Mind a három ábra esetén ugyanaz a helyzet áll fenn: az 50 éves korig tartó időszakokban a szórás nagyobb, mint 50 éves kor felett. Természetesen betudhatjuk ezt a véletlennek, bár kevésbé tűnik valószínűnek, hiszen nemcsak a nőknél, hanem a férfiak és az unisex halandósági táblájára is ugyanezt állapíthatjuk meg. A teljesség igényét figyelembe véve célszerű leírni a módosított Lee–Carter-modellt mortalitási valószínűségekre. Az eredetihez képest kétféle $\varepsilon_{(x,t)}$ reziduálist tartunk érdemesnek a modellben alkalmazni:

$$\varepsilon_{(x,t)} = \begin{cases} \varepsilon_{(x,t)}^{(1)} \sim \mathcal{N}(0, \sigma_{\varepsilon_1}^2) & \text{ha } x \leq 50 \\ \varepsilon_{(x,t)}^{(2)} \sim \mathcal{N}(0, \sigma_{\varepsilon_2}^2) & \text{ha } x > 50 \end{cases} \quad (46)$$

ahol $\varepsilon_{(x,t)}^{(1)}$ és $\varepsilon_{(x,t)}^{(2)}$ egymástól is független változók. A (8) egyenletet tehát az alábbi alakra hoznánk:

$$\ln(q_{x,t}) = \begin{cases} a_x^{(1)} + b_x^{(1)}k_t^{(1)} + \varepsilon_{(x,t)}^{(1)} & \text{ha } x \leq 50 \\ a_x^{(2)} + b_x^{(2)}k_t^{(2)} + \varepsilon_{(x,t)}^{(2)} & \text{ha } x > 50 \end{cases}, \quad (47)$$

ahol hasonló megkötéseket vennénk, mint a (10) egyenletekben:

$$\sum_{x=0}^{50} b_x^{(1)} = \sum_{x=51}^{\omega} b_x^{(2)} = 1 \quad \text{és} \quad \sum_{t \in L} k_t^{(1)} = \sum_{t \in L} k_t^{(2)} = 0 \quad (48)$$

3.4. A módosított Lee–Carter-modell mortalitási valószínűségek becslése

Ebben a fejezet részben a módosított feladatban található paraméterek maximum likelihood becsléseit szeretnénk meghatározni. Ehhez jelentős segítséget nyújt VÉKÁS [2] doktori értekezésében az eredeti Lee–Carter-modellben a (11), (12) és a (13) becslések bizonyítása, amelynek egészét nem részletezzük, hiszen az ott írva található. Jelölje $L(\theta)$ a módosított Lee–Carter-modell likelihood-függvényét, ahol

$$\theta = \left([\mathbf{a}^{(1)}]^\top, [\mathbf{a}^{(2)}]^\top, [\mathbf{b}^{(1)}]^\top, [\mathbf{b}^{(2)}]^\top, [\mathbf{k}^{(1)}]^\top, [\mathbf{k}^{(2)}]^\top, \sigma_{\varepsilon_1}^2, \sigma_{\varepsilon_2}^2 \right) \quad (49)$$

$[2(\omega + 1) + 2|L| + 2]$ -dimenziós paramétervektor. Ekkor kihasználva $\varepsilon_{(x,t)}^{(1)}$ és $\varepsilon_{(x,t)}^{(2)}$ függetlenségét és hogy rendre $\mathcal{N}(0, \sigma_{\varepsilon_1}^2)$ és $\mathcal{N}(0, \sigma_{\varepsilon_2}^2)$ eloszlásból származnak, felírjuk a likelihood-függvényt:

$$L(\theta) = \left(\frac{1}{\sqrt{2\pi\sigma_{\varepsilon_1}^2}} \right)^{51 \cdot |L|} \prod_{x=0}^{50} \prod_{t \in L} \exp \left(- \frac{\left(\ln(q_{x,t}) - a_x^{(1)} - b_x^{(1)}k_t^{(1)} \right)^2}{2\sigma_{\varepsilon_1}^2} \right) \cdot \left(\frac{1}{\sqrt{2\pi\sigma_{\varepsilon_2}^2}} \right)^{(\omega-50) \cdot |L|} \prod_{x=51}^{\omega} \prod_{t \in L} \exp \left(- \frac{\left(\ln(q_{x,t}) - a_x^{(2)} - b_x^{(2)}k_t^{(2)} \right)^2}{2\sigma_{\varepsilon_2}^2} \right) \quad (50)$$

Ebből a loglikelihood-függvény ($l(\theta) = \ln(L(\theta))$) is könnyen adódik:

$$l(\theta) = -\ln\left(\sqrt{2\pi}\right)^{(\omega+1)|L|} - 51|L|\ln(\sigma_{\varepsilon_1}) - (\omega-50)|L|\ln(\sigma_{\varepsilon_2}) - \sum_{x=0}^{50} \sum_{t \in L} \frac{\left(\ln(q_{x,t}) - a_x^{(1)} - b_x^{(1)} k_t^{(1)}\right)^2}{2\sigma_{\varepsilon_1}^2} - \sum_{x=51}^{\omega} \sum_{t \in L} \frac{\left(\ln(q_{x,t}) - a_x^{(2)} - b_x^{(2)} k_t^{(2)}\right)^2}{2\sigma_{\varepsilon_2}^2} \quad (51)$$

Fölismerhető, hogy végsősoron hasonló maximalizálási feladathoz jutottunk, mint VÉKÁS [2] a (4.3.1)-es részfejezetben azzal a különbséggel, hogy nálunk a loglikelihood-függvénybeli tagok két részre oszthatók: vannak, amelyek csak az 50 év alattiakat és vannak, amelyek csak az 50 év felettieket jellemző paraméterek határoznak meg. Így megengedhető, hogy a kétféle szummát külön-külön maximalizáljuk. A becslések tehát a következők (hasonlóan a (11) és a (13) egyenletekhez):

$$\hat{a}_x^{(1)} = \frac{1}{|L|} \sum_{t \in L} \ln q_{x,t} \quad \text{és} \quad \hat{a}_x^{(2)} = \frac{1}{|L|} \sum_{t \in L} \ln q_{x,t} \quad (52)$$

$$\hat{\mathbf{b}}^{(1)} = \mathbf{M}_1 \mathbf{M}_1^T \text{ normált domináns sajátvektora,} \quad \hat{\mathbf{k}}^{(1)} = \mathbf{M}_1^T \hat{\mathbf{b}}^{(1)} \quad (53)$$

$$\hat{\mathbf{b}}^{(2)} = \mathbf{M}_2 \mathbf{M}_2^T \text{ normált domináns sajátvektora,} \quad \hat{\mathbf{k}}^{(2)} = \mathbf{M}_2^T \hat{\mathbf{b}}^{(2)} \quad (54)$$

$$\hat{\sigma}_{\varepsilon_1}^2 = \frac{1}{51|L|} \sum_{x=0}^{50} \sum_{t \in L} \left(\ln(q_{x,t}) - \hat{a}_x^{(1)} - \hat{\mathbf{b}}_x^{(1)} \hat{\mathbf{k}}_t^{(1)}\right)^2 \quad (55)$$

$$\hat{\sigma}_{\varepsilon_2}^2 = \frac{1}{(\omega-50)|L|} \sum_{x=51}^{\omega} \sum_{t \in L} \left(\ln(q_{x,t}) - \hat{a}_x^{(2)} - \hat{\mathbf{b}}_x^{(2)} \hat{\mathbf{k}}_t^{(2)}\right)^2, \quad (56)$$

ahol $\mathbf{M}_1 = \left[\ln(q_{x,t}) - \frac{1}{|L|} \sum_{t \in L} \ln(q_{x,t}) \right]_{x=0}^{50}$ $51 \times |L|$ -dimenziós és $\mathbf{M}_2 = \left[\ln(q_{x,t}) - \frac{1}{|L|} \sum_{t \in L} \ln(q_{x,t}) \right]_{x=51}^{\omega}$ $(\omega-50) \times |L|$ -dimenziós mátrixok. Ez esetben a korábban már kiszámolt $e_{n,t}^{(0)}$, $e_{n,t}^{(1)}$ és $e_{n,t}^{(2)}$ előrevetítések és a $\xi_{(2012+t)}^{(n)}$ valószínűségi modell módosításra szorúlnak, jelölje azokat rendre $e_{n,t}^{(0),m}$, $e_{n,t}^{(1),m}$, $e_{n,t}^{(2),m}$ és $\xi_{(2012+t)}^{(n),m}$. Minthogy az eredeti modellben az a feltétel, hogy a mortalitási indexek véletlenbolyongásos, driftes folyamatot követnek, ezért itt is ugyanezt fogjuk feltenni azzal a megkötéssel, hogy $k_t^{(1)}$ és $k_t^{(2)}$ egymástól és az individuális reziduálisoktól is függetlenek. Lévén, hogy azok véletlen folyamatként értelmezhetőek 2012 után, érdemes inkább a $K_t^{(1)}$ és $K_t^{(2)}$ jelöléseket alkalmazni utalva arra, hogy azok valószínűségi változók. Szintén hasonló alakra hozhatók a becslések:

$$\hat{c}^{(r)} = \frac{\hat{\mathbf{k}}_{|L|}^{(r)} - \hat{\mathbf{k}}_1^{(r)}}{|L|-1}, \quad \hat{\sigma}_{Z_r}^2 = \frac{1}{|L|-1} \sum_{t \in M} \left(\hat{\mathbf{k}}_{t+1}^{(1)} - \hat{\mathbf{k}}_t^{(r)} - \hat{c}^{(r)}\right)^2 \quad (r \in \{1, 2\}) \quad (57)$$

Visszagondolva $e_{n,t}^{(0)}$ levezetésére egyszerűbbnek mutatkozik az az eset, ha az egyén 51 éves vagy annál idősebb. Átlapozva a (2) fejezetben bemutatott lépéseket láthatóvá válik az explicit alak:

$$\begin{aligned}
e_{n,t}^{(0),m} \Big|_{n \geq 51} &= \mathbb{E} \left(\xi_{(2012+t)}^{(n),m} \Big| \mathcal{F}_{2012}, n \geq 51 \right) \\
&= \sum_{i=1}^{\omega-n} i \cdot \left(\exp \left(\hat{\mathbf{a}}_{n+i}^{(2)} + \hat{\mathbf{b}}_{n+i}^{(2)}(i+t)\hat{c}^{(2)} + \hat{\mathbf{b}}_{n+i}^{(2)} \hat{\mathbf{k}}_{2012}^{(2)} + \frac{\sigma_{\varepsilon_2}^2}{2} + (i+t) \left(\hat{\mathbf{b}}_{n+i}^{(2)} \right)^2 \frac{\sigma_{Z_2}^2}{2} \right) \right. \\
&\quad + \sum_{k=1}^i (-1)^k \sum_{0 \leq j_1 < \dots < j_k \leq i-1} \exp \left(\left[\hat{\mathbf{a}}_{n+i}^{(2)} + \sum_{l=1}^k \hat{\mathbf{a}}_{n+j_l}^{(2)} \right] + \left[\left(\hat{\mathbf{b}}_{n+i}^{(2)} + \sum_{l=1}^k \hat{\mathbf{b}}_{n+j_l}^{(2)} \right) \hat{\mathbf{k}}_{2012}^{(2)} \right] \right. \\
&\quad \left. + \left[\left((t+i)\hat{\mathbf{b}}_{n+i}^{(2)} + \sum_{l=1}^k (t+j_l) \hat{\mathbf{b}}_{n+j_l}^{(2)} \right) \hat{c}^{(2)} \right] + (k+1) \frac{\sigma_{\varepsilon_2}^2}{2} \right. \\
&\quad \left. \left. + \frac{\sigma_{Z_2}^2}{2} \left[(j_1+t) \left(\hat{\mathbf{b}}_{n+i}^{(2)} + \sum_{l=1}^k \hat{\mathbf{b}}_{n+j_l}^{(2)} \right)^2 + (j_2-j_1) \left(\hat{\mathbf{b}}_{n+i}^{(2)} + \sum_{l=2}^k \hat{\mathbf{b}}_{n+j_l}^{(2)} \right)^2 + \dots + (i-j_k) \left(\hat{\mathbf{b}}_{n+i}^{(2)} \right)^2 \right] \right) \right) \quad (58)
\end{aligned}$$

Ennél is nagyobb odafigyelést igényel az az eset, ha $n \leq 50$, hiszen a várható élettartamot az 50 év alatti és az 51 év feletti tényezők egyszerre befolyásolják. Szükségszerű tehát a szummát kétfelé bontani: az egyikben az i -szerinti indexelés 0-tól $50-n$ -ig, a másikban pedig $51-n$ -től $\omega-n$ -ig tart lévén, hogy várható élettartamról van szó. Az utóbbi esetben is számos összegzést kell végrehajtani és azokon belül is két esetet kell megkülönböztetni, nevezetesen hogy a mortalitási valószínűségek 50 év alatti vagy az 51 év feletti korcsoporthoz tartoznak-e.

$$\begin{aligned}
e_{n,t}^{(0),m} \Big|_{n \leq 50} &= \mathbb{E} \left(\xi_{(2012+t)}^{(n),m} \Big| \mathcal{F}_{2012}, n \leq 50 \right) \\
&= \sum_{i=0}^{50-n} i \cdot \left(\exp \left(\hat{\mathbf{a}}_{n+i}^{(1)} + \hat{\mathbf{b}}_{n+i}^{(1)}(i+t)\hat{c}^{(1)} + \hat{\mathbf{b}}_{n+i}^{(1)} \hat{\mathbf{k}}_{2012}^{(1)} + \frac{\sigma_{\varepsilon_1}^2}{2} + (i+t) \left(\hat{\mathbf{b}}_{n+i}^{(1)} \right)^2 \frac{\sigma_{Z_1}^2}{2} \right) \right. \\
&\quad + \sum_{k=1}^i (-1)^k \sum_{0 \leq j_1 < \dots < j_k \leq i-1} \exp \left(\left[\hat{\mathbf{a}}_{n+i}^{(1)} + \sum_{l=1}^k \hat{\mathbf{a}}_{n+j_l}^{(1)} \right] + \left[\left(\hat{\mathbf{b}}_{n+i}^{(1)} + \sum_{l=1}^k \hat{\mathbf{b}}_{n+j_l}^{(1)} \right) \hat{\mathbf{k}}_{2012}^{(1)} \right] \right. \\
&\quad \left. + \left[\left((t+i)\hat{\mathbf{b}}_{n+i}^{(1)} + \sum_{l=1}^k (t+j_l) \hat{\mathbf{b}}_{n+j_l}^{(1)} \right) \hat{c}^{(1)} \right] + (k+1) \frac{\sigma_{\varepsilon_1}^2}{2} \right. \\
&\quad \left. \left. + \frac{\sigma_{Z_1}^2}{2} \left[(j_1+t) \left(\hat{\mathbf{b}}_{n+i}^{(1)} + \sum_{l=1}^k \hat{\mathbf{b}}_{n+j_l}^{(1)} \right)^2 + (j_2-j_1) \left(\hat{\mathbf{b}}_{n+i}^{(1)} + \sum_{l=2}^k \hat{\mathbf{b}}_{n+j_l}^{(1)} \right)^2 + \dots + (i-j_k) \left(\hat{\mathbf{b}}_{n+i}^{(1)} \right)^2 \right] \right) \right) \quad (59)
\end{aligned}$$

$$\begin{aligned}
& + \sum_{i=51-n}^{\omega-n} i \cdot \left(\exp \left(\hat{\mathbf{a}}_{n+i}^{(2)} + \hat{\mathbf{b}}_{n+i}^{(2)}(i+t)\hat{c}^{(2)} + \hat{\mathbf{b}}_{n+i}^{(2)}\hat{\mathbf{k}}_{2012}^{(2)} + \frac{\sigma_{\varepsilon_2}^2}{2} + (i+t) \left(\hat{\mathbf{b}}_{n+i}^{(2)} \right)^2 \frac{\sigma_{Z_2}^2}{2} \right) \right. \\
& \quad + \sum_{k=1}^i (-1)^k \sum_{\substack{0 \leq j_1 < \dots < j_{k_1} \leq 50 \\ 51 \leq r_1 < \dots < r_{k_2} \leq i-1 \\ k_1 + k_2 = k}} \exp \left(\left[\hat{\mathbf{a}}_{n+i}^{(2)} + \sum_{l=1}^{k_2} \hat{\mathbf{a}}_{n+r_l}^{(2)} + \sum_{l=1}^{k_1} \hat{\mathbf{a}}_{n+j_l}^{(1)} \right] \right. \\
& \quad \left. + \left[\left(\hat{\mathbf{b}}_{n+i}^{(2)} + \sum_{l=1}^{k_2} \hat{\mathbf{b}}_{n+r_l}^{(2)} \right) \hat{\mathbf{k}}_{2012}^{(2)} + \left(\sum_{l=1}^{k_1} \hat{\mathbf{b}}_{n+j_l}^{(1)} \right) \hat{\mathbf{k}}_{2012}^{(1)} \right] \right. \\
& \quad + \left[\left((t+i)\hat{\mathbf{b}}_{n+i}^{(2)} + \sum_{l=1}^{k_2} (t+r_l)\hat{\mathbf{b}}_{n+r_l}^{(2)} \right) \hat{c}^{(2)} + \left(\sum_{l=1}^{k_1} (t+j_l)\hat{\mathbf{b}}_{n+j_l}^{(1)} \right) \hat{c}^{(1)} \right] + (k_2+1)\frac{\sigma_{\varepsilon_2}^2}{2} + k_1\frac{\sigma_{\varepsilon_1}^2}{2} \\
& \quad + \frac{\sigma_{Z_2}^2}{2} \left[(r_1+t) \left(\hat{\mathbf{b}}_{n+i}^{(2)} + \sum_{l=1}^{k_2} \hat{\mathbf{b}}_{n+r_l}^{(2)} \right)^2 + (r_2-r_1) \left(\hat{\mathbf{b}}_{n+i}^{(2)} + \sum_{l=2}^{k_2} \hat{\mathbf{b}}_{n+r_l}^{(2)} \right)^2 + \dots + (i-r_{k_2}) \left(\hat{\mathbf{b}}_{n+i}^{(2)} \right)^2 \right] \\
& \quad \left. + \frac{\sigma_{Z_1}^2}{2} \left[(j_1+t) \left(\sum_{l=1}^{k_1} \hat{\mathbf{b}}_{n+j_l}^{(1)} \right)^2 + (j_2-j_1) \left(\sum_{l=2}^k \hat{\mathbf{b}}_{n+j_l}^{(1)} \right)^2 + \dots + (j_{k_1}-j_{k_1-1}) \left(\hat{\mathbf{b}}_{n+j_{k_1}}^{(1)} \right)^2 \right] \right) \\
& \hspace{15em} (60)
\end{aligned}$$

A kifejezés az eredeti $e_{n,t}^{(0)}$ -hoz képest még komplikáltabb, nyilvánvalóan annak explicit kiszámítására nincs lehetőség. Tisztáznunk kell, hogy végül maradunk-e az eredeti modellnél vagy a módosított verziót fogjuk-e használni. A χ^2 -próba – amit a bázisidőszakok kiválasztásánál alkalmaztunk a (3.2) részfejezetben – is segíthet eldönteni a dolgot sorsát. Először szükséges kiszámítani a módosított modellből származó mortalitási ráták várható értékét, mint a jövő halálzási folyamatait prezentáló, lehetséges legjobb becsléseket a 2000-ig rendelkezésre álló információk alapján a 2001, 2002, ..., 2012-es évekre. Mint ahogyan azt korábban is tettük, jobb felső indexben m jelöli majd a változókra vonatkozó módosításokat és a (39)-es egyenlethez hasonló átalakításokkal kapom a következőket (az érthetőség kedvéért a paraméterbecsléseknél a bázisidőszakokhoz tartozó indexek (i) feltüntetése nélkül):

$$\begin{aligned}
\hat{q}_{x,t}^{(i),m} \Big|_{x \leq 50} &= \mathbb{E} \left(Q_{x,t}^m \Big| F_{2000}^{(i)}, x \leq 50 \right) \\
&= \exp \left(\hat{\mathbf{a}}_x^{(1)} + \hat{\mathbf{b}}_x^{(1)} \left(\hat{\mathbf{k}}_{2000}^{(1)} + (t-2000)\hat{c}^{(1)} \right) + (t-2000) \frac{\left(\hat{\mathbf{b}}_x^{(1)} \right)^2 \hat{\sigma}_{Z_1}^2}{2} + \frac{\hat{\sigma}_{\varepsilon_1}^2}{2} \right) \Big|_{x \leq 50} \hspace{2em} (61)
\end{aligned}$$

$$\begin{aligned} \hat{q}_{x,t}^{(i),m} \Big|_{x \geq 51} &= \mathbb{E} \left(Q_{x,t}^m \mid F_{2000}^{(i)}, x \geq 51 \right) \\ &= \exp \left(\hat{\mathbf{a}}_x^{(2)} + \hat{\mathbf{b}}_x^{(2)} \left(\hat{\mathbf{k}}_{2000}^{(2)} + (t-2000)\hat{c}^{(2)} \right) + (t-2000) \frac{\left(\hat{\mathbf{b}}_x^{(2)} \right)^2 \hat{\sigma}_{Z_2}^2 + \hat{\sigma}_{\varepsilon_2}^2}{2} \right) \Big|_{x \geq 51} \end{aligned} \quad (62)$$

A numerikus elemzéseket ebben a fejezetben is az unisex táblára készítjük el, hogy a (3.2) részfejezetben kapott eredményeket összehasonlíthassuk. Az alábbi táblázatban bemutatjuk az individuális és a véletlenbolyongásos reziduálisok tapasztalati szórásait valamint a χ^2 -statisztikákat. Az utóbbinál az összegzést szintén 25 éves korig fogjuk venni, mivel az öregkori kitettségekről kevés információ áll rendelkezésre.

Bázisidőszak	$\hat{\sigma}_{\varepsilon_1}$	$\hat{\sigma}_{Z_1}$	$\hat{\sigma}_{\varepsilon_2}$	$\hat{\sigma}_{Z_2}$
'50–2000	0,1684	1,3697	0,0637	0,1544
'60–2000	0,1359	0,2525	0,0532	0,0246
'70–2000	0,1381	0,4307	0,0526	0,0361
'80–2000	0,1241	1,4791	0,0405	0,6701
'89–2000	0,1197	1,7749	0,0346	1,0068

χ^2	'50–2000	'60–2000	'70–2000	'80–2000	'89–2000
2001	11975,30	129,42	80,45	33,61	32,39
2002	9907,05	127,27	68,06	19,29	23,30
2003	10714,29	129,56	71,29	23,98	32,84
2004	9454,66	141,44	75,98	24,98	36,57
2005	9032,27	154,53	85,47	41,06	63,95
2006	7980,03	178,11	100,84	44,85	53,67
2007	8670,22	148,13	80,84	41,78	69,08
2008	8575,46	154,17	87,54	52,12	84,23
2009	7326,65	184,16	104,80	55,29	85,44
2010	8078,42	221,32	136,83	64,08	74,40
2011	7460,70	188,77	103,19	46,20	81,39
2012	8106,63	191,07	110,74	65,23	106,09
szumma	107281,70	1947,97	1106,02	512,47	743,35

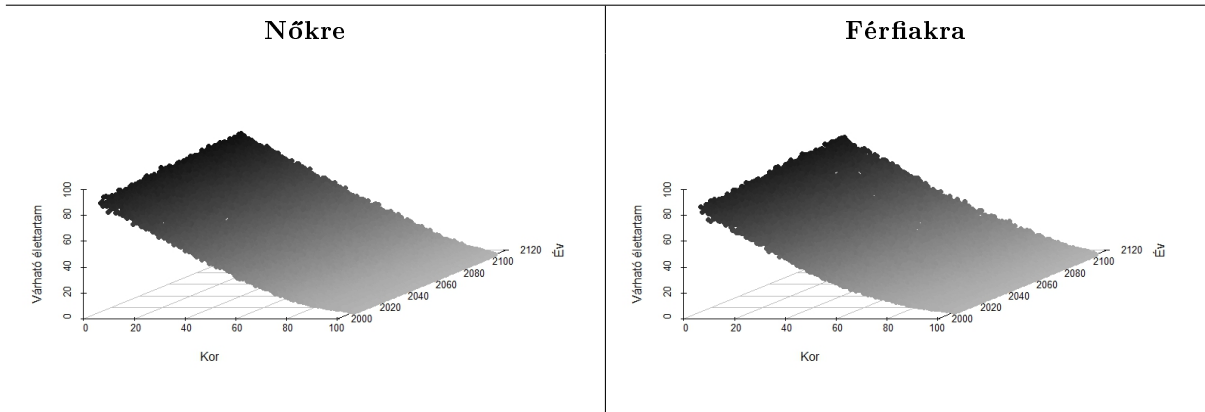
Számításaink igazolják, hogy a fiatalabb korosztályban majdnem háromszor annyi az individuális szórás, mint az 50 év felettiekénél. Az egymástól függetlennek tekintett véletlenbolyongásos folyamatoknál az idősebbek szórásparamétereire a fiatalabbakéhoz viszonyítva kisebb értékeket kaptunk. A 26 és a $12 \times 26 = 312$ szabadságfokokhoz tartozó χ^2 -eloszlás 95%-os kvantilisei 38,88 és 354,19,

amelyek 5%-os szignifikanciaszinten nézve éppen a kritikus értékeket adják meg. A pirossal szedett számokhoz tartozó értékeknél elutasítjuk azt a hipotézist, hogy a várt halálozási valószínűségek származhatnak a magyarországi lakosságra vonatkozó halálozási adatokból. Az elfogadási és elutasítási tartományokba eső részek pontosan ugyanazok, mint az eredeti modellnél, ezért a két modell előrejelzési pontossága 5%-os szignifikanciaszinten megegyezik unisex tábla esetén. Mivel a módosított modellnél nem szolgálhatunk lényegileg pontosabb előrejelzésekkel, ezért a továbbiakban az eredeti, kevesebb paramétert tartalmazó Lee–Carter-moddal fogunk dolgozni.

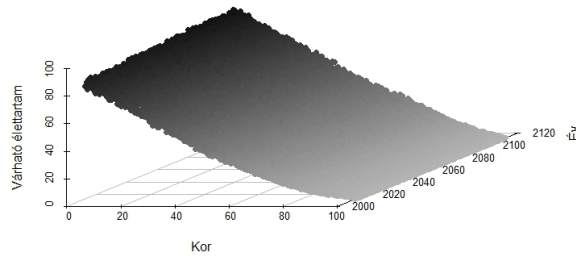
3.5. Az $e_{n,t}^{(0)}$, $e_{n,t}^{(1)}$ és $e_{n,t}^{(2)}$ előrevetítések megrajzolása

A (3.2) részfejezetben sikerült megállapítani, hogy mely bázisidőszakok esetében alkalmazzuk a Lee–Carter-modellt nőkre, férfiakra és a teljes lakosságra nézve, valamint az azokhoz tartozó paraméterbecsléseket is. Azt is megállapítottuk, hogy a módosított Lee–Carter-modell szimulációs előrejelzései nem mutattak számottevő eltérést az eredetihez képest, ezért a többi (rész)fejezetet az eredetileg felállított modellnek megfelelően fogjuk részletezni. Most már minden szükséges adat rendelkezésre áll, hogy megnézzük, mégis milyen viszonyban állnak a bonyolult kifejezésekkel leírható $e_{n,t}^{(0)}$, $e_{n,t}^{(1)}$ és $e_{n,t}^{(2)}$ várható élettartamok, ahol $e_{n,t}^{(0)}$ az explicit, $e_{n,t}^{(1)}$ az általam végzett közelítő módszerrel, $e_{n,t}^{(2)}$ pedig a gyakorlatban alkalmazott módszerrel meghatározható várható élettartamok n éves egyénre a $(2012+t)$. évben. A következő ábrán megmutatjuk, hogy 100 éven át, 2018-tól 2117-ig hogyan alakul születéstől egészen a feltételezett legnagyobb életkorig (esetünkben $\omega = 100$ évig) a modell szerint nőkre, férfiakra és unisex táblára a várható élettartam.

Várható élettartamok



Unisex táblára



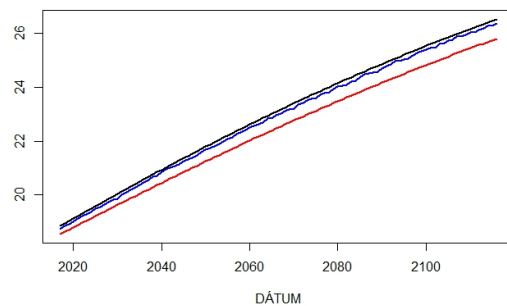
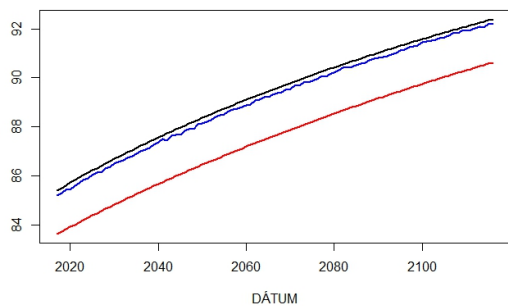
A megszorodott műveletek időigénye miatt csak 25-szörös replikációval hajtjuk végre a Monte-Carlo-szimulációt a 16. oldaltól ismertetett lépéseknek megfelelően. Minél sötétebb a felület, annál nagyobb az érték. Nem meglepő, hogy bármely évet tekintve a kor függvényében csökken a várható évek száma. Kevésbé észrevehető, de szemügyre véve az ábrákat főleg a nőknél az idő előrehaladtával a születéskori várható élettartamok lassan emelkednek. A modell a várakozásnak megfelelően tehát növekvő tendenciát mutat időben előrehaladva a várható értékeknél. Bár nincsen technikai akadálya a szimulációval kapott $e_{n,t}^{(0)}$ értékeken felül berajzolni az $e_{n,t}^{(1)}$ és $e_{n,t}^{(2)}$ várható értékeket (esetleg más, egymástól elütő színezéssel), azonban nehezen átlátható ábrát kapnánk a pontthalmazok egymásba olvadása miatt. Célszerű csak születéskori és 65 éves keresztmetszetben ábrázolni azokat. Feketével a gyakorlatban alkalmazott, pirossal a saját közelítéssel és kézzel pedig a 2000 db. replikációval kapott várható élettartamok szerepelnek.

2018–2117

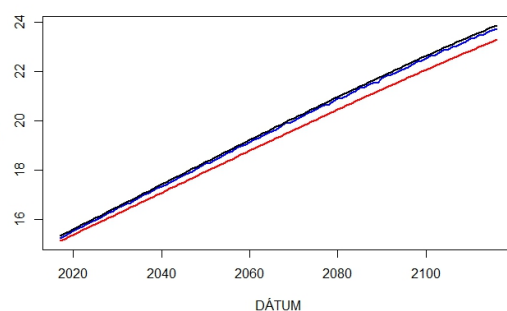
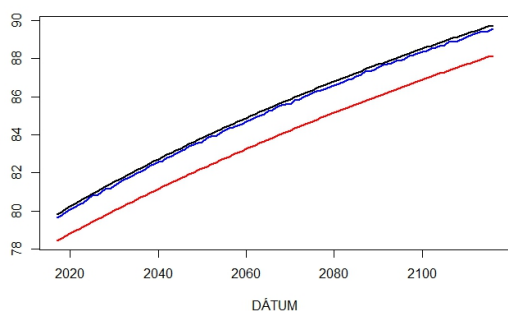
Születés kori várható élettartamok
($n = 0$)

65 éves egyének várható élettartamai
($n = 65$)

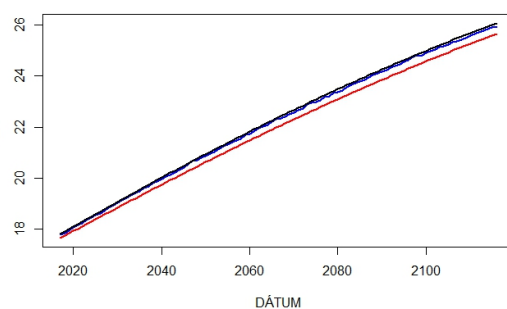
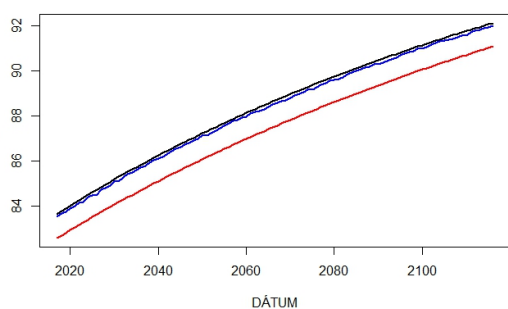
Nőkre



Férfiakra



Unisex táblára



Az ábrákon látszik, hogy sikerült a szimulációval kapott $e_{n,t}^{(0)}$ várható éveket úgy kiszámítani, hogy azok legfeljebb negyedévet torzítsanak. Sőt mind a hat ábrán az is észrevehető, hogy a modellbéli várható értéket a saját közelítésem alul-, a gyakorlatban vett közelítés pedig fölülbecsli, vagyis $e_{0,t}^{(1)} < e_{0,t}^{(0)} < e_{0,t}^{(2)}$ és $e_{65,t}^{(1)} < e_{65,t}^{(0)} < e_{65,t}^{(2)}$ ($\forall t \in \{6, \dots, 106\}$) A saját becsléseim 65 éves korban 1

évnél kisebb hibával közelítik a modell szerinti tényleges értékeket, míg születés kori esetekben akár 2 éves különbségek is adódnak. A gyakorlatban alkalmazott módszer sokkal pontosabbnak bizonyul, alig nagyobb a replikációval kapott várható évektől. Így három releváns információhoz jutottunk: egyrészt hatásos becslést kapunk $e_{n,t}^{(2)}$ -vel, mert sok helyen elhanyagolható az $e_{n,t}^{(0)}$ -vel vett különbség. Másrészt ahol nem tekinthető kicsinek, ott felső határt kapunk. Továbbá az ábrák bizonyítékul szolgálnak arra, hogy az idő előrehaladtával majdnem minden évben (t -ben) szigorúan monoton nőnek a várható élettartamokhoz tartozó görbék majdnem mindig csökkenő meredekséggel. Természetesen a modell megalkotottsága miatt az elérhető legmagasabb életkort ($\omega = 100$ -at) nem haladhatják meg, hiszen a (14) egyenletlenség csak akkor teljesül, ha a 100. évben az egyének 1 valószínűséggel halnak meg.

Bár sikerült kiszámítani a jövőbeli halálozási ráták időbeli alakulásainak várható értékét, sajnos tudomásul kell venni, hogy a realizáció azoktól eltérhet. Érdekes számszerűsíteni, hogy az egyes években és életkorokban hogyan ingadoznak a szimulációval kapott várható élettartamok. A 2018-es évre ($t = 6$ -ra) vonatkozóan 5000-es replikációt generálunk a születés kori és 65 éves korban várható élettartamokra női–férfi bontásban és a teljes lakosságra egyaránt, majd azokra hisztogramot készítünk:⁸

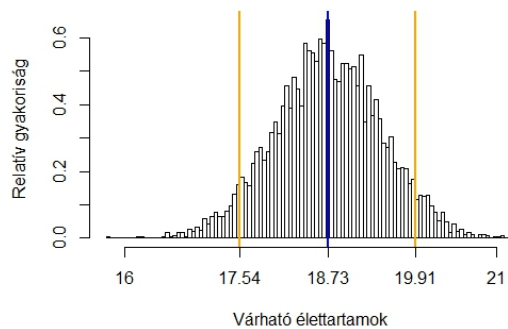
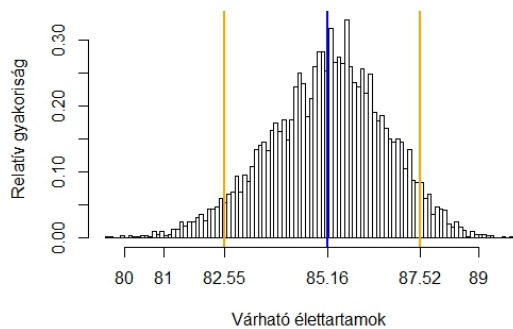
⁸A generált mintákra vonatkozó valamennyi statisztikai mutatót a függelékben az I/(a), I/(b) és I/(c) táblázatok tartalmazzák.

($t = 6$)

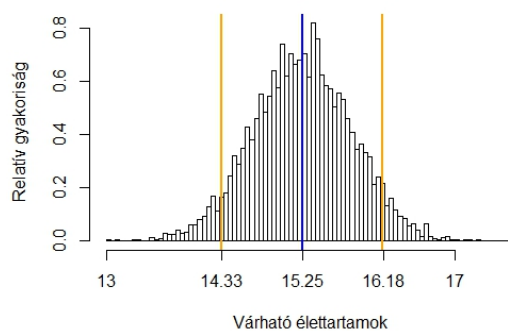
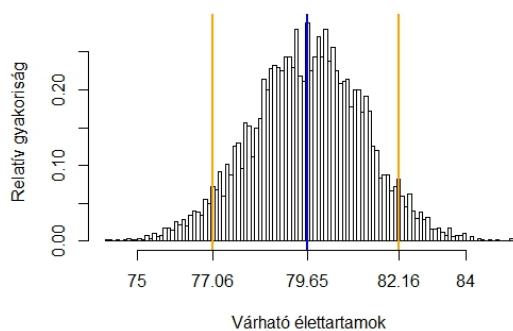
Születéskori ($n = 0$)

65 éves korra ($n = 65$)

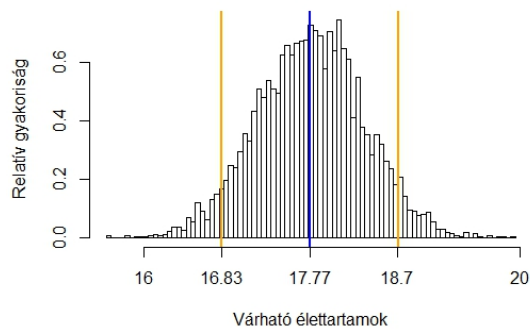
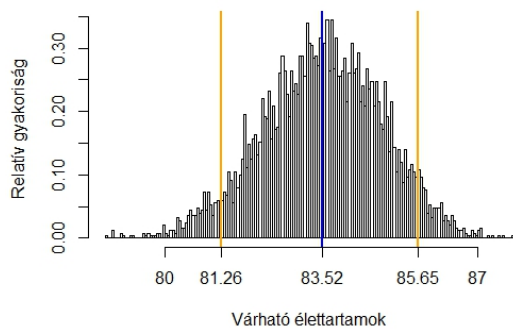
Nőkre



Férfiakra



Unisex táblára



A kék vertikális vonalak az egyes mintákon vett átlagot, a narancssárga vertikális vonalak pedig a sorrendbe rakott minták 5%- és 95%-hoz tartozó kvantiliseit jelölik. Ha visszagondolunk a (19) kifejezésre, akkor látható, hogy a születéskori várható élettartam kiszámításánál több véletlen ta-

got kellett figyelembe venni, mint az idősebbeknél. Ez okozza azt a jelenséget, hogy a születéskori hisztogramok alapján az 5%- és 95%-hoz tartozó kvantilisok differenciája abszolút értékben körülbelül két évvel nagyobb a 65 évesekéhez képest. Az előbbi esetben ez a különbség nagyjából 4 év, míg az utóbbinál majdnem 2 év. Érdekes, hogy születéskor a női várható élettartam szórása (1,50 év) kisebb a férfiakénál (1,55 év), de 65 éves korban ez megváltozik: a férfiaknál 0,57 évet, a nőknél pedig 0,72 évet realizáltunk. Ez a váltás főleg a mortalitási index (K_t) véletlen tagjainak (Z_t) is köszönhető, amelyeknek a szórása nők esetében nagyobb. A hisztogramról leolvasható, hogy a nők várhatóan többet fognak élni mint a férfiak és ennek megfelelően unisex táblánál a várható élet születéskor és 65 éves korban a két nem megfelelő várható értékei közé esik.

3.6. Egy- és kétfázisú élettartamok generálása

Ez a fejezet azt hivatott bemutatni, hogy a nyers halálozási valószínűségekre alkalmazott Lee–Carter-modell mennyire robusztus. Már nem az számít, hogy az 1950-től 2012-ig éppen hogyan alakultak a tényleges halálozási ráták. Úgy tekintünk a fent megbecsült Lee–Carter-paraméterekre, hogy a mortalitási ráták időbeli alakulásának véletlen folyamatát megfelelően jellemzik. Eredetileg úgy jártunk el, hogy a 2012-ig rendelkezésre álló halálozási valószínűségekből maximum likelihood becslést vettünk a paraméterekre. Vajon ha újragenerálnánk 2012-ig az eddigi értékeket és újra vennénk a paraméterek becsléseit, akkor $2012 + t$ év múlva az n éves egyének várható élettartami hogyan változnának ahhoz képest, hogy az újragenerált értékeket 2012 után az eredeti paraméterezéssel folytatnánk tovább? Másképp szólva a Lee–Carter-modellben a paraméterbecsléseknél a maximum likelihood elv torzítatlan-e a várható élettartamra nézve? A modell jósága tehát nem összekeverendő a paraméterbecslések statisztikai torzítatlanságával, ahol az számít, hogy a mintasokaságból vett statisztikai mutató várható értékben a tényleges paraméterértékeket adja.

Legyen tehát $\hat{\theta} = (\hat{\mathbf{a}}^\top, \hat{\mathbf{b}}^\top, \hat{c}, \hat{\sigma}_Z^2, \hat{\sigma}_\varepsilon^2)$ a becslült paraméterek nőkre, férfiakra és unisex táblára a csoportoknak megfelelő bázisidőszakokon. Emlékeztetőül a nők esetében 1950-től 2012-ig, a másik két csoportnál pedig 1980-tól 2012-ig tartó időszakon alkalmaztuk a paraméterek becslését. A férfiak és unisex tábla esetén nincs értelme 1950-től kezdeni az újragenerálást, hiszen az 1950 és az 1980 közötti évekből nem használtunk fel információt. Tekintettel arra, hogy a mortalitási indexek driftes, véletlen bolyongásos folyamatot követnek, nem egyértelmű, hogy az 1950-es ill. az 1980-as mortalitási index milyen eloszlást követ. Sok más ötlet mellett célszerű talán egyetlen értéket adni azoknak, nevezetesen a becslült mortalitási indexek 1950-es és 1980-as évekhez tartozó értékeit.

Jelölje tehát

$$\mathbf{Q}_{2012} = \begin{pmatrix} Q_{0,1950} & Q_{0,1951} & \dots & Q_{0,2012} \\ Q_{1,1950} & Q_{1,1951} & \dots & Q_{1,2012} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{100,1950} & Q_{100,1951} & \dots & Q_{100,2012} \end{pmatrix} \text{ vagy } \begin{pmatrix} Q_{0,1980} & Q_{0,1981} & \dots & Q_{0,2012} \\ Q_{1,1980} & Q_{1,1981} & \dots & Q_{1,2012} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{100,1980} & Q_{100,1981} & \dots & Q_{100,2012} \end{pmatrix} \quad (63)$$

a megfelelő bázisidőszakokhoz tartozó mortalitási rátákat, mint valószínűségi változókat, amelyeket az első fázisban legenerálunk. Majd ezekből újabb paraméterbecsléseket képzünk az alábbi jelöléssel:

$$\hat{\theta}(\mathbf{Q}_{2012}) = \left[\hat{\mathbf{a}}^\top(\mathbf{Q}_{2012}), \hat{\mathbf{b}}^\top(\mathbf{Q}_{2012}), \hat{c}(\mathbf{Q}_{2012}), \hat{\sigma}_Z^2(\mathbf{Q}_{2012}), \hat{\sigma}_\varepsilon^2(\mathbf{Q}_{2012}) \right] \quad (64)$$

Végül a második fázisban a 2013, 2014, ..., 2100, 2101, stb. évekre továbbgeneráljuk a mintát kettéágaztatva aszerint, hogy az eredeti $\hat{\theta}$ vagy a $\hat{\theta}(\mathbf{Q}_{2012})$ vektorparamétert alkalmazzuk a modellben. Ha $\hat{\theta}$ szerint generálunk, akkor a

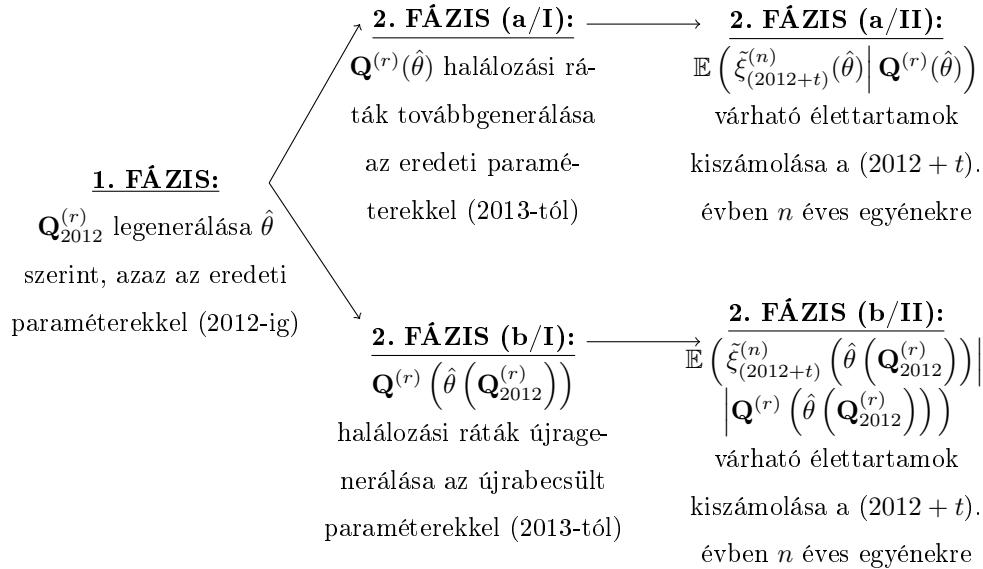
$$\mathbf{Q}(\hat{\theta}) = \begin{pmatrix} Q_{0,2013}(\hat{\theta}) & Q_{0,2014}(\hat{\theta}) & \dots & Q_{0,2100}(\hat{\theta}) & \dots \\ Q_{1,2013}(\hat{\theta}) & Q_{1,2014}(\hat{\theta}) & \dots & Q_{1,2100}(\hat{\theta}) & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ Q_{100,2013}(\hat{\theta}) & Q_{100,2014}(\hat{\theta}) & \dots & Q_{100,2100}(\hat{\theta}) & \dots \end{pmatrix} \quad (65)$$

mintát kapjuk eredményül, a másik esetben pedig jelölje azt

$$\mathbf{Q}(\hat{\theta}(\mathbf{Q}_{2012})) = \begin{pmatrix} Q_{0,2013}(\hat{\theta}(\mathbf{Q}_{2012})) & Q_{0,2014}(\hat{\theta}(\mathbf{Q}_{2012})) & \dots & Q_{0,2100}(\hat{\theta}(\mathbf{Q}_{2012})) & \dots \\ Q_{1,2013}(\hat{\theta}(\mathbf{Q}_{2012})) & Q_{1,2014}(\hat{\theta}(\mathbf{Q}_{2012})) & \dots & Q_{1,2100}(\hat{\theta}(\mathbf{Q}_{2012})) & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ Q_{100,2013}(\hat{\theta}(\mathbf{Q}_{2012})) & Q_{100,2014}(\hat{\theta}(\mathbf{Q}_{2012})) & \dots & Q_{100,2100}(\hat{\theta}(\mathbf{Q}_{2012})) & \dots \end{pmatrix} \quad (66)$$

Persze lehetne újabb előrevetítéseket nézni, de ennek nincs értelme, hiszen egy újragenerált minta nem feltétlen ugyanazzal az információval fog rendelkezni, mint a Human Mortality Database online felületéről szerzett halálzási valószínűsősokaság. Nemhiába kapnánk más és más paraméterbecslést. Korábban nem volt arra szükség, hogy az előrejelzéseknél mutassuk, mely paraméterbecslésekből származnak a várható élettartamok közelítései, most viszont az érthetőség kedvéért meg kell különböztetni azokat. Összességében tehát a modell változik, így a $\xi_{(2012+t)}^{(n)}$ valószínűségi változót át kell értelmezni a duplafázisú rendszernek megfelelően, ezért bevezetjük az alábbi jelöléseket: $\tilde{\xi}_{(2012+t)}^{(n)}(\hat{\theta})$ és $\tilde{\xi}_{(2012+t)}^{(n)}(\hat{\theta}(\mathbf{Q}_{2012}))$ jelentsék azokat változókat, amelyek $(2012+t)$ évben az n éves egyének halálukig tartó tartamait magyarázza. Míg az elsónél a bázisidőszakon vett újragenerálást 2013-tól az eredeti $\hat{\theta}$ paraméterezéssel folytatjuk, addig a másíknál a 2012-ig tartó bázisidőszak(ok)on maximum likelihood becsléssel újraszámolt paraméterekkel hajtjuk végre a

2013-tól vett halálozási valószínűségek generálását. Fontos leszögezni, hogy átírodtak a bázisidőszakok halálozási rátái, ennek a részfejezetnek nem célja a magyarországi előrejelzések pontosítása. Nem is szükségszerű, hogy a második fázisba történő átlépés ideje éppen 2012-re essen. Ezt a referenciaévet önkényesen választottuk ki, sőt így visszamenőleg $t \in \{-1, -2, \dots\}$ -re is a helyzetnek megfelelően értelmezhető $\tilde{\xi}_{(2012+t)}^{(n)}(\theta)$ ($\theta \in \{\hat{\theta}, \hat{\theta}(\mathbf{Q}_{2012})\}$) változót kapunk egészen az egyes bázisidőszakok kezdőévéig. Az $e_{n,t}^{(i)}$ ($i \in \{0, 1, 2\}$) várható értékek átértelmezése pedig okafogyottá vált, mert az \mathcal{F}_{2012} információnak kizárólagos magyarországi vonatkozása volt a múltbeli halálozási adatokon. Az eredmények értelmezését a numerikus elemzések segítik, ezért érdemesnek mutatkozik az előző fejezetekben is használt Monte-Carlo-módszer. Ennek kapcsán a továbbiakban az r . replikáció és r . forgatókönyv kifejezéseket szinonimaként is használjuk azzal a különbséggel, hogy az egymástól független forgatókönyvek mögött bonyolultabb struktúrák húzódnak meg. Lássuk tehát egy sematikus ábrán, hogy kellően nagy, N darabszámú forgatókönyv szimulációjánál az egyes forgatókönyvekben pontosan milyen lépéseket hajtunk végre:



Arra vagyunk kíváncsiak, hogy $\mathbb{E} \left(\tilde{\xi}_{(2012+t)}^{(n)}(\hat{\theta}(\mathbf{Q}_{2012}^{(r)})) \mid \mathbf{Q}^{(r)}(\hat{\theta}(\mathbf{Q}_{2012}^{(r)})) \right)$ várhatóan mennyire fog torzítani $\mathbb{E} \left(\tilde{\xi}_{(2012+t)}^{(n)}(\hat{\theta}) \mid \mathbf{Q}^{(r)}(\hat{\theta}) \right)$ -hoz képest, matematikailag megfogalmazva $\mathbb{E} \left(\tilde{\xi}_{(2012+t)}^{(n)}(\hat{\theta}) \mid \mathbf{Q}^{(r)}(\hat{\theta}) \right) - \mathbb{E} \left(\tilde{\xi}_{(2012+t)}^{(n)}(\hat{\theta}(\mathbf{Q}_{2012}^{(r)})) \mid \mathbf{Q}^{(r)}(\hat{\theta}(\mathbf{Q}_{2012}^{(r)})) \right)$ eloszlását kívánom meghatározni. 5000 db. forgatókönyvet készítettünk nőkre, férfiakra és a teljes lakosságra egyaránt, majd a 2018-as évre ($t = 6$ -ra) vonatkozóan születéskori ($n = 0$) és a 65 éves kor feletti ($n = 65$) esetekben mutatunk hisztogramot.⁹

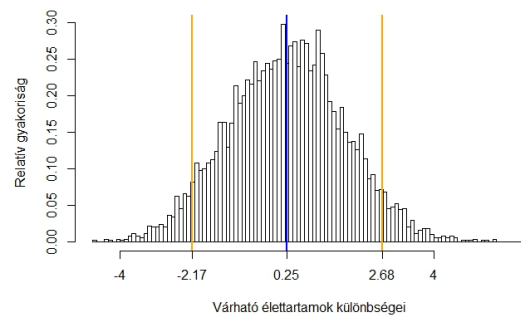
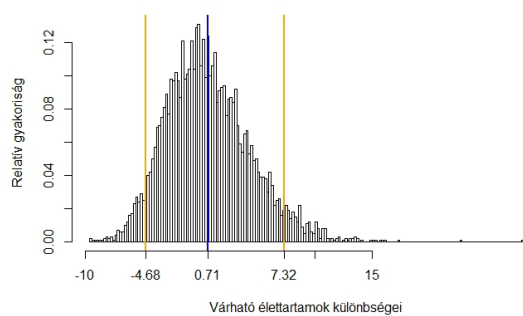
⁹ A generált mintákra vonatkozó valamennyi statisztikai mutatót a függelékben az II/(a), II/(b) és II/(c) táblázatok tartalmazzák.

($t = 6$)

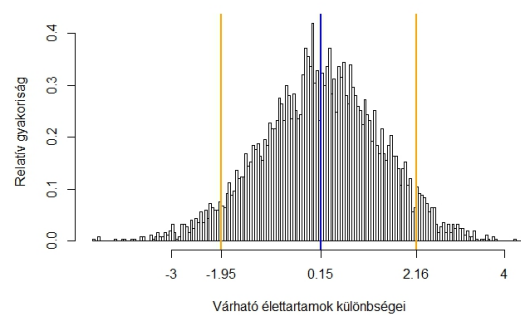
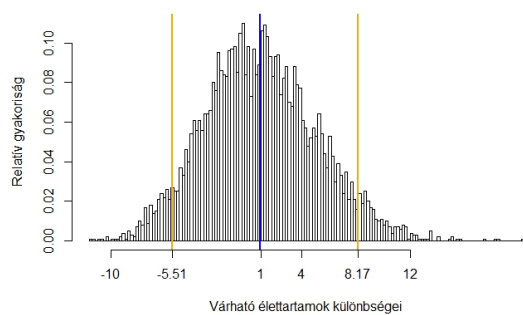
Születéskori ($n = 0$)

65 éves korra ($n = 65$)

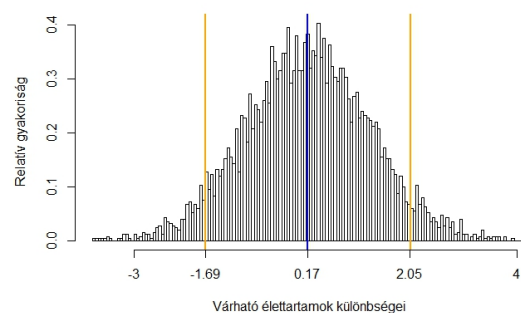
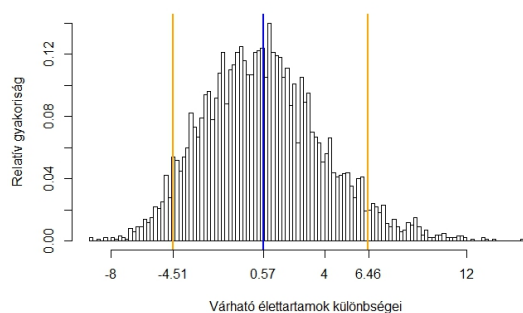
Nőkre



Férfiakra



Unisex táblára



A születéskori hisztogramoknál az 5%-os és a 95%-os kvantilisokhoz (amiket a sárga vertikális vonalak jeleznek) tartozó értékek különbségei átlagosan 13 év, míg a 65 éveseknél kevesebb, mint 5 év. Ez a jelenség azzal magyarázható, hogy 65 éves korban természetesen sokkal kisebb a hátralévő élettartam, mint születésnél. Látható, hogy a születéskori várható élettartamok különbségeinek az átlaga (amit kék vertikális vonal jelez) 0,5 év és 1 év között alakul, míg az idősebbeknél 0,15 év és 0,25 év

között. A torzítás várható értékben tehát nem jelentős, amely sejtésünk szerint nem lehet több 1 évnél (a 2018-as évre vonatkoztatva), mivel a születéskortól vett összes halandósági ráta tartalmazza mindegyik individuális és véletlenbolyongásos változót. Másképpen minél nagyobb az egyén életkora, annál kisebb a különbség. Mi több, a várható értékek pozitív átlagának volta miatt úgy tűnik, hogy várhatóan $\mathbb{E} \left(\tilde{\xi}_{(2012+t)}^{(n)}(\hat{\theta}) \mid \mathbf{Q}^{(r)}(\hat{\theta}) \right)$ felső határa $\mathbb{E} \left(\tilde{\xi}_{(2012+t)}^{(n)}(\hat{\theta}) \mid \mathbf{Q}^{(r)} \left(\hat{\theta} \left(\mathbf{Q}_{2012}^{(r)} \right) \right) \right)$ -nek. Elmondható tehát, hogy a modell a maximum likelihood becslésre nézve megbízható paraméterekkel szolgál figyelembe véve a második fázisban generált várható élettartamok torzítását.

4. Az eredmények összegzése

A dolgozat tehát az élettartam-kockázatokat vette górcső alá: miként lehet azokat modellezni és milyen eredményekkel számolhatunk be. Az alapfogalmak és a Lee–Carter-modell struktúrájának részletezése után sikerült általános keretek közt megkapni a szóban forgó modelltől n éves egyénekre $2012 + t$ év múlva várható élettartamuk explicit alakjait $\left(e_{n,t}^{(0)} = \mathbb{E}\left(\xi_{(2012+t)}^{(n)} \mid \mathcal{F}_{2012}\right)\right)$ anélkül, hogy kiválasztottuk volna a megfelelő bázisidőszakot. Fiatalkori élettartamok esetén a (25) egyenlet rendkívül bonyolult, ezért kísérletet tettünk arra, hogy azt egyszerűbb alakra hozzuk $\left(e_{n,t}^{(1)}\right)$. Felvázoltuk, hogy a gyakorlatban milyen módszert használnak az explicit alak közelítéséhez $\left(e_{n,t}^{(2)}\right)$. A matematikai formulák meghatározása után numerikus számításokat végeztünk, hogy megtudjuk, az előbb említett várható élettartamok milyen relációba hozhatók egymással. Magyarországi adatokra nézve a BAJKÓ–MAKNICS–TÓTH–VÉKÁS [4] cikk ajánlásával nőkre és férfiakra rendre az 1950-től illetve az 1980-tól 2012-ig tartó időszakot vettük, saját számításokkal χ^2 -próba alapján a teljes lakosságra nézve pedig az [1980, 2012]-es intervallumot. Éves, majd korcsoportos bontásban nemenként ábrázoltuk az individuális reziduálisokat, ahol észrevettük, hogy az 50 év alattiak szórása lényegesen elüt az idősebbekétől, ezért célszerűnek tartottunk felírni egy ún. módosított Lee–Carter-modellt. A korábban kapott χ^2 -próbában az elfogadási és elutasítási eredményeket összehasonlítva arra következtetésre jutottunk, hogy nem szükséges kiterjeszteni modellt, elegendőnek mutatkozik az eredeti. Az eredeti modellen nemenként, a megfelelő bázisidőszakokra születéskori és 65 évesek esetében a 2018-tól 2117-ig tartó időszakban a modellbéli várható értéket a saját közelítésünk alul-, a gyakorlatban vett közelítés pedig fölülbecsli, vagyis $e_{0,t}^{(1)} < e_{0,t}^{(0)} < e_{0,t}^{(2)}$ és $e_{65,t}^{(1)} < e_{65,t}^{(0)} < e_{65,t}^{(2)}$ ($\forall t \in \{6, \dots, 106\}$). A gyakorlatban használt módszer viszont negyedévnél is kevesebb torzítással becsli a tényleges várható élettartamokat, mint a sajátunk, ahol éven túli különbségeket is realizáltunk. Végül a kétfázisú forgatókönyvek generálásával megállapítottuk, hogy a Lee–Carter-modellben a paraméterbecsléseknél a maximum likelihood elv elhanyagolható mértékben torzít a várható élettartamra nézve, ami a kor előrehaladtával csökken. Sőt az eredeti illetve a 2012-ben újrabecslült paraméterekkel generált halandósági valószínűségekből számolt várható értékek $\left(\text{rendre } \mathbb{E}\left(\tilde{\xi}_{(2012+t)}^{(n)}(\hat{\theta}) \mid \mathbf{Q}^{(r)}(\hat{\theta})\right) \text{ és } \mathbb{E}\left(\tilde{\xi}_{(2012+t)}^{(n)}\left(\hat{\theta}\left(\mathbf{Q}_{2012}^{(r)}\right)\right) \mid \mathbf{Q}^{(r)}\left(\hat{\theta}\left(\mathbf{Q}_{2012}^{(r)}\right)\right)\right)\right)$ különbségeinek az átlaga pozitív. Azonban láthattuk azt is, hogy az előrejelzési intervallumok nagyon szélesek. Például a 35. oldalon található hisztogramok egyikéről leolvasható, hogy a 65 éves nők várható hátralévő élettartamára a 2018-as évre vonatkozó előrejelzés 18,73 év, viszont a 90%-os megbízhatóságú intervallum [17, 54 (év); 19, 91 (év)]. Ez azt mutatja, hogy ezeket az előrejelzéseket csak nagyon óvatosan és megfontoltan szabad figyelembe venni fontos döntések meghozatalához.

Hivatkozások

- [1] *Human Mortality Database*. University of California, Berkeley (USA), és Max Planck Institute for Demographic Research (Germany). (A felhasznált mortalitási adatokat 2017.10.01-én töltöttem le.) www.mortality.org, www.humanmortality.de
- [2] VÉKÁS Péter (2016). *Az élettartam kockázat modellezése*, doktori értekezés. Budapesti Corvinus Egyetem Általános és Kvantitatív Közgazdaságtan Doktori Iskola. Budapest.
<http://dx.doi.org/10.14267/phd.2017018>, letöltés dátuma: 2018.04.26.
- [3] MÁJER István & Dr. KOVÁCS Erzsébet (2011). *Élettartam-kockázat – a nyugdíjrendszerre nehezedő egyik teher*. Statisztikai Szemle, 89. évfolyam 7-8. szám: 790-812.
http://www.ksh.hu/statszemle_archive/2011/2011_07-08/2011_07-08_790.pdf, letöltés dátuma: 2018.04.26.
- [4] BAJKÓ Attila, MAKNICS Anita, TÓTH Krisztián & VÉKÁS Péter (2015). *A magyar nyugdíjrendszer fenntarthatóságáról*. Közgazdasági Szemle, 62. évfolyam, december: 1229-1257.
<http://dx.doi.org/10.18414/KSZ.2015.12.1229>, letöltés dátuma: 2018.04.26.
- [5] LEE, Ronald D. & CARTER, Lawrence R. (1992). *Modeling and forecasting U.S. mortality*. Journal of the American Statistical Association, 87:659–671.
<http://dx.doi.org/10.2307/2290201>, letöltés dátuma: 2018.04.26.
- [6] LEE, Ronald D. (2000). *The Lee–Carter Method for Forecasting Mortality, with Various Extensions and Applications*. North American Actuarial Journal. 4. 10.1080/10920277.2000.10595882.
<https://doi.org/10.1080/10920277.2000.10595883>, letöltés dátuma: 2018.04.26.
- [7] BARAN S., GÁLL J., ISPÁNY M. & PAP G. (2007). *Forecasting Hungarian mortality rates using the Lee-Carter method*. Acta Oeconomica, Vol. 57 (1), 21–34.
<http://doi.org/10.1556/A0econ.57.2007.1.3>, letöltés dátuma: 2018.04.26.
- [8] VILLEGAS, A. M., KAISHEV, V. & MILLOSOVICH, P. (2016). *StMoMo: An R Package for Stochastic Mortality Modelling*.
<https://cran.r-project.org/web/packages/StMoMo/vignettes/StMoMoVignette.pdf>, letöltés dátuma: 2018.04.26.

5. Függelék

I/(a) táblázat:

$e_{n,t}^{(0)}$ várható élettartamokból a generált mintákra
vonatkozó leíró statisztikák nőkre 2018-ban

Leíró statisztikák	Születés kori ($n = 0$)	65 éves korra ($n = 65$)
minimum	79,51	15,77
maximum	89,88	21,20
átlag	85,16	18,73
szórás	1,50	0,72
terjedelem	10,37	5,43
5%-os kvantilis	82,55	17,54
25%-os kvantilis	84,20	18,25
50%-os kvantilis	85,25	18,72
75%-os kvantilis	86,19	19,20
95%-os kvantilis	87,52	19,91
ferdeség	-0,25	-0,02
csúcsosság	3,01	2,99

I/(b) táblázat:

$e_{n,t}^{(0)}$ várható élettartamokból generált mintákra
vonatkozó leíró statisztikák férfiakra 2018-ban

Leíró statisztikák	Születés kori ($n = 0$)	65 éves korra ($n = 65$)
minimum	74,16	13,04
maximum	85,23	17,65
átlag	79,65	15,25
szórás	1,55	0,57
terjedelem	11,06	4,61
5%-os kvantilis	77,06	14,33
25%-os kvantilis	78,63	14,87
50%-os kvantilis	79,67	15,26
75%-os kvantilis	80,71	15,64
95%-os kvantilis	82,16	16,18
ferdeség	-0,06	0,01
csúcsosság	3,04	3,04

I/(c) táblázat:

$e_{n,t}^{(0)}$ várható élettartamokból generált mintákra
 vonatkozó leíró statisztikák a teljes lakosságra 2018-ban

Leíró statisztikák	Születéskori ($n = 0$)	65 éves korra ($n = 65$)
minimum	78,69	15,63
maximum	87,76	19,95
átlag	83,52	17,77
szórás	1,33	0,57
terjedelem	9,67	4,32
5%-os kvantilis	81,26	16,83
25%-os kvantilis	82,63	17,38
50%-os kvantilis	83,55	17,78
75%-os kvantilis	84,46	18,15
95%-os kvantilis	85,65	18,70
ferdeség	-0,16	-0,01
csúcosság	2,93	3,06

II/(a) táblázat:

Egy- és kétfázisú várható élettartamok különbségeiből generált
 mintákra vonatkozó leíró statisztikák nőkre 2018-ban

Leíró statisztikák	Születéskori ($n = 0$)	65 éves korra ($n = 65$)
minimum	-9,56	-4,61
maximum	28,04	6,34
átlag	0,71	0,25
szórás	3,69	1,48
terjedelem	37,60	10,95
5%-os kvantilis	-4,68	-2,17
25%-os kvantilis	-1,92	-0,77
50%-os kvantilis	0,35	0,26
75%-os kvantilis	2,95	1,21
95%-os kvantilis	7,32	2,68
ferdeség	0,63	0,05
csúcosság	4,16	2,96

II/(b) táblázat:

Egy- és kétfázisú várható élettartamok különbségeiből generált mintákra vonatkozó leíró statisztikák férfiakra 2018-ban

Leíró statisztikák	Születés kori ($n = 0$)	65 éves korra ($n = 65$)
minimum	-11,43	-4,61
maximum	20,36	4,47
átlag	1,00	0,15
szórás	4,12	1,25
terjedelem	31,79	9,07
5%-os kvantilis	-5,51	-1,95
25%-os kvantilis	-1,88	-0,68
50%-os kvantilis	0,84	0,17
75%-os kvantilis	3,68	0,99
95%-os kvantilis	8,17	2,16
ferdeség	0,28	-0,12
csúcsosság	3,12	3,02

II/(c) táblázat:

Egy- és kétfázisú várható élettartamok különbségeiből generált mintákra vonatkozó leíró statisztikák a teljes lakosságra 2018-ban

Leíró statisztikák	Születés kori ($n = 0$)	65 éves korra ($n = 65$)
minimum	-9,06	-3,73
maximum	15,03	4,15
átlag	0,57	0,17
szórás	3,36	1,14
terjedelem	24,10	7,89
5%-os kvantilis	-4,51	-1,69
25%-os kvantilis	-1,81	-0,57
50%-os kvantilis	0,39	0,18
75%-os kvantilis	2,71	0,93
95%-os kvantilis	6,46	2,05
ferdeség	0,42	(-)0,00
csúcsosság	3,20	3,08

Forráskód

```
#####  
#  
# ---- Numerical analyzis for the diploma thesis ----- #  
# ---- (ERRORS OF MORTALITY PROJECTIONS) ----- #  
#  
# ---- Made by: Benyi Gabor ----- #  
#  
#  
#####  
  
#####  
#  
# ---- (SUB) SECTION: (4.2) ---- #  
#  
#  
#####  
  
# setting destination folder:  
# -----  
setwd("D:\\Tanulmányok\\BCE-ELTE\\2017_2018_1_félév\\Szakszeminárium,  
kutatómódszertan")  
  
# reading mortality datas (Hungarian datas (1950-2012 1x1 is the  
necessary period) are available at www.mortality.org or www.ksh.hu):  
# -----  
-----  
data_mortality_female<-read.csv("Mortality_female.csv", header=TRUE,  
sep=";")  
data_mortality_male<-read.csv("Mortality_male.csv", header=TRUE, sep=";")  
data_mortality_unisex<-read.csv("Mortality_unisex.csv", header=TRUE,  
sep=";")  
  
# creating 'demogdata' objects from mortality datas:  
# -----  
### creating mortality probability matrixes from mortality datas:  
### -----  
matrix_mortality_female<-matrix(data_mortality_female$qx ,nrow=111)  
matrix_mortality_male<-matrix(data_mortality_male$qx ,nrow=111)  
matrix_mortality_unisex<-matrix(data_mortality_unisex$qx ,nrow=111)  
  
### creating exposure matrixes from mortality datas:  
### -----  
matrix_exposure_female<-matrix(data_mortality_female$Lx ,nrow=111)  
matrix_exposure_male<-matrix(data_mortality_male$Lx ,nrow=111)  
matrix_exposure_unisex<-matrix(data_mortality_unisex$Lx ,nrow=111)  
  
### creating vectors from mortality datas including years:  
### -----  
vector_ages<-0:110  
vector_years<-  
type.convert(attributes(factor(data_mortality_female$Year))$levels)  
is.vector(vector_years<-  
type.convert(attributes(factor(data_mortality_female$Year))$levels))  
  
### installing 'demogdata' package or reading that from the library:
```

```

### -----
install.packages("demography")
library("demography")

### at last creating 'demogdata' objects:
### -----
demogdata_mortality_female<-demogdata(data=matrix_mortality_female,
pop=matrix_exposure_female, ages=vector_ages, years=vector_years,
type="mortality", label="Human Mortality Database", name="female",
lambda=0)
demogdata_mortality_male<-demogdata(data=matrix_mortality_male,
pop=matrix_exposure_male, ages=vector_ages, years=vector_years,
type="mortality", label="Human Mortality Database", name="male",
lambda=0)
demogdata_mortality_unisex<-demogdata(data=matrix_mortality_unisex,
pop=matrix_exposure_unisex, ages=vector_ages, years=vector_years,
type="mortality", label="Human Mortality Database", name="female",
lambda=0)

# building up the Lee--Carter-model:
# -----

### creating 'lca' objects for women (this is the best fitting period
from article [4]):
### -----
lca_female_1950_2012<-lca(demogdata_mortality_female,years=(1950:2012),
adjust = "none")

### creating 'lca' objects for men (this is the best fitting period from
article [4]):
### -----
lca_male_1980_2012<-lca(demogdata_mortality_male,years=(1980:2012),
adjust = "none")

### creating 'lca' objects for unisex table trying to choose the best
fitting period (because there is no fitting period in article [4]):
### -----
lca_unisex_1950_2000<-lca(demogdata_mortality_unisex,years=(1950:2000),
adjust = "none")
lca_unisex_1960_2000<-lca(demogdata_mortality_unisex,years=(1960:2000),
adjust = "none")
lca_unisex_1970_2000<-lca(demogdata_mortality_unisex,years=(1970:2000),
adjust = "none")
lca_unisex_1980_2000<-lca(demogdata_mortality_unisex,years=(1980:2000),
adjust = "none")
lca_unisex_1989_2000<-lca(demogdata_mortality_unisex,years=(1989:2000),
adjust = "none")

### creating function 'ML_becslesek' from 'lca' objects in order to
estimate the standard deviations from the individual variables and the
random walk process and the drift length (using equations (12), (13)):
### -----
-----
ML_becslesek<-function(lca){

```



```

matrix_residuals<-lca$residuals$y

length_row<-length(matrix_residuals[,1])
length_column<-length(matrix_residuals[1,])

szigma_epszilon_negyzet<-(t(rep(1,
times=length_row))%*%matrix_residuals^2*%*rep(1,times=length_column))[1,1]
)/(length_row*length_column)

c<-(lca$kt[length(lca$kt)]-lca$kt[1])/(length(lca$kt)-1)

szigma_Z_negyzet<-0
for(i in seq(length(lca$kt)-1)){
  szigma_Z_negyzet = szigma_Z_negyzet + (lca$kt[i+1]-lca$kt[i]-
c)^2/(length(lca$kt)-1)
}
ML_parameterek<-c(c,szigma_epszilon_negyzet,szigma_Z_negyzet)
names(ML_parameterek)<-
c("c","szigma_epszilon_negyzet","szigma_Z_negyzet")
return(ML_parameterek)
}

### projecting unisex mortality probabilities (using equation (39)):
### -----
forecast_mortality_unisex_1950<-
matrix(nrow=length(lca_unisex_1950_2000$ax),ncol=12)
forecast_mortality_unisex_1960<-
matrix(nrow=length(lca_unisex_1960_2000$ax),ncol=12)
forecast_mortality_unisex_1970<-
matrix(nrow=length(lca_unisex_1970_2000$ax),ncol=12)
forecast_mortality_unisex_1980<-
matrix(nrow=length(lca_unisex_1980_2000$ax),ncol=12)
forecast_mortality_unisex_1989<-
matrix(nrow=length(lca_unisex_1989_2000$ax),ncol=12)
### -----
for(i in seq(length(lca_unisex_1950_2000$ax))){
  ML_becslesek_unisex<-ML_becslesek(lca_unisex_1950_2000)
  for(j in seq(12)){
    forecast_mortality_unisex_1950[i,j]<-
exp(lca_unisex_1950_2000$ax[i]+lca_unisex_1950_2000$bx[i]*(lca_unisex_1950_2000$kt[length(lca_unisex_1950_2000$kt)]+j*ML_becslesek_unisex["c"])+j*(lca_unisex_1950_2000$bx[i])^2*ML_becslesek_unisex["szigma_Z_negyzet"])/2+ML_becslesek_unisex["szigma_epszilon_negyzet"]/2)
  }
}
for(i in seq(length(lca_unisex_1960_2000$ax))){
  ML_becslesek_unisex<-ML_becslesek(lca_unisex_1960_2000)
  for(j in seq(12)){
    forecast_mortality_unisex_1960[i,j]<-
exp(lca_unisex_1960_2000$ax[i]+lca_unisex_1960_2000$bx[i]*(lca_unisex_1960_2000$kt[length(lca_unisex_1960_2000$kt)]+j*ML_becslesek_unisex["c"])+j*(lca_unisex_1960_2000$bx[i])^2*ML_becslesek_unisex["szigma_Z_negyzet"])/2+ML_becslesek_unisex["szigma_epszilon_negyzet"]/2)
  }
}
for(i in seq(length(lca_unisex_1970_2000$ax))){
  ML_becslesek_unisex<-ML_becslesek(lca_unisex_1970_2000)
  for(j in seq(12)){

```

```

    forecast_mortality_unisex_1970[i,j]<-
exp(lca_unisex_1970_2000$ax[i]+lca_unisex_1970_2000$bx[i]*(lca_unisex_197
0_2000$kt[length(lca_unisex_1970_2000$kt)]+j*ML_becslesek_unisex["c"])+j*
(lca_unisex_1970_2000$bx[i])^2*ML_becslesek_unisex["szigma_Z_negyzet"])/2+
ML_becslesek_unisex["szigma_epsilon_negyzet"]/2)
}
}
for(i in seq(length(lca_unisex_1980_2000$ax))){
  ML_becslesek_unisex<-ML_becslesek(lca_unisex_1980_2000)
  for(j in seq(12)){
    forecast_mortality_unisex_1980[i,j]<-
exp(lca_unisex_1980_2000$ax[i]+lca_unisex_1980_2000$bx[i]*(lca_unisex_198
0_2000$kt[length(lca_unisex_1980_2000$kt)]+j*ML_becslesek_unisex["c"])+j*
(lca_unisex_1980_2000$bx[i])^2*ML_becslesek_unisex["szigma_Z_negyzet"])/2+
ML_becslesek_unisex["szigma_epsilon_negyzet"]/2)
  }
}
for(i in seq(length(lca_unisex_1989_2000$ax))){
  ML_becslesek_unisex<-ML_becslesek(lca_unisex_1989_2000)
  for(j in seq(12)){
    forecast_mortality_unisex_1989[i,j]<-
exp(lca_unisex_1989_2000$ax[i]+lca_unisex_1989_2000$bx[i]*(lca_unisex_198
9_2000$kt[length(lca_unisex_1989_2000$kt)]+j*ML_becslesek_unisex["c"])+j*
(lca_unisex_1989_2000$bx[i])^2*ML_becslesek_unisex["szigma_Z_negyzet"])/2+
ML_becslesek_unisex["szigma_epsilon_negyzet"]/2)
  }
}
}

```

creating scatters from the differences between the projected and the real mortality probabilities:

```

scatters_unisex_1950<-matrix(ncol=3,nrow=12*101)
colnames(scatters_unisex_1950)<-c("Kor", "Év", "Különbség")
for(j in seq(12)){
  for(i in seq(101)){
    scatters_unisex_1950[(j-1)*101+i,1]<-(i-1)
    scatters_unisex_1950[(j-1)*101+i,2]<-(2000+j)
    scatters_unisex_1950[(j-1)*101+i,3]<-
(forecast_mortality_unisex_1950[i,j]-matrix_mortality_unisex[i,51+j])
  }
}
scatters_unisex_1960<-matrix(ncol=3,nrow=12*101)
colnames(scatters_unisex_1960)<-c("Kor", "Év", "Különbség")
for(j in seq(12)){
  for(i in seq(101)){
    scatters_unisex_1960[(j-1)*101+i,1]<-(i-1)
    scatters_unisex_1960[(j-1)*101+i,2]<-(2000+j)
    scatters_unisex_1960[(j-1)*101+i,3]<-
(forecast_mortality_unisex_1960[i,j]-matrix_mortality_unisex[i,51+j])
  }
}
scatters_unisex_1970<-matrix(ncol=3,nrow=12*101)
colnames(scatters_unisex_1970)<-c("Kor", "Év", "Különbség")
for(j in seq(12)){
  for(i in seq(101)){
    scatters_unisex_1970[(j-1)*101+i,1]<-(i-1)
    scatters_unisex_1970[(j-1)*101+i,2]<-(2000+j)

```

```

    scatters_unisex_1970[(j-1)*101+i,3]<-
(forecast_mortality_unisex_1970[i,j]-matrix_mortality_unisex[i,51+j])
  }
}
scatters_unisex_1980<-matrix(ncol=3,nrow=12*101)
colnames(scatters_unisex_1980)<-c("Kor","Év","Különbség")
for(j in seq(12)){
  for(i in seq(101)){
    scatters_unisex_1980[(j-1)*101+i,1]<-(i-1)
    scatters_unisex_1980[(j-1)*101+i,2]<-(2000+j)
    scatters_unisex_1980[(j-1)*101+i,3]<-
(forecast_mortality_unisex_1980[i,j]-matrix_mortality_unisex[i,51+j])
  }
}
scatters_unisex_1989<-matrix(ncol=3,nrow=12*101)
colnames(scatters_unisex_1989)<-c("Kor","Év","Különbség")
for(j in seq(12)){
  for(i in seq(101)){
    scatters_unisex_1989[(j-1)*101+i,1]<-(i-1)
    scatters_unisex_1989[(j-1)*101+i,2]<-(2000+j)
    scatters_unisex_1989[(j-1)*101+i,3]<-
(forecast_mortality_unisex_1989[i,j]-matrix_mortality_unisex[i,51+j])
  }
}

```

```
### installing 'scatterplot3d' package or reading that from the library:
```

```
### -----
install.packages("scatterplot3d")
library("scatterplot3d")

```

```
### plotting differences between the projected and the real mortality
probabilities:
```

```
### -----
-----
colors<-NULL
for(i in seq(12*101)){
  colors[i]<-paste("gray",toString(60-(scatters_unisex_1950[i,2]-
2001)*5))
}
par(mfrow=c(1,2))
scatterplot3d(x=scatters_unisex_1950,pch=16,color=colors,box=FALSE,grid=T
RUE,main="1950-2000-es bázisidőszakra\n vonatkozó
eltérések",lab=c(5,2,5))
scatterplot3d(x=scatters_unisex_1960,pch=16,color=colors,box=FALSE,grid=T
RUE,main="1960-2000-es bázisidőszakra\n vonatkozó
eltérések",lab=c(5,2,5))
scatterplot3d(x=scatters_unisex_1970,pch=16,color=colors,box=FALSE,grid=T
RUE,main="1970-2000-es bázisidőszakra\n vonatkozó
eltérések",lab=c(5,2,5))
scatterplot3d(x=scatters_unisex_1980,pch=16,color=colors,box=FALSE,grid=T
RUE,main="1980-2000-es bázisidőszakra\n vonatkozó
eltérések",lab=c(5,2,5))
par(mfrow=c(1,1))
scatterplot3d(x=scatters_unisex_1989,pch=16,color=colors,box=FALSE,grid=T
RUE,main="1989-2000-es bázisidőszakra\n vonatkozó
eltérések",lab=c(5,2,5))

```

```
### plotting 'b(x) '-s on each basis period:
```

```

### -----
scatters_bx<-matrix(ncol=3,nrow=5*101)
colnames(scatters_bx)<-c("Kor","Bázisidőszak","b(x)")
for(j in seq(5)){
  for(i in seq(101)){
    scatters_bx[(j-1)*101+i,1]<-(i-1)
    if(j==1){
      scatters_bx[(j-1)*101+i,2]<-1950
      scatters_bx[(j-1)*101+i,3]<-lca_unisex_1950_2000$bx[i]
    }
    if(j==2){
      scatters_bx[(j-1)*101+i,2]<-1960
      scatters_bx[(j-1)*101+i,3]<-lca_unisex_1960_2000$bx[i]
    }
    if(j==3){
      scatters_bx[(j-1)*101+i,2]<-1970
      scatters_bx[(j-1)*101+i,3]<-lca_unisex_1970_2000$bx[i]
    }
    if(j==4){
      scatters_bx[(j-1)*101+i,2]<-1980
      scatters_bx[(j-1)*101+i,3]<-lca_unisex_1980_2000$bx[i]
    }
    if(j==5){
      scatters_bx[(j-1)*101+i,2]<-1989
      scatters_bx[(j-1)*101+i,3]<-lca_unisex_1989_2000$bx[i]
    }
  }
}
colors_bx<-NULL
for(i in seq(5*101)){
  colors_bx[i]<-paste("gray",toString(60-round((scatters_bx[i,2]-
1950)*1.5)))
}
scatterplot3d(x=scatters_bx,pch=16,color=colors_bx,box=FALSE,grid=TRUE,mai
in="A mortalitási indexhez tartozó szorzók (b(x))\n koronkénti alakulása
az egyes bázisidőszakokban",lab=c(5,2,5))

### creating vectors of khi-squared tests (using equation (37)):
### -----
khi_negyzet_1950<-rep(0,13)
for(j in seq(12)){
  for(i in seq(25)){
    khi_negyzet_1950[j]<-
khi_negyzet_1950[j]+(matrix_exposure_unisex[i,51+j]*matrix_mortality_unis
ex[i,51+j]-
matrix_exposure_unisex[i,51+j]*forecast_mortality_unisex_1950[i,j])^2/(ma
trix_exposure_unisex[i,51+j]*forecast_mortality_unisex_1950[i,j]*(1-
forecast_mortality_unisex_1950[i,j]))
  }
}
khi_negyzet_1950[13]<-sum(khi_negyzet_1950)
khi_negyzet_1950
### -----
khi_negyzet_1960<-rep(0,13)
for(j in seq(12)){
  for(i in seq(25)){
    khi_negyzet_1960[j]<-
khi_negyzet_1960[j]+(matrix_exposure_unisex[i,51+j]*matrix_mortality_unis

```

```

ex[i,51+j]-
matrix_exposure_unisex[i,51+j]*forecast_mortality_unisex_1960[i,j])^2/(ma
trix_exposure_unisex[i,51+j]*forecast_mortality_unisex_1960[i,j]*(1-
forecast_mortality_unisex_1960[i,j]))
}
}
khi_negyzet_1960[13]<-sum(khi_negyzet_1960)
khi_negyzet_1960
### -----
khi_negyzet_1970<-rep(0,13)
for(j in seq(12)){
  for(i in seq(25)){
    khi_negyzet_1970[j]<-
khi_negyzet_1970[j]+(matrix_exposure_unisex[i,51+j]*matrix_mortality_unis
ex[i,51+j]-
matrix_exposure_unisex[i,51+j]*forecast_mortality_unisex_1970[i,j])^2/(ma
trix_exposure_unisex[i,51+j]*forecast_mortality_unisex_1970[i,j]*(1-
forecast_mortality_unisex_1970[i,j]))
  }
}
khi_negyzet_1970[13]<-sum(khi_negyzet_1970)
khi_negyzet_1970
### -----
khi_negyzet_1980<-rep(0,13)
for(j in seq(12)){
  for(i in seq(25)){
    khi_negyzet_1980[j]<-
khi_negyzet_1980[j]+(matrix_exposure_unisex[i,51+j]*matrix_mortality_unis
ex[i,51+j]-
matrix_exposure_unisex[i,51+j]*forecast_mortality_unisex_1980[i,j])^2/(ma
trix_exposure_unisex[i,51+j]*forecast_mortality_unisex_1980[i,j]*(1-
forecast_mortality_unisex_1980[i,j]))
  }
}
khi_negyzet_1980[13]<-sum(khi_negyzet_1980)
khi_negyzet_1980
### -----
khi_negyzet_1989<-rep(0,13)
for(j in seq(12)){
  for(i in seq(25)){
    khi_negyzet_1989[j]<-
khi_negyzet_1989[j]+(matrix_exposure_unisex[i,51+j]*matrix_mortality_unis
ex[i,51+j]-
matrix_exposure_unisex[i,51+j]*forecast_mortality_unisex_1989[i,j])^2/(ma
trix_exposure_unisex[i,51+j]*forecast_mortality_unisex_1989[i,j]*(1-
forecast_mortality_unisex_1989[i,j]))
  }
}
khi_negyzet_1989[13]<-sum(khi_negyzet_1989)
khi_negyzet_1989

### calculating the 5% and 95% quantization of the 26 and 312 freedom-
rate khi-squared distribution:
### -----
-----
qchisq(.05, df=26)
qchisq(.95, df=26)
qchisq(.05, df=312)

```

```

qchisq(.95, df=312)

### making 'lca' objects for unisex table (this is the best fitting
period due to the considerations in the diploma thesis):
### -----
lca_unisex_1980_2012<-lca(demogdata_mortality_unisex,years=(1980:2012),
adjust = "none")

#####
# #
# ---- (SUB)SECTION: (4.3) ---- #
# #
#####

# calculating maximum likelihood estimates for men, women and unisex
table on the best fitting periods (using equations (12), (13) and
function 'ML_becselek'):
# -----
-----
ML_becselek_female_1950_2012<-ML_becselek(lca_female_1950_2012)
ML_becselek_male_1980_2012<-ML_becselek(lca_male_1980_2012)
ML_becselek_unisex_1980_2012<-ML_becselek(lca_unisex_1980_2012)
# -----
-----

szigma_epszilon_female<-
sqrt(ML_becselek_female_1950_2012["szigma_epszilon_negyzet"])
szigma_Z_female<-sqrt(ML_becselek_female_1950_2012["szigma_Z_negyzet"])
szigma_epszilon_male<-
sqrt(ML_becselek_male_1980_2012["szigma_epszilon_negyzet"])
szigma_Z_male<-sqrt(ML_becselek_male_1980_2012["szigma_Z_negyzet"])
szigma_epszilon_unisex<-
sqrt(ML_becselek_unisex_1980_2012["szigma_epszilon_negyzet"])
szigma_Z_unisex<-sqrt(ML_becselek_unisex_1980_2012["szigma_Z_negyzet"])
# -----
-----

names(szigma_epszilon_female)<-"szigma epszilon"
names(szigma_Z_female)<-"szigma Z"
names(szigma_epszilon_male)<-"szigma epszilon"
names(szigma_Z_male)<-"szigma Z"
names(szigma_epszilon_unisex)<-"szigma epszilon"
names(szigma_Z_unisex)<-"szigma Z"
# -----
-----

szigma_epszilon_female
szigma_Z_female
szigma_epszilon_male
szigma_Z_male
szigma_epszilon_unisex
szigma_Z_unisex

```

```

# constructing confidence intervals for individual residuals (using
equations (42)):
# -----
-----
confidence_interval_female_individual<-NULL
confidence_interval_female_individual[1]<-
mean(lca_female_1950_2012$residuals$y) -
qnorm(0.975,0,1)*ML_becslesek_female_1950_2012["szigma_epsilon_negyzet"]
/sqrt(length(lca_female_1950_2012$residuals$y))
confidence_interval_female_individual[2]<-
mean(lca_female_1950_2012$residuals$y)+qnorm(0.975,0,1)*ML_becslesek_fema
le_1950_2012["szigma_epsilon_negyzet"]/sqrt(length(lca_female_1950_2012$
residuals$y))
names(confidence_interval_female_individual)<-c("low limit","upper
limit")
confidence_interval_female_individual
# -----
-----
confidence_interval_male_individual<-NULL
confidence_interval_male_individual[1]<-
mean(lca_male_1980_2012$residuals$y) -
qnorm(0.975,0,1)*ML_becslesek_male_1980_2012["szigma_epsilon_negyzet"]/s
qrt(length(lca_male_1980_2012$residuals$y))
confidence_interval_male_individual[2]<-
mean(lca_male_1980_2012$residuals$y)+qnorm(0.975,0,1)*ML_becslesek_male_1
980_2012["szigma_epsilon_negyzet"]/sqrt(length(lca_male_1980_2012$residu
als$y))
names(confidence_interval_male_individual)<-c("low limit","upper limit")
confidence_interval_male_individual
# -----
-----
confidence_interval_unisex_individual<-NULL
confidence_interval_unisex_individual[1]<-
mean(lca_unisex_1980_2012$residuals$y) -
qnorm(0.975,0,1)*ML_becslesek_unisex_1980_2012["szigma_epsilon_negyzet"]
/sqrt(length(lca_unisex_1980_2012$residuals$y))
confidence_interval_unisex_individual[2]<-
mean(lca_unisex_1980_2012$residuals$y)+qnorm(0.975,0,1)*ML_becslesek_unis
ex_1980_2012["szigma_epsilon_negyzet"]/sqrt(length(lca_unisex_1980_2012$
residuals$y))
names(confidence_interval_unisex_individual)<-c("low limit","upper
limit")
confidence_interval_unisex_individual

# constructing confidence intervals for random walk residuals (using
equations (44)):
# -----
-----
confidence_interval_female_random_walk<-NULL
confidence_interval_female_random_walk[1]<-
mean(diff(lca_female_1950_2012$kt)) -
qnorm(0.975,0,1)*ML_becslesek_female_1950_2012["szigma_Z_negyzet"]/sqrt(l
ength(diff(lca_female_1950_2012$kt)))
confidence_interval_female_random_walk[2]<-
mean(diff(lca_female_1950_2012$kt))+qnorm(0.975,0,1)*ML_becslesek_female_
1950_2012["szigma_Z_negyzet"]/sqrt(length(diff(lca_female_1950_2012$kt)))
names(confidence_interval_female_random_walk)<-c("low limit","upper
limit")

```

```

confidence_interval_female_random_walk
# -----
-----
confidence_interval_male_random_walk<-NULL
confidence_interval_male_random_walk[1]<-
mean(diff(lca_male_1980_2012$kt))-
qnorm(0.975,0,1)*ML_becslesek_male_1980_2012["szigma_Z_negyzet"]/sqrt(len
gth(diff(lca_male_1980_2012$kt)))
confidence_interval_male_random_walk[2]<-
mean(diff(lca_male_1980_2012$kt))+qnorm(0.975,0,1)*ML_becslesek_male_1980
_2012["szigma_Z_negyzet"]/sqrt(length(diff(lca_male_1980_2012$kt)))
names(confidence_interval_male_random_walk)<-c("low limit","upper limit")
confidence_interval_male_random_walk
# -----
-----
confidence_interval_unisex_random_walk<-NULL
confidence_interval_unisex_random_walk[1]<-
mean(diff(lca_unisex_1980_2012$kt))-
qnorm(0.975,0,1)*ML_becslesek_unisex_1980_2012["szigma_Z_negyzet"]/sqrt(l
ength(diff(lca_unisex_1980_2012$kt)))
confidence_interval_unisex_random_walk[2]<-
mean(diff(lca_unisex_1980_2012$kt))+qnorm(0.975,0,1)*ML_becslesek_unisex
_1980_2012["szigma_Z_negyzet"]/sqrt(length(diff(lca_unisex_1980_2012$kt)))
names(confidence_interval_unisex_random_walk)<-c("low limit","upper
limit")
confidence_interval_unisex_random_walk

# plotting individual residuals by years:
# -----
ts_res_female_by_years<-NULL
labels_ts_res_female_by_years<-seq(from=1950, to=2012)
at_ts_res_female_by_years<-NULL
for(j in seq(63)){
  for(i in seq(101)){
    if(i==1){
      at_ts_res_female_by_years[j]<-(j-1)*101+i
    }
    ts_res_female_by_years[(j-1)*101+i]<-
lca_female_1950_2012$residuals$y[i,j]
  }
}
plot(ts_res_female_by_years, xaxt = "n", main="Női reziduálisok éves
bontásokban", xlab="Év", ylab="")
axis(1, at=at_ts_res_female_by_years,
labels=labels_ts_res_female_by_years)
abline(lwd=2,h=sqrt(ML_becslesek(lca_female_1950_2012) ["szigma_epsilon_n
egyzet"]), col="orange")
abline(lwd=2,h=-
sqrt(ML_becslesek(lca_female_1950_2012) ["szigma_epsilon_negyzet"]), col="
orange")
# -----
ts_res_male_by_years<-NULL
labels_ts_res_male_by_years<-seq(from=1980, to=2012)
at_ts_res_male_by_years<-NULL
for(j in seq(33)){
  for(i in seq(101)){
    if(i==1){
      at_ts_res_male_by_years[j]<-(j-1)*101+i
    }
  }
}

```



```

    }
    ts_res_male_by_years[(j-1)*101+i]<-
lca_male_1980_2012$residuals$y[i,j]
  }
}
plot(ts_res_male_by_years, xaxt = "n", main="Férfi reziduálisok éves
bontásokban", xlab="Év", ylab="")
axis(1, at=at_ts_res_male_by_years, labels=labels_ts_res_male_by_years)
abline(lwd=2, h=sqrt(ML_becslesek(lca_male_1980_2012)["szigma_epsilon_neg
yzet"]), col="orange")
abline(lwd=2, h=-
sqrt(ML_becslesek(lca_male_1980_2012)["szigma_epsilon_negyzet"]), col="or
ange")
# -----
ts_res_unisex_by_years<-NULL
labels_ts_res_unisex_by_years<-seq(from=1980, to=2012)
at_ts_res_unisex_by_years<-NULL
for(j in seq(33)){
  for(i in seq(101)){
    if(i==1){
      at_ts_res_unisex_by_years[j]<-(j-1)*101+i
    }
    ts_res_unisex_by_years[(j-1)*101+i]<-
lca_unisex_1980_2012$residuals$y[i,j]
  }
}
plot(ts_res_unisex_by_years, xaxt = "n", main="Unisex reziduálisok éves
bontásokban", xlab="Év", ylab="")
axis(1, at=at_ts_res_unisex_by_years,
labels=labels_ts_res_unisex_by_years)
abline(lwd=2, h=sqrt(ML_becslesek(lca_unisex_1980_2012)["szigma_epsilon_n
egyzet"]), col="orange")
abline(lwd=2, h=-
sqrt(ML_becslesek(lca_unisex_1980_2012)["szigma_epsilon_negyzet"]), col="
orange")

# plotting individual residuals by ages:
# -----
ts_res_female_by_ages<-NULL
labels_ts_res_female_by_ages<-seq(from=0, to=100)
at_ts_res_female_by_ages<-NULL
for(j in seq(63)){
  for(i in seq(101)){
    if(j==1){
      at_ts_res_female_by_ages[i]<-(i-1)*63+j
    }
    ts_res_female_by_ages[(i-1)*63+j]<-
lca_female_1950_2012$residuals$y[i,j]
  }
}
plot(ts_res_female_by_ages, xaxt = "n", main="Női reziduálisok kor
szerinti felosztásban", xlab="Kor", ylab="")
axis(1, at=at_ts_res_female_by_ages, labels=labels_ts_res_female_by_ages)
abline(lwd=2, h=sqrt(ML_becslesek(lca_female_1950_2012)["szigma_epsilon_n
egyzet"]), col="orange")
abline(lwd=2, h=-
sqrt(ML_becslesek(lca_female_1950_2012)["szigma_epsilon_negyzet"]), col="
orange")

```

```

# -----
ts_res_male_by_ages<-NULL
labels_ts_res_male_by_ages<-seq(from=0, to=100)
at_ts_res_male_by_ages<-NULL
for(j in seq(33)){
  for(i in seq(101)){
    if(j==1){
      at_ts_res_male_by_ages[i]<-(i-1)*33+j
    }
    ts_res_male_by_ages[(i-1)*33+j]<-lca_male_1980_2012$residuals$y[i,j]
  }
}
plot(ts_res_male_by_ages, xaxt = "n", main="Férfi reziduálisok kor
szerinti felosztásban", xlab="Kor", ylab="")
axis(1, at=at_ts_res_male_by_ages, labels=labels_ts_res_male_by_ages)
abline(lwd=2,h=sqrt(ML_becslesek(lca_male_1980_2012)["szigma_epsilon_neg
yzet"]),col="orange")
abline(lwd=2,h=-
sqrt(ML_becslesek(lca_male_1980_2012)["szigma_epsilon_negyzet"]),col="or
ange")
# -----
ts_res_unisex_by_ages<-NULL
labels_ts_res_unisex_by_ages<-seq(from=0, to=100)
at_ts_res_unisex_by_ages<-NULL
for(j in seq(33)){
  for(i in seq(101)){
    if(j==1){
      at_ts_res_unisex_by_ages[i]<-(i-1)*33+j
    }
    ts_res_unisex_by_ages[(i-1)*33+j]<-
lca_unisex_1980_2012$residuals$y[i,j]
  }
}
plot(ts_res_unisex_by_ages, xaxt = "n", main="Unisex reziduálisok kor
szerinti felosztásban", xlab="Kor", ylab="")
axis(1, at=at_ts_res_unisex_by_ages, labels=labels_ts_res_unisex_by_ages)
abline(lwd=2,h=sqrt(ML_becslesek(lca_unisex_1980_2012)["szigma_epsilon_n
egyzet"]),col="orange")
abline(lwd=2,h=-
sqrt(ML_becslesek(lca_unisex_1980_2012)["szigma_epsilon_negyzet"]),col="
orange")

# plotting random walk residuals:
# -----
res_Z_female<-NULL
for(i in seq(2012-1950)){
  res_Z_female[i]<-lca_female_1950_2012$kt[i+1]-
lca_female_1950_2012$kt[i]-ML_becslesek(lca_female_1950_2012)["c"]
}
plot(seq(from=1950,to=2011),res_Z_female,xlab="Év",ylab="",main="Női
reziduálisok")
abline(lwd=2,h=sqrt(ML_becslesek(lca_female_1950_2012)["szigma_Z_negyzet"
]),col="orange")
abline(lwd=2,h=-
sqrt(ML_becslesek(lca_female_1950_2012)["szigma_Z_negyzet"]),col="orange"
)
# -----
res_Z_male<-NULL

```

```

for(i in seq(2012-1980)){
  res_Z_male[i]<-lca_male_1980_2012$kt[i+1]-lca_male_1980_2012$kt[i]-
ML_becslesek(lca_male_1980_2012) ["c"]
}
plot(seq(from=1980,to=2011),res_Z_male,xlab="Év",ylab="",main="Férfi
reziduálisok")
abline(lwd=2,h=sqrt(ML_becslesek(lca_male_1980_2012) ["szigma_Z_negyzet"])
,col="orange")
abline(lwd=2,h=-
sqrt(ML_becslesek(lca_male_1980_2012) ["szigma_Z_negyzet"]),col="orange")
# -----
res_Z_unisex<-NULL
for(i in seq(2012-1980)){
  res_Z_unisex[i]<-lca_unisex_1980_2012$kt[i+1]-
lca_unisex_1980_2012$kt[i]-ML_becslesek(lca_unisex_1980_2012) ["c"]
}
plot(seq(from=1980,to=2011),res_Z_unisex,xlab="Év",ylab="",main="Unisex
reziduálisok")
abline(lwd=2,h=sqrt(ML_becslesek(lca_unisex_1980_2012) ["szigma_Z_negyzet"
]),col="orange")
abline(lwd=2,h=-
sqrt(ML_becslesek(lca_unisex_1980_2012) ["szigma_Z_negyzet"]),col="orange"
)

```

```

#####
# #
# ---- (SUB)SECTION: (4.4) ---- #
# #
#####

```

```

# creating 'lca' objects for unisex table trying to choose the best
fitting period in the modified Lee--Carter-model:
# -----

```

```

### 'lca' objects up to 50 ages:
### -----
lca_unisex_1950_2000_50<-
lca(demogdata_mortality_unisex,years=(1950:2000), adjust = "none",
ages=(0:50))
lca_unisex_1960_2000_50<-
lca(demogdata_mortality_unisex,years=(1960:2000), adjust = "none",
ages=(0:50))
lca_unisex_1970_2000_50<-
lca(demogdata_mortality_unisex,years=(1970:2000), adjust = "none",
ages=(0:50))
lca_unisex_1980_2000_50<-
lca(demogdata_mortality_unisex,years=(1980:2000), adjust = "none",
ages=(0:50))
lca_unisex_1989_2000_50<-
lca(demogdata_mortality_unisex,years=(1989:2000), adjust = "none",
ages=(0:50))

```

```

### 'lca' objects from 51 ages:
### -----

```

```

lca_unisex_1950_2000_51<-
lca(demogdata_mortality_unisex,years=(1950:2000), adjust = "none",
ages=(51:110))
lca_unisex_1960_2000_51<-
lca(demogdata_mortality_unisex,years=(1960:2000), adjust = "none",
ages=(51:110))
lca_unisex_1970_2000_51<-
lca(demogdata_mortality_unisex,years=(1970:2000), adjust = "none",
ages=(51:110))
lca_unisex_1980_2000_51<-
lca(demogdata_mortality_unisex,years=(1980:2000), adjust = "none",
ages=(51:110))
lca_unisex_1989_2000_51<-
lca(demogdata_mortality_unisex,years=(1989:2000), adjust = "none",
ages=(51:110))

# projecting unisex mortality probabilities from the modified Lee--
Carter-model (using equations (61), (62)):
# -----
forecast_mortality_unisex_1950_mod<-matrix(nrow=101,ncol=12)
forecast_mortality_unisex_1960_mod<-matrix(nrow=101,ncol=12)
forecast_mortality_unisex_1970_mod<-matrix(nrow=101,ncol=12)
forecast_mortality_unisex_1980_mod<-matrix(nrow=101,ncol=12)
forecast_mortality_unisex_1989_mod<-matrix(nrow=101,ncol=12)
# -----

for(i in seq(101)){
  ML_becslesek_unisex_50<-ML_becslesek(lca_unisex_1950_2000_50)
  ML_becslesek_unisex_51<-ML_becslesek(lca_unisex_1950_2000_51)
  for(j in seq(12)){
    if(i<=51){
      forecast_mortality_unisex_1950_mod[i,j]<-
exp(lca_unisex_1950_2000_50$ax[i]+lca_unisex_1950_2000_50$bx[i]*(lca_unis
ex_1950_2000_50$kt[length(lca_unisex_1950_2000_50$kt)]+j*ML_becslesek_uni
sex_50["c"])+j*(lca_unisex_1950_2000_50$bx[i])^2*ML_becslesek_unisex_50["
szigma_Z_negyzet"])/2+ML_becslesek_unisex_50["szigma_epsilon_negyzet"]/2)
    }
    else{
      forecast_mortality_unisex_1950_mod[i,j]<-
exp(lca_unisex_1950_2000_51$ax[i]+lca_unisex_1950_2000_51$bx[i]*(lca_unis
ex_1950_2000_51$kt[length(lca_unisex_1950_2000_51$kt)]+j*ML_becslesek_uni
sex_51["c"])+j*(lca_unisex_1950_2000_51$bx[i])^2*ML_becslesek_unisex_51["
szigma_Z_negyzet"])/2+ML_becslesek_unisex_51["szigma_epsilon_negyzet"]/2)
    }
  }
}
for(i in seq(101)){
  ML_becslesek_unisex_50<-ML_becslesek(lca_unisex_1960_2000_50)
  ML_becslesek_unisex_51<-ML_becslesek(lca_unisex_1960_2000_51)
  for(j in seq(12)){
    if(i<=51){
      forecast_mortality_unisex_1960_mod[i,j]<-
exp(lca_unisex_1960_2000_50$ax[i]+lca_unisex_1960_2000_50$bx[i]*(lca_unis
ex_1960_2000_50$kt[length(lca_unisex_1960_2000_50$kt)]+j*ML_becslesek_uni
sex_50["c"])+j*(lca_unisex_1960_2000_50$bx[i])^2*ML_becslesek_unisex_50["
szigma_Z_negyzet"])/2+ML_becslesek_unisex_50["szigma_epsilon_negyzet"]/2)
    }
  }
}

```

```

else{
  forecast_mortality_unisex_1960_mod[i,j]<-
exp(lca_unisex_1960_2000_51$ax[i]+lca_unisex_1960_2000_51$bx[i]*(lca_unis
ex_1960_2000_51$kt[length(lca_unisex_1960_2000_51$kt)]+j*ML_becslesek_uni
sex_51["c"])+j*(lca_unisex_1960_2000_51$bx[i])^2*ML_becslesek_unisex_51["
szigma_Z_negyzet"])/2+ML_becslesek_unisex_51["szigma_epszilon_negyzet"])/2)
}
}
}
for(i in seq(101)){
  ML_becslesek_unisex_50<-ML_becslesek(lca_unisex_1970_2000_50)
  ML_becslesek_unisex_51<-ML_becslesek(lca_unisex_1970_2000_51)
  for(j in seq(12)){
    if(i<=51){
      forecast_mortality_unisex_1970_mod[i,j]<-
exp(lca_unisex_1970_2000_50$ax[i]+lca_unisex_1970_2000_50$bx[i]*(lca_unis
ex_1970_2000_50$kt[length(lca_unisex_1970_2000_50$kt)]+j*ML_becslesek_uni
sex_50["c"])+j*(lca_unisex_1970_2000_50$bx[i])^2*ML_becslesek_unisex_50["
szigma_Z_negyzet"])/2+ML_becslesek_unisex_50["szigma_epszilon_negyzet"])/2)
}
else{
  forecast_mortality_unisex_1970_mod[i,j]<-
exp(lca_unisex_1970_2000_51$ax[i]+lca_unisex_1970_2000_51$bx[i]*(lca_unis
ex_1970_2000_51$kt[length(lca_unisex_1970_2000_51$kt)]+j*ML_becslesek_uni
sex_51["c"])+j*(lca_unisex_1970_2000_51$bx[i])^2*ML_becslesek_unisex_51["
szigma_Z_negyzet"])/2+ML_becslesek_unisex_51["szigma_epszilon_negyzet"])/2)
}
}
}
}
for(i in seq(101)){
  ML_becslesek_unisex_50<-ML_becslesek(lca_unisex_1980_2000_50)
  ML_becslesek_unisex_51<-ML_becslesek(lca_unisex_1980_2000_51)
  for(j in seq(12)){
    if(i<=51){
      forecast_mortality_unisex_1980_mod[i,j]<-
exp(lca_unisex_1980_2000_50$ax[i]+lca_unisex_1980_2000_50$bx[i]*(lca_unis
ex_1980_2000_50$kt[length(lca_unisex_1980_2000_50$kt)]+j*ML_becslesek_uni
sex_50["c"])+j*(lca_unisex_1980_2000_50$bx[i])^2*ML_becslesek_unisex_50["
szigma_Z_negyzet"])/2+ML_becslesek_unisex_50["szigma_epszilon_negyzet"])/2)
}
else{
  forecast_mortality_unisex_1980_mod[i,j]<-
exp(lca_unisex_1980_2000_51$ax[i]+lca_unisex_1980_2000_51$bx[i]*(lca_unis
ex_1980_2000_51$kt[length(lca_unisex_1980_2000_51$kt)]+j*ML_becslesek_uni
sex_51["c"])+j*(lca_unisex_1980_2000_51$bx[i])^2*ML_becslesek_unisex_51["
szigma_Z_negyzet"])/2+ML_becslesek_unisex_51["szigma_epszilon_negyzet"])/2)
}
}
}
}
}
for(i in seq(101)){
  ML_becslesek_unisex_50<-ML_becslesek(lca_unisex_1989_2000_50)
  ML_becslesek_unisex_51<-ML_becslesek(lca_unisex_1989_2000_51)
  for(j in seq(12)){
    if(i<=51){
      forecast_mortality_unisex_1989_mod[i,j]<-
exp(lca_unisex_1989_2000_50$ax[i]+lca_unisex_1989_2000_50$bx[i]*(lca_unis
ex_1989_2000_50$kt[length(lca_unisex_1989_2000_50$kt)]+j*ML_becslesek_uni

```

```

sex_50["c"])+j*(lca_unisex_1989_2000_50$bx[i])^2*ML_becslesek_unisex_50["
szigma_Z_negyzet"])/2+ML_becslesek_unisex_50["szigma_epszilon_negyzet"])/2)
}
else{
  forecast_mortality_unisex_1989_mod[i,j]<-
exp(lca_unisex_1989_2000_51$ax[i]+lca_unisex_1989_2000_51$bx[i]*(lca_unis
ex_1989_2000_51$kt[length(lca_unisex_1989_2000_51$kt)]+j*ML_becslesek_uni
sex_51["c"])+j*(lca_unisex_1989_2000_51$bx[i])^2*ML_becslesek_unisex_51["
szigma_Z_negyzet"])/2+ML_becslesek_unisex_51["szigma_epszilon_negyzet"])/2)
}
}
}

```

```

# calculating maximum likelihood estimates for unisex table in the
modified Lee--Carter-model (using equations (55), (56) and function
'ML_becslesek'):

```

```

# -----
-----
-----

```

```

szigma_epszilon_unisex_mod_1950<-NULL
szigma_epszilon_unisex_mod_1950[1]<-
sqrt(ML_becslesek(lca_unisex_1950_2000_50)["szigma_epszilon_negyzet"])
szigma_epszilon_unisex_mod_1950[2]<-
sqrt(ML_becslesek(lca_unisex_1950_2000_51)["szigma_epszilon_negyzet"])
names(szigma_epszilon_unisex_mod_1950)<-c("Up to 50 ages","from 51 ages")
szigma_epszilon_unisex_mod_1950

```

```

# -----
-----
-----

```

```

szigma_epszilon_unisex_mod_1960<-NULL
szigma_epszilon_unisex_mod_1960[1]<-
sqrt(ML_becslesek(lca_unisex_1960_2000_50)["szigma_epszilon_negyzet"])
szigma_epszilon_unisex_mod_1960[2]<-
sqrt(ML_becslesek(lca_unisex_1960_2000_51)["szigma_epszilon_negyzet"])
names(szigma_epszilon_unisex_mod_1960)<-c("Up to 50 ages","from 51 ages")
szigma_epszilon_unisex_mod_1960

```

```

# -----
-----
-----

```

```

szigma_epszilon_unisex_mod_1970<-NULL
szigma_epszilon_unisex_mod_1970[1]<-
sqrt(ML_becslesek(lca_unisex_1970_2000_50)["szigma_epszilon_negyzet"])
szigma_epszilon_unisex_mod_1970[2]<-
sqrt(ML_becslesek(lca_unisex_1970_2000_51)["szigma_epszilon_negyzet"])
names(szigma_epszilon_unisex_mod_1970)<-c("Up to 50 ages","from 51 ages")
szigma_epszilon_unisex_mod_1970

```

```

# -----
-----
-----

```

```

szigma_epszilon_unisex_mod_1980<-NULL
szigma_epszilon_unisex_mod_1980[1]<-
sqrt(ML_becslesek(lca_unisex_1980_2000_50)["szigma_epszilon_negyzet"])
szigma_epszilon_unisex_mod_1980[2]<-
sqrt(ML_becslesek(lca_unisex_1980_2000_51)["szigma_epszilon_negyzet"])
names(szigma_epszilon_unisex_mod_1980)<-c("Up to 50 ages","from 51 ages")
szigma_epszilon_unisex_mod_1980

```

```

# -----
-----
-----
szigma_epsilon_unisex_mod_1989<-NULL
szigma_epsilon_unisex_mod_1989[1]<-
sqrt(ML_becslesek(lca_unisex_1989_2000_50)["szigma_epsilon_negyzet"])
szigma_epsilon_unisex_mod_1989[2]<-
sqrt(ML_becslesek(lca_unisex_1989_2000_51)["szigma_epsilon_negyzet"])
names(szigma_epsilon_unisex_mod_1989)<-c("Up to 50 ages","from 51 ages")
szigma_epsilon_unisex_mod_1989
# -----
-----
-----
szigma_Z_unisex_mod_1950<-NULL
szigma_Z_unisex_mod_1950[1]<-
sqrt(ML_becslesek(lca_unisex_1950_2000_50)["szigma_Z_negyzet"])
szigma_Z_unisex_mod_1950[2]<-
sqrt(ML_becslesek(lca_unisex_1950_2000_51)["szigma_Z_negyzet"])
names(szigma_Z_unisex_mod_1950)<-c("Up to 50 ages","from 51 ages")
szigma_Z_unisex_mod_1950
# -----
-----
-----
szigma_Z_unisex_mod_1960<-NULL
szigma_Z_unisex_mod_1960[1]<-
sqrt(ML_becslesek(lca_unisex_1960_2000_50)["szigma_Z_negyzet"])
szigma_Z_unisex_mod_1960[2]<-
sqrt(ML_becslesek(lca_unisex_1960_2000_51)["szigma_Z_negyzet"])
names(szigma_Z_unisex_mod_1960)<-c("Up to 50 ages","from 51 ages")
szigma_Z_unisex_mod_1960
# -----
-----
-----
szigma_Z_unisex_mod_1970<-NULL
szigma_Z_unisex_mod_1970[1]<-
sqrt(ML_becslesek(lca_unisex_1970_2000_50)["szigma_Z_negyzet"])
szigma_Z_unisex_mod_1970[2]<-
sqrt(ML_becslesek(lca_unisex_1970_2000_51)["szigma_Z_negyzet"])
names(szigma_Z_unisex_mod_1970)<-c("Up to 50 ages","from 51 ages")
szigma_Z_unisex_mod_1970
# -----
-----
-----
szigma_Z_unisex_mod_1980<-NULL
szigma_Z_unisex_mod_1980[1]<-
sqrt(ML_becslesek(lca_unisex_1980_2000_50)["szigma_Z_negyzet"])
szigma_Z_unisex_mod_1980[2]<-
sqrt(ML_becslesek(lca_unisex_1980_2000_51)["szigma_Z_negyzet"])
names(szigma_Z_unisex_mod_1980)<-c("Up to 50 ages","from 51 ages")
szigma_Z_unisex_mod_1980
# -----
-----
-----
szigma_Z_unisex_mod_1989<-NULL
szigma_Z_unisex_mod_1989[1]<-
sqrt(ML_becslesek(lca_unisex_1989_2000_50)["szigma_Z_negyzet"])
szigma_Z_unisex_mod_1989[2]<-
sqrt(ML_becslesek(lca_unisex_1989_2000_51)["szigma_Z_negyzet"])

```

```

names(szigma_Z_unisex_mod_1989)<-c("Up to 50 ages","from 51 ages")
szigma_Z_unisex_mod_1989

# creating vectors of khi-squared tests from the modified Lee--Carter-
model (using equation (37)):
# -----
-----
khi_negyzet_1950_mod<-rep(0,13)
for(j in seq(12)){
  for(i in seq(25)){
    khi_negyzet_1950_mod[j]<-
khi_negyzet_1950_mod[j]+(matrix_exposure_unisex[i,51+j]*matrix_mortality_
unisex[i,51+j]-
matrix_exposure_unisex[i,51+j]*forecast_mortality_unisex_1950_mod[i,j])^2
/(matrix_exposure_unisex[i,51+j]*forecast_mortality_unisex_1950_mod[i,j]*
(1-forecast_mortality_unisex_1950_mod[i,j]))
  }
}
khi_negyzet_1950_mod[13]<-sum(khi_negyzet_1950_mod)
khi_negyzet_1950_mod
# -----
-----
khi_negyzet_1960_mod<-rep(0,13)
for(j in seq(12)){
  for(i in seq(25)){
    khi_negyzet_1960_mod[j]<-
khi_negyzet_1960_mod[j]+(matrix_exposure_unisex[i,51+j]*matrix_mortality_
unisex[i,51+j]-
matrix_exposure_unisex[i,51+j]*forecast_mortality_unisex_1960_mod[i,j])^2
/(matrix_exposure_unisex[i,51+j]*forecast_mortality_unisex_1960_mod[i,j]*
(1-forecast_mortality_unisex_1960_mod[i,j]))
  }
}
khi_negyzet_1960_mod[13]<-sum(khi_negyzet_1960_mod)
khi_negyzet_1960_mod
# -----
-----
khi_negyzet_1970_mod<-rep(0,13)
for(j in seq(12)){
  for(i in seq(25)){
    khi_negyzet_1970_mod[j]<-
khi_negyzet_1970_mod[j]+(matrix_exposure_unisex[i,51+j]*matrix_mortality_
unisex[i,51+j]-
matrix_exposure_unisex[i,51+j]*forecast_mortality_unisex_1970_mod[i,j])^2
/(matrix_exposure_unisex[i,51+j]*forecast_mortality_unisex_1970_mod[i,j]*
(1-forecast_mortality_unisex_1970_mod[i,j]))
  }
}
khi_negyzet_1970_mod[13]<-sum(khi_negyzet_1970_mod)
khi_negyzet_1970_mod
# -----
-----
khi_negyzet_1980_mod<-rep(0,13)
for(j in seq(12)){
  for(i in seq(25)){
    khi_negyzet_1980_mod[j]<-
khi_negyzet_1980_mod[j]+(matrix_exposure_unisex[i,51+j]*matrix_mortality_
unisex[i,51+j]-

```



```

matrix_exposure_unisex[i,51+j]*forecast_mortality_unisex_1980_mod[i,j])^2
/(matrix_exposure_unisex[i,51+j]*forecast_mortality_unisex_1980_mod[i,j]*
(1-forecast_mortality_unisex_1980_mod[i,j]))
}
}
khi_negyzet_1980_mod[13]<-sum(khi_negyzet_1980_mod)
khi_negyzet_1980_mod
# -----
-----
khi_negyzet_1989_mod<-rep(0,13)
for(j in seq(12)){
  for(i in seq(25)){
    khi_negyzet_1989_mod[j]<-
khi_negyzet_1989_mod[j]+(matrix_exposure_unisex[i,51+j]*matrix_mortality_
unisex[i,51+j]-
matrix_exposure_unisex[i,51+j]*forecast_mortality_unisex_1989_mod[i,j])^2
/(matrix_exposure_unisex[i,51+j]*forecast_mortality_unisex_1989_mod[i,j]*
(1-forecast_mortality_unisex_1989_mod[i,j]))
  }
}
khi_negyzet_1989_mod[13]<-sum(khi_negyzet_1989_mod)
khi_negyzet_1989_mod

### calculating the 5% and 95% quantization of the 26 and 312 freedom-
rate khi-squared distribution:
### -----
-----
qchisq(.05, df=26)
qchisq(.95, df=26)
qchisq(.05, df=312)
qchisq(.95, df=312)

#####
# #
# ---- (SUB)SECTION: (4.5) ---- #
# #
#####

# creating functions 'atlagolt_varhato_elet', 'varhato_elet_max' and
'varhato_elet_gyakorlatban' in order to generate remaining lifetimes
(e(0)) and calculate the remaining lifetimes in practise (e(2)) and in
our approach (e(1)) in ages:
# -----
-----
-----
-----
varhato_elet_generalas<-
function(ax,bx,c,k_2000,t,kor,szigma_epszilon_negyzet,szigma_Z_negyzet){
  if(kor<length(ax)-1){
    vector_random_epszilon<-rnorm(n=((length(ax)-1)-kor),mean=0,
sd=sqrt(szigma_epszilon_negyzet))
    vector_random_Z<-rnorm(n=(length(ax)-1)-kor+t-1,mean=0,
sd=sqrt(szigma_Z_negyzet))
    vector_q<-NULL
    for (ev in seq(from=0, to=(length(ax)-1)-kor-1)){

```

```

    vector_q[ev+1]<-
exp(ax[kor+ev+1]+bx[kor+ev+1]*(k_2000+sum(vector_random_Z[seq(from=1,by=1
,length.out=(t+ev)])))+(t+ev)*c)+vector_random_epsilon[ev+1])
}
varhato_elet<-0
for (ev in seq(from=1, to=(length(ax)-1)-kor)){
  seged<-1
  for (j in seq(from=0, to=ev-1)){
    seged = seged*(1-vector_q[j+1])
  }
  if(ev<(length(ax)-1)-kor){
    varhato_elet = varhato_elet + ev*vector_q[ev+1]*seged
  }
  else{
    varhato_elet = varhato_elet + ev*1*seged
  }
}
return(varhato_elet)
}
else{
  return(0)
}
}
# -----
-----
-----
-----
atlagolt_varhato_elet<-
function(ax,bx,c,k_2000,t,kor,szigma_epszilon_negyzet,szigma_Z_negyzet,
replikacio){
  atlag = 0
  for(n in seq(replikacio)){
    atlag = atlag +
varhato_elet_generalas(ax,bx,c,k_2000,t,kor,szigma_epszilon_negyzet,szigm
a_Z_negyzet)/replikacio
  }
  return(atlag)
}
# -----
-----
-----
-----
varhato_elet_max<-
function(ax,bx,c,k_2000,t,kor,szigma_epszilon_negyzet,szigma_Z_negyzet){
  if(kor<length(ax)-1){
    varhato_elet<-0
    for (ev in seq(from=1, to=(length(ax)-1)-kor)){
      seged<-1
      for (j in seq(from=0, to=ev-1)){
        seged = seged*(1-
exp(ax[kor+j+1]+bx[kor+j+1]*(j+t)*c+bx[kor+j+1]*k_2000+szigma_epszilon_ne
gyzet/2+szigma_Z_negyzet/2*(2*bx[kor+ev+1]*bx[kor+j+1]*(j+t)+bx[kor+j+1]^
2*(j+t))))
      }
      if(ev<length(ax)-kor-1){
        varhato_elet = varhato_elet +
ev*exp(ax[kor+ev+1]+bx[kor+ev+1]*(ev+t)*c+bx[kor+ev+1]*k_2000+szigma_epsz
ilon_negyzet/2+(ev+t)*bx[kor+ev+1]^2*szigma_Z_negyzet/2)*seged

```

```

    }
    else{
      varhato_elet = varhato_elet + ev*1*seged
    }
  }
  return(varhato_elet)
}
else{
  return(0)
}
}
# -----
-----
-----
-----
varhato_elet_gyakorlatban<-function(ax,bx,c,k_2000,t,kor){
  if(kor<length(ax)-1){
    varhato_elet<-0
    for (ev in seq(from=1, to=(length(ax)-1)-kor)){
      seged<-1
      for (j in seq(from=0, to=ev-1)){
        seged = seged*(1-
exp(ax[kor+j+1]+bx[kor+j+1]*(j+t)*c+bx[kor+j+1]*k_2000))
      }
      if(ev<length(ax)-kor-1){
        varhato_elet = varhato_elet +
ev*exp(ax[kor+ev+1]+bx[kor+ev+1]*(ev+t)*c+bx[kor+ev+1]*k_2000)*seged
      }
      else{
        varhato_elet = varhato_elet + ev*1*seged
      }
    }
    return(varhato_elet)
  }
  else{
    return(0)
  }
}

# Generating remaining lifetimes (e(0)) in ages (it may take a few
minutes):
# -----
---
varhato_eletek_rep_female<-matrix(nrow=101,ncol=101)
for(i in seq(101)){
  for(j in seq(101)){
    varhato_eletek_rep_female[i,j]<-
atlagolt_varhato_elet(lca_female_1950_2012$ax,lca_female_1950_2012$bx,ML_
becslesek_female_1950_2012["c"],lca_female_1950_2012$kt[length(lca_female
_1950_2012$kt)],j-1,i-
1,ML_becslesek_female_1950_2012["szigma_epsilon_negyzet"],ML_becslesek_f
emale_1950_2012["szigma_z_negyzet"],25)
  }
}
scatters_varhato_eletek_rep_female<-matrix(ncol=3,nrow=101*101)
colnames(scatters_varhato_eletek_rep_female)<-c("Kor","Év","Várható
élettartam")
for(j in seq(101)){

```

```

for(i in seq(101)){
  scatters_varhato_eletek_rep_female[(j-1)*101+i,1]<-(i-1)
  scatters_varhato_eletek_rep_female[(j-1)*101+i,2]<-(2012-1+j)
  scatters_varhato_eletek_rep_female[(j-1)*101+i,3]<-
varhato_eletek_rep_female[i,j]
}
}
# -----
---
varhato_eletek_rep_male<-matrix(nrow=101,ncol=101)
for(i in seq(101)){
  for(j in seq(101)){
    varhato_eletek_rep_male[i,j]<-
atlagolt_varhato_elet(lca_male_1980_2012$ax,lca_male_1980_2012$bx,ML_becs
lesekek_male_1980_2012["c"],lca_male_1980_2012$kt[length(lca_male_1980_2012
$kt)],j-1,i-
1,ML_becslesek_male_1980_2012["szigma_epsilon_negyzet"],ML_becslesek_mal
e_1980_2012["szigma_Z_negyzet"],25)
  }
}
scatters_varhato_eletek_rep_male<-matrix(ncol=3,nrow=101*101)
colnames(scatters_varhato_eletek_rep_male)<-c("Kor","Év","Várható
élettartam")
for(j in seq(101)){
  for(i in seq(101)){
    scatters_varhato_eletek_rep_male[(j-1)*101+i,1]<-(i-1)
    scatters_varhato_eletek_rep_male[(j-1)*101+i,2]<-(2012-1+j)
    scatters_varhato_eletek_rep_male[(j-1)*101+i,3]<-
varhato_eletek_rep_male[i,j]
  }
}
# -----
---
varhato_eletek_rep_unisex<-matrix(nrow=101,ncol=101)
for(i in seq(101)){
  for(j in seq(101)){
    varhato_eletek_rep_unisex[i,j]<-
atlagolt_varhato_elet(lca_unisex_1980_2012$ax,lca_unisex_1980_2012$bx,ML_
becslesek_unisex_1980_2012["c"],lca_unisex_1980_2012$kt[length(lca_unisex
_1980_2012$kt)],j-1,i-
1,ML_becslesek_unisex_1980_2012["szigma_epsilon_negyzet"],ML_becslesek_u
nisex_1980_2012["szigma_Z_negyzet"],25)
  }
}
scatters_varhato_eletek_rep_unisex<-matrix(ncol=3,nrow=101*101)
colnames(scatters_varhato_eletek_rep_unisex)<-c("Kor","Év","Várható
élettartam")
for(j in seq(101)){
  for(i in seq(101)){
    scatters_varhato_eletek_rep_unisex[(j-1)*101+i,1]<-(i-1)
    scatters_varhato_eletek_rep_unisex[(j-1)*101+i,2]<-(2012-1+j)
    scatters_varhato_eletek_rep_unisex[(j-1)*101+i,3]<-
varhato_eletek_rep_unisex[i,j]
  }
}
}

# plotting the generated remainig lifetimes (e(0)) in ages in 3D:
# -----

```

```

colors_rep_female<-NULL
for(i in seq(101*101)){
  colors_rep_female[i]<-paste("gray",toString(70-
round((scatters_varhato_eletek_rep_female[i,3])*0.68)))
}
colors_rep_male<-NULL
for(i in seq(101*101)){
  colors_rep_male[i]<-paste("gray",toString(70-
round((scatters_varhato_eletek_rep_male[i,3])*0.68)))
}
colors_rep_unisex<-NULL
for(i in seq(101*101)){
  colors_rep_unisex[i]<-paste("gray",toString(70-
round((scatters_varhato_eletek_rep_unisex[i,3])*0.68)))
}
# -----
scatterplot3d(x=scatters_varhato_eletek_rep_female,pch=16,color=colors_re
p_female,box=FALSE,grid=TRUE)
scatterplot3d(x=scatters_varhato_eletek_rep_male,pch=16,color=colors_rep_
male,box=FALSE,grid=TRUE)
scatterplot3d(x=scatters_varhato_eletek_rep_unisex,pch=16,color=colors_re
p_unisex,box=FALSE,grid=TRUE)

# Generating remaining lifetimes (e(0)) and calculating the remaining
lifetimes in practise (e(2)) and in our approach (e(1)) in ages:
# -----

### creating function 'abrazolas_3_felekeppen' in order to do generating:
### -----
abrazolas_3_felekeppen<-
function(ax,bx,c,k_2000,t_start,t_length,kor,szigma_epsilon_negyzet,szig
ma_Z_negyzet,rep,group){
  ## Megj.: A 't_start' a 2012-es évhez képest vett tartam.
  varhato_eletek_rep_2012<-NULL
  for(i in seq(t_length)){
    varhato_eletek_rep_2012[i]<-
atlagolt_varhato_elet(ax,bx,c,k_2000,t_start+i,kor,szigma_epsilon_negyze
t,szigma_Z_negyzet,rep)
  }

  varhato_eletek_max_2012<-NULL
  for(i in seq(t_length)){
    varhato_eletek_max_2012[i]<-
varhato_elet_max(ax,bx,c,k_2000,t_start+i,kor,szigma_epsilon_negyzet,szi
gma_Z_negyzet)
  }

  varhato_eletek_gyakorlatban_2012<-NULL
  for(i in seq(t_length)){
    varhato_eletek_gyakorlatban_2012[i]<-
varhato_elet_gyakorlatban(ax,bx,c,k_2000,t_start+i,kor)
  }

  matrix_varhato_eletek_2012<-
rbind(varhato_eletek_rep_2012,varhato_eletek_gyakorlatban_2012,varhato_el
etek_max_2012)

```

```

rownames(matrix_varhato_eletek_2012)<-c("Replikációval
generált","Gyakorlatban alkalmazott","Saját közelítés")
ts_rep_2012<-ts(data=t(matrix_varhato_eletek_2012),start=2012+t_start)
par(mfrow=c(1,1))
if(group==1){
  if(kor==0){
    plot(x=ts_rep_2012, plot.type = "s",type=c("l"),lwd=2,
main=paste("Férfiak születéskori várható
élettartamai\n",toString(2012+t_start+1),"-tól/től
",toString(2012+t_start+t_length),"-ig"), xlab="DÁTUM", ylab="ÉV",
col=c("blue","black","red"))
  }
  else{
    plot(x=ts_rep_2012, plot.type = "s",type=c("l"),lwd=2,
main=paste(toString(kor)," éves férfiak várható
élettartamai\n",toString(2012+t_start+1),"-tól/től
",toString(2012+t_start+t_length),"-ig"), xlab="DÁTUM", ylab="ÉV",
col=c("blue","black","red"))
  }
}

```

```

legend(x=2012+t_start,y=varhato_elet_gyakorlatban(ax,bx,c,k_2000,t_start+
t_length,kor),legend=c("Gyakorlatban alkalmazott","Replikációval
generált","Saját közelítés"),
col=c("black","blue","red"),lty=c(1,1,1),lwd=c(3,3,3))
}
if(group==2){
  if(kor==0){
    plot(x=ts_rep_2012, plot.type = "s",type=c("l"),lwd=2,
main=paste("Nők születéskori várható
élettartamai\n",toString(2012+t_start+1),"-tól/től
",toString(2012+t_start+t_length),"-ig"), xlab="DÁTUM", ylab="ÉV",
col=c("blue","black","red"))
  }
  else{
    plot(x=ts_rep_2012, plot.type = "s",type=c("l"),lwd=2,
main=paste(toString(kor)," éves nők várható
élettartamai\n",toString(2012+t_start+1),"-tól/től
",toString(2012+t_start+t_length),"-ig"), xlab="DÁTUM", ylab="ÉV",
col=c("blue","black","red"))
  }
}

```

```

legend(x=2012+t_start,y=varhato_elet_gyakorlatban(ax,bx,c,k_2000,t_start+
t_length,kor),legend=c("Gyakorlatban alkalmazott","Replikációval
generált","Saját közelítés"),
col=c("black","blue","red"),lty=c(1,1,1),lwd=c(3,3,3))
}
if(group==3){
  if(kor==0){
    plot(x=ts_rep_2012, plot.type = "s",type=c("l"),lwd=2,
main=paste("Unisex születéskori várható
élettartamok\n",toString(2012+t_start+1),"-tól/től
",toString(2012+t_start+t_length),"-ig"), xlab="DÁTUM", ylab="ÉV",
col=c("blue","black","red"))
  }
  else{

```

```

        plot(x=ts_rep_2012, plot.type = "s", type=c("l"), lwd=2,
main=paste(toString(kor), " éves korban unisex várható
élettartamok\n", toString(2012+t_start+1), "-tól/től
", toString(2012+t_start+t_length), "-ig"), xlab="DÁTUM", ylab="ÉV",
col=c("blue", "black", "red"))
    }

legend(x=2012+t_start, y=varhato_elet_gyakorlatban(ax,bx,c,k_2000,t_start+
t_length, kor), legend=c("Gyakorlatban alkalmazott", "Replikációval
generált", "Saját közelítés"),
col=c("black", "blue", "red"), lty=c(1,1,1), lwd=c(3,3,3))
    }
    return(ts_rep_2012)
}

### creating time series from generated and calculated remaining
lifetimes (it may take a few minutes):
### -----
ts_rep_female_szuleteskori<-
abrazolas_3_felekeppen(lca_female_1950_2012$ax, lca_female_1950_2012$bx, ML_
_becslesek_female_1950_2012["c"], lca_female_1950_2012$kt[length(lca_femal
e_1950_2012$kt)], 5, 100, 0, ML_becslesek_female_1950_2012["szigma_epsilon_n
egyzet"], ML_becslesek_female_1950_2012["szigma_Z_negyzet"], 2000, group=2)
ts_rep_female_45<-
abrazolas_3_felekeppen(lca_female_1950_2012$ax, lca_female_1950_2012$bx, ML_
_becslesek_female_1950_2012["c"], lca_female_1950_2012$kt[length(lca_femal
e_1950_2012$kt)], 5, 100, 45, ML_becslesek_female_1950_2012["szigma_epsilon_
negyzet"], ML_becslesek_female_1950_2012["szigma_Z_negyzet"], 2000, group=2)
ts_rep_female_65<-
abrazolas_3_felekeppen(lca_female_1950_2012$ax, lca_female_1950_2012$bx, ML_
_becslesek_female_1950_2012["c"], lca_female_1950_2012$kt[length(lca_femal
e_1950_2012$kt)], 5, 100, 65, ML_becslesek_female_1950_2012["szigma_epsilon_
negyzet"], ML_becslesek_female_1950_2012["szigma_Z_negyzet"], 2000, group=2)
### -----
ts_rep_male_szuleteskori<-
abrazolas_3_felekeppen(lca_male_1980_2012$ax, lca_male_1980_2012$bx, ML_bec
slesek_male_1980_2012["c"], lca_male_1980_2012$kt[length(lca_male_1980_201
2$kt)], 5, 100, 0, ML_becslesek_male_1980_2012["szigma_epsilon_negyzet"], ML_
becslesek_male_1980_2012["szigma_Z_negyzet"], 2000, group=1)
ts_rep_male_45<-
abrazolas_3_felekeppen(lca_male_1980_2012$ax, lca_male_1980_2012$bx, ML_bec
slesek_male_1980_2012["c"], lca_male_1980_2012$kt[length(lca_male_1980_201
2$kt)], 5, 100, 45, ML_becslesek_male_1980_2012["szigma_epsilon_negyzet"], ML_
becslesek_male_1980_2012["szigma_Z_negyzet"], 2000, group=1)
ts_rep_male_65<-
abrazolas_3_felekeppen(lca_male_1980_2012$ax, lca_male_1980_2012$bx, ML_bec
slesek_male_1980_2012["c"], lca_male_1980_2012$kt[length(lca_male_1980_201
2$kt)], 5, 100, 65, ML_becslesek_male_1980_2012["szigma_epsilon_negyzet"], ML_
becslesek_male_1980_2012["szigma_Z_negyzet"], 2000, group=1)
### -----
ts_rep_unisex_szuleteskori<-
abrazolas_3_felekeppen(lca_unisex_1980_2012$ax, lca_unisex_1980_2012$bx, ML_
becslesek_unisex_1980_2012["c"], lca_unisex_1980_2012$kt[length(lca_unise

```

```

x_1980_2012$kt)],5,100,0,ML_becslesek_unisex_1980_2012["szigma_epsilon_n
egyzet"],ML_becslesek_unisex_1980_2012["szigma_Z_negyzet"],2000,group=3)
ts_rep_unisex_45<-
abrazolas_3_felekeppen(lca_unisex_1980_2012$ax,lca_unisex_1980_2012$bx,ML
_becslesek_unisex_1980_2012["c"],lca_unisex_1980_2012$kt[length(lca_unise
x_1980_2012$kt)],5,100,45,ML_becslesek_unisex_1980_2012["szigma_epsilon_
negyzet"],ML_becslesek_unisex_1980_2012["szigma_Z_negyzet"],2000,group=3)
ts_rep_unisex_65<-
abrazolas_3_felekeppen(lca_unisex_1980_2012$ax,lca_unisex_1980_2012$bx,ML
_becslesek_unisex_1980_2012["c"],lca_unisex_1980_2012$kt[length(lca_unise
x_1980_2012$kt)],5,100,65,ML_becslesek_unisex_1980_2012["szigma_epsilon_
negyzet"],ML_becslesek_unisex_1980_2012["szigma_Z_negyzet"],2000,group=3)

### plotting created time series above:
### -----
plot(x=ts_rep_female_szuleteskori, plot.type = "s",type=c("l"),lwd=2,
xlab="DÁTUM", ylab="ÉV", col=c("blue","black","red"))
plot(x=ts_rep_female_45, plot.type = "s",type=c("l"),lwd=2, xlab="DÁTUM",
ylab="ÉV", col=c("blue","black","red"))
plot(x=ts_rep_female_65, plot.type = "s",type=c("l"),lwd=2, xlab="DÁTUM",
ylab="ÉV", col=c("blue","black","red"))
### -----
plot(x=ts_rep_male_szuleteskori, plot.type = "s",type=c("l"),lwd=2,
xlab="DÁTUM", ylab="ÉV", col=c("blue","black","red"))
plot(x=ts_rep_male_45, plot.type = "s",type=c("l"),lwd=2, xlab="DÁTUM",
ylab="ÉV", col=c("blue","black","red"))
plot(x=ts_rep_male_65, plot.type = "s",type=c("l"),lwd=2, xlab="DÁTUM",
ylab="ÉV", col=c("blue","black","red"))
### -----
plot(x=ts_rep_unisex_szuleteskori, plot.type = "s",type=c("l"),lwd=2,
xlab="DÁTUM", ylab="ÉV", col=c("blue","black","red"))
plot(x=ts_rep_unisex_45, plot.type = "s",type=c("l"),lwd=2, xlab="DÁTUM",
ylab="ÉV", col=c("blue","black","red"))
plot(x=ts_rep_unisex_65, plot.type = "s",type=c("l"),lwd=2, xlab="DÁTUM",
ylab="ÉV", col=c("blue","black","red"))

# creating function 'hisztogramkeszites' in order to create histograms:
# -----
hisztogramkeszites<-
function(ax,bx,c,k_2000,t,kor,szigma_epsilon_negyzet,szigma_Z_negyzet,rep,group){
  varhato_eletek_generalas<-NULL
  for(i in seq(rep)){
    varhato_eletek_generalas[i]<-
varhato_elet_generalas(ax,bx,c,k_2000,t,kor,szigma_epsilon_negyzet,szigm
a_Z_negyzet)
  }
  if(group==1){
    if(kor==0){
hist(breaks=rep/35,freq=FALSE,right=FALSE,x=varhato_eletek_generalas,main
=paste("Férfiak születés kori várható élettartamaiból
generált\nmintahisztogramja ",toString(2012+t),"-ban/ben"), xlab="Várható
élettartamok",ylab="Sűrűségi mérték")
    }
  }
  else{
hist(breaks=rep/35,freq=FALSE,right=FALSE,x=varhato_eletek_generalas,main

```



```

=paste(toString(kor)," éves férfiak várható élettartamaiból
generált\nminta hisztogramja ",toString(2012+t),"-ban/ben"),
xlab="Várható élettartamok", ylab="Sűrűségi mérték")
}
}
if(group==2) {
  if(kor==0) {

hist(breaks=rep/35,freq=FALSE,right=FALSE,x=varhato_eletek_generalas,main
=paste("Nők születés kori várható élettartamaiból generált\nminta
hisztogramja ",toString(2012+t),"-ban/ben"), xlab="Várható
élettartamok",ylab="Sűrűségi mérték")
}
else{

hist(breaks=rep/35,freq=FALSE,right=FALSE,x=varhato_eletek_generalas,main
=paste(toString(kor)," éves nők várható élettartamaiból generált\nminta
hisztogramja ",toString(2012+t),"-ban/ben"), xlab="Várható élettartamok",
ylab="Sűrűségi mérték")
}
}
if(group==3) {
  if(kor==0) {

hist(breaks=rep/35,freq=FALSE,right=FALSE,x=varhato_eletek_generalas,main
=paste("Születés kori várható élettartamokból generált\nminta hisztogramja
",toString(2012+t),"-ban/ben"), xlab="Várható
élettartamok",ylab="Sűrűségi mérték")
}
else{

hist(breaks=rep/35,freq=FALSE,right=FALSE,x=varhato_eletek_generalas,main
=paste(toString(kor)," évesek várható élettartamaiból generált\nminta
hisztogramja ",toString(2012+t),"-ban/ben"), xlab="Várható élettartamok",
ylab="Sűrűségi mérték")
}
}
return(varhato_eletek_generalas)
}

# generating remaining lifetimes (e(0)) in order to create histograms:
# -----
hiszt_rep_5000_male_sz<-
hisztogramkeszites(lca_male_1980_2012$ax,lca_male_1980_2012$bx,ML_becslesek_male_1980_2012["c"],lca_male_1980_2012$kt[length(lca_male_1980_2012$kt)],6,0,ML_becslesek_male_1980_2012["szigma_epszilon_negyzet"],ML_becslesek_male_1980_2012["szigma_Z_negyzet"],5000,1)
hiszt_rep_5000_female_sz<-
hisztogramkeszites(lca_female_1950_2012$ax,lca_female_1950_2012$bx,ML_becslesek_female_1950_2012["c"],lca_female_1950_2012$kt[length(lca_female_1950_2012$kt)],6,0,ML_becslesek_female_1950_2012["szigma_epszilon_negyzet"],ML_becslesek_female_1950_2012["szigma_Z_negyzet"],5000,2)
hiszt_rep_5000_unisex_sz<-
hisztogramkeszites(lca_unisex_1980_2012$ax,lca_unisex_1980_2012$bx,ML_becslesek_unisex_1980_2012["c"],lca_unisex_1980_2012$kt[length(lca_unisex_1980_2012$kt)],6,0,ML_becslesek_unisex_1980_2012["szigma_epszilon_negyzet"],ML_becslesek_unisex_1980_2012["szigma_Z_negyzet"],5000,3)

```

```

hiszt_rep_5000_male_65<-
hisztogramkészítés(lca_male_1980_2012$ax,lca_male_1980_2012$bx,ML_becsles
ek_male_1980_2012["c"],lca_male_1980_2012$kt[length(lca_male_1980_2012$kt
)],6,65,ML_becslesek_male_1980_2012["szigma_epsilon_negyzet"],ML_becsles
ek_male_1980_2012["szigma_Z_negyzet"],5000,1)
hiszt_rep_5000_female_65<-
hisztogramkészítés(lca_female_1950_2012$ax,lca_female_1950_2012$bx,ML_bec
slesek_female_1950_2012["c"],lca_female_1950_2012$kt[length(lca_female_19
50_2012$kt)],6,65,ML_becslesek_female_1950_2012["szigma_epsilon_negyzet"
],ML_becslesek_female_1950_2012["szigma_Z_negyzet"],5000,2)
hiszt_rep_5000_unisex_65<-
hisztogramkészítés(lca_unisex_1980_2012$ax,lca_unisex_1980_2012$bx,ML_bec
slesek_unisex_1980_2012["c"],lca_unisex_1980_2012$kt[length(lca_unisex_19
80_2012$kt)],6,65,ML_becslesek_unisex_1980_2012["szigma_epsilon_negyzet"
],ML_becslesek_unisex_1980_2012["szigma_Z_negyzet"],5000,3)

# creating histograms:
# -----
sorted_hiszt_rep_5000_female_sz<-sort(hiszt_rep_5000_female_sz)
sorted_hiszt_rep_5000_male_sz<-sort(hiszt_rep_5000_male_sz)
sorted_hiszt_rep_5000_unisex_sz<-sort(hiszt_rep_5000_unisex_sz)
sorted_hiszt_rep_5000_female_65<-sort(hiszt_rep_5000_female_65)
sorted_hiszt_rep_5000_male_65<-sort(hiszt_rep_5000_male_65)
sorted_hiszt_rep_5000_unisex_65<-sort(hiszt_rep_5000_unisex_65)
# -----
hist(breaks=5000/35,freq=FALSE,right=FALSE,x=hiszt_rep_5000_female_sz,mai
n="",xaxt="n", xlab="Várható élettartamok",ylab="Relatív gyakoriság")
abline(v=mean(sorted_hiszt_rep_5000_female_sz), lwd=2, col="blue")
abline(v=sorted_hiszt_rep_5000_female_sz[5000*0.05], lwd=2, col="orange")
abline(v=sorted_hiszt_rep_5000_female_sz[5000*0.95], lwd=2, col="orange")
axis(1,at=c(80,81,round(sorted_hiszt_rep_5000_female_sz[5000*0.05],digits
=2),round(mean(hiszt_rep_5000_female_sz),digits=2),round(sorted_hiszt_rep
_5000_female_sz[5000*0.95],digits=2),89),labels=c(80,81,round(sorted_hisz
t_rep_5000_female_sz[5000*0.05],digits=2),round(mean(hiszt_rep_5000_femal
e_sz),digits=2),round(sorted_hiszt_rep_5000_female_sz[5000*0.95],digits=2
),89))
# -----
hist(breaks=5000/35,freq=FALSE,right=FALSE,x=hiszt_rep_5000_female_65,mai
n="",xaxt="n", xlab="Várható élettartamok",ylab="Relatív gyakoriság")
abline(v=mean(sorted_hiszt_rep_5000_female_65), lwd=2, col="blue")
abline(v=sorted_hiszt_rep_5000_female_65[5000*0.05], lwd=2, col="orange")
abline(v=sorted_hiszt_rep_5000_female_65[5000*0.95], lwd=2, col="orange")
axis(1,at=c(16,round(sorted_hiszt_rep_5000_female_65[5000*0.05],digits=2
),round(mean(hiszt_rep_5000_female_65),digits=2),round(sorted_hiszt_rep_50
00_female_65[5000*0.95],digits=2),21),labels=c(16,round(sorted_hiszt_rep_
5000_female_65[5000*0.05],digits=2),round(mean(hiszt_rep_5000_female_65)
,digits=2),round(sorted_hiszt_rep_5000_female_65[5000*0.95],digits=2),21))
# -----
hist(breaks=5000/35,freq=FALSE,right=FALSE,x=hiszt_rep_5000_male_sz,main=
"",xaxt="n", xlab="Várható élettartamok",ylab="Relatív gyakoriság")
abline(v=mean(sorted_hiszt_rep_5000_male_sz), lwd=2, col="blue")
abline(v=sorted_hiszt_rep_5000_male_sz[5000*0.05], lwd=2, col="orange")
abline(v=sorted_hiszt_rep_5000_male_sz[5000*0.95], lwd=2, col="orange")
axis(1,at=c(75,round(sorted_hiszt_rep_5000_male_sz[5000*0.05],digits=2),r
ound(mean(hiszt_rep_5000_male_sz),digits=2),round(sorted_hiszt_rep_5000_m
ale_sz[5000*0.95],digits=2),84),labels=c(75,round(sorted_hiszt_rep_5000_m
ale_sz[5000*0.05],digits=2),round(mean(hiszt_rep_5000_male_sz),digits=2)
,round(sorted_hiszt_rep_5000_male_sz[5000*0.95],digits=2),84))

```

```

# -----
hist(breaks=5000/35,freq=FALSE,right=FALSE,x=hiszt_rep_5000_male_65,main=
"",xaxt="n", xlab="Várható élettartamok",ylab="Relatív gyakoriság")
abline(v=mean(sorted_hiszt_rep_5000_male_65), lwd=2, col="blue")
abline(v=sorted_hiszt_rep_5000_male_65[5000*0.05], lwd=2, col="orange")
abline(v=sorted_hiszt_rep_5000_male_65[5000*0.95], lwd=2, col="orange")
axis(1,at=c(13,round(sorted_hiszt_rep_5000_male_65[5000*0.05],digits=2),r
ound(mean(hiszt_rep_5000_male_65),digits=2),round(sorted_hiszt_rep_5000_m
ale_65[5000*0.95],digits=2),17),labels=c(13,round(sorted_hiszt_rep_5000_m
ale_65[5000*0.05],digits=2),round(mean(hiszt_rep_5000_male_65),digits=2),
round(sorted_hiszt_rep_5000_male_65[5000*0.95],digits=2),17))
# -----
hist(breaks=5000/35,freq=FALSE,right=FALSE,x=hiszt_rep_5000_unisex_sz,mai
n="",xaxt="n", xlab="Várható élettartamok",ylab="Relatív gyakoriság")
abline(v=mean(sorted_hiszt_rep_5000_unisex_sz), lwd=2, col="blue")
abline(v=sorted_hiszt_rep_5000_unisex_sz[5000*0.05], lwd=2, col="orange")
abline(v=sorted_hiszt_rep_5000_unisex_sz[5000*0.95], lwd=2, col="orange")
axis(1,at=c(80,round(sorted_hiszt_rep_5000_unisex_sz[5000*0.05],digits=2)
,round(mean(hiszt_rep_5000_unisex_sz),digits=2),round(sorted_hiszt_rep_50
00_unisex_sz[5000*0.95],digits=2),87),labels=c(80,round(sorted_hiszt_rep_
5000_unisex_sz[5000*0.05],digits=2),round(mean(hiszt_rep_5000_unisex_sz),
digits=2),round(sorted_hiszt_rep_5000_unisex_sz[5000*0.95],digits=2),87))
# -----
hist(breaks=5000/35,freq=FALSE,right=FALSE,x=hiszt_rep_5000_unisex_65,mai
n="",xaxt="n", xlab="Várható élettartamok",ylab="Relatív gyakoriság")
abline(v=mean(sorted_hiszt_rep_5000_unisex_65), lwd=2, col="blue")
abline(v=sorted_hiszt_rep_5000_unisex_65[5000*0.05], lwd=2, col="orange")
abline(v=sorted_hiszt_rep_5000_unisex_65[5000*0.95], lwd=2, col="orange")
axis(1,at=c(16,round(sorted_hiszt_rep_5000_unisex_65[5000*0.05],digits=2)
,round(mean(hiszt_rep_5000_unisex_65),digits=2),round(sorted_hiszt_rep_50
00_unisex_65[5000*0.95],digits=2),20),labels=c(16,round(sorted_hiszt_rep_
5000_unisex_65[5000*0.05],digits=2),round(mean(hiszt_rep_5000_unisex_65),
digits=2),round(sorted_hiszt_rep_5000_unisex_65[5000*0.95],digits=2),20))

#####
# #
# ---- (SUB) SECTION: (4.6) ---- #
# #
#####

# creating function 'ujrageneralas_varhatoertekszamitas_duplafazis' in
order to create remining lifetimes from double-phase algorithm:
# -----
-----
ujrageneralas_varhatoertekszamitas_duplafazis<-
function(ax_gen,bx_gen,c_gen,k_1950_gen,t,kor,sigma_epsilon_negyzet_gen
,sigma_Z_negyzet_gen,lee_Carter_start,rep_varh_ertek,group){
  if(kor<length(ax_gen)-1){
    # legeneráljuk az 'epsilon' és a 'Z' független változókat:
    vector_random_epsilon<-rnorm(n=((2012-
lee_Carter_start+1)+t+(length(ax_gen)-1)-kor)*length(ax_gen)),mean=0,
sd=sqrt(sigma_epsilon_negyzet_gen))
    matrix_random_epsilon<-
matrix(vector_random_epsilon,nrow=length(ax_gen))
    vector_random_Z<-rnorm(n=((2012-lee_Carter_start)+t+(length(ax_gen)-
1)-kor),mean=0, sd=sqrt(sigma_Z_negyzet_gen))

```

```

# A fenti független változókból mortalitási rátákat készítünk 1950-
től 2012-ig:
matrix_mortality<-matrix(nrow=length(ax_gen),ncol=((2012-
Lee_Carter_start+1)+t+(length(ax_gen)-1)-kor)) # Üres mátrixot készítünk,
majd azt feltöltjük elemekkel.

```

```

for(ev in seq(from=Lee_Carter_start, to=2012+t+(length(ax_gen)-1)-
kor, by=1)){
  for(korev in seq(from=0,by=1,length.out=length(ax_gen))){ #
Vigyázat! A 'korev'-futóindex korábban 'kor' volt, csak rájöttem, hogy
nem szabad azt névnek adni, különben a függvény argumentumában szereplő
'kor' felülíródik.
    matrix_mortality[(korev+1),(ev-Lee_Carter_start+1)]<-
exp(ax_gen[korev+1]+bx_gen[korev+1]*(k_1950_gen+sum(vector_random_Z[seq(f
rom=1,by=1,length.out=(ev-Lee_Carter_start))])+(ev-
Lee_Carter_start)*c_gen)+matrix_random_epsilon[korev+1,ev-
Lee_Carter_start+1])
  }
}

```

```

varhato_elet_1<-0
for (ev in seq(from=1, to=(length(ax_gen)-1)-kor)){
  seged<-1
  for (j in seq(from=0, to=ev-1)){
    seged = seged*(1-matrix_mortality[kor+j+1,(2012-
Lee_Carter_start+1)+t+j+1])
  }
  if(ev<(length(ax_gen)-1)-kor){
    varhato_elet_1 = varhato_elet_1 +
ev*matrix_mortality[kor+ev+1,(2012-Lee_Carter_start+1)+t+ev+1]*seged
  }
  else{
    varhato_elet_1 = varhato_elet_1 + ev*1*seged
  }
}

```

```

# Újból alkalmazzuk a Lee-Carter modellt, azaz a megfelelő
paramétereket újból meghatározzuk ML-becsléssel:
# (!! ) Annyi, hogy a kitétség most tetszőleges lesz, később lehet,
hogy paraméterként fogom belevenni, de nem most!!!!

```

```

matrix_exposure<-matrix_exposure_female[seq(length(ax_gen)),seq(2012-
Lee_Carter_start+1)]

```

```

## evekbol keszitett vektorok:
vector_ages<-seq(from=0,by=1,length.out=length(ax_gen))
vector_years<-seq(from=Lee_Carter_start,to=2012,by=1)

```

```

if(group==1){
  name<-"male"
}
if(group==2){
  name<-"female"
}
if(group==3){
  name<-"unisex"
}

```

```

library("demography")

## es vegul a 'demogdata'-objektumok:
demogdata_mortality<-
demogdata(data=matrix_mortality[seq(length(ax_gen)),seq(2012-
Lee_Carter_start+1)], pop=matrix_exposure, ages=vector_ages,
years=vector_years, type="mortality", label="Human Mortality Database",
name=name, lambda=0)

### elkeszítjük a Lee-Carter-modellt:
lca<-lca(demogdata_mortality,years=(Lee_Carter_start:2012), adjust =
"dt")
ax<-lca$ax
bx<-lca$bx
kt<-lca$kt

ML_becslesek<-ML_becslesek(lca)
c<-ML_becslesek["c"]
szigma_epsilon_negyzet<-ML_becslesek["szigma_epsilon_negyzet"]
szigma_Z_negyzet<-ML_becslesek["szigma_Z_negyzet"]
varhato_elet_2<-
atlagolt_varhato_elet(ax,bx,c,kt[length(kt)],t,kor,szigma_epsilon_negyze
t,szigma_Z_negyzet,rep_varh_ertek)

varhato_eletek<-c(varhato_elet_1,varhato_elet_2)
names(varhato_eletek)<-c("elso_fazisbol","dupla_fazisbol")

return(varhato_eletek)
}
else{
return(0)
}
}

# generating remaining lifetimes from double-phase algorithm:
# -----
duplafazis_mintahalmaz_female_sz_2018<-matrix(nrow=5000,ncol=2)
colnames(duplafazis_mintahalmaz_female_sz_2018)<-
c("elso_fazisbol","dupla_fazisbol")
for(i in seq(5000)){
duplafazis_minta<-
ujrageneralas_varhatoertekszamitas_duplafazis(lca_female_1950_2012$ax,lca
_female_1950_2012$bx,ML_becslesek_female_1950_2012["c"],lca_female_1950_2
012$kt[1],6,0,ML_becslesek_female_1950_2012["szigma_epsilon_negyzet"],ML
_becslesek_female_1950_2012["szigma_Z_negyzet"],1950,1,2)
duplafazis_mintahalmaz_female_sz_2018[i,1]<-duplafazis_minta[1]
duplafazis_mintahalmaz_female_sz_2018[i,2]<-duplafazis_minta[2]
}
distance_female_sz_2018<-duplafazis_mintahalmaz_female_sz_2018[,1]-
duplafazis_mintahalmaz_female_sz_2018[,2]
# -----
duplafazis_mintahalmaz_female_65_2018<-matrix(nrow=5000,ncol=2)
colnames(duplafazis_mintahalmaz_female_65_2018)<-
c("elso_fazisbol","dupla_fazisbol")
for(i in seq(5000)){
duplafazis_minta<-
ujrageneralas_varhatoertekszamitas_duplafazis(lca_female_1950_2012$ax,lca

```

```

female_1950_2012$bx,ML_becslesek_female_1950_2012["c"],lca_female_1950_2
012$kt[1],6,65,ML_becslesek_female_1950_2012["szigma_epsilon_negyzet"],M
L_becslesek_female_1950_2012["szigma_Z_negyzet"],1950,1,2)
  duplafazis_mintahalmaz_female_65_2018[i,1]<-duplafazis_minta[1]
  duplafazis_mintahalmaz_female_65_2018[i,2]<-duplafazis_minta[2]
}
distance_female_65_2018<-duplafazis_mintahalmaz_female_65_2018[,1]-
duplafazis_mintahalmaz_female_65_2018[,2]
# -----
duplafazis_mintahalmaz_male_sz_2018<-matrix(nrow=5000,ncol=2)
colnames(duplafazis_mintahalmaz_male_sz_2018)<-
c("elso_fazisbol","dupla_fazisbol")
for(i in seq(5000)){
  duplafazis_minta<-
ujrageneralas_varhatoertekszamitas_duplafazis(lca_male_1980_2012$ax,lca_m
ale_1980_2012$bx,ML_becslesek_male_1980_2012["c"],lca_male_1980_2012$kt[1
],6,0,ML_becslesek_male_1980_2012["szigma_epsilon_negyzet"],ML_becslesek
_male_1980_2012["szigma_Z_negyzet"],1980,1,1)
  duplafazis_mintahalmaz_male_sz_2018[i,1]<-duplafazis_minta[1]
  duplafazis_mintahalmaz_male_sz_2018[i,2]<-duplafazis_minta[2]
}
distance_male_sz_2018<-duplafazis_mintahalmaz_male_sz_2018[,1]-
duplafazis_mintahalmaz_male_sz_2018[,2]
# -----
duplafazis_mintahalmaz_male_65_2018<-matrix(nrow=5000,ncol=2)
colnames(duplafazis_mintahalmaz_male_65_2018)<-
c("elso_fazisbol","dupla_fazisbol")
for(i in seq(5000)){
  duplafazis_minta<-
ujrageneralas_varhatoertekszamitas_duplafazis(lca_male_1980_2012$ax,lca_m
ale_1980_2012$bx,ML_becslesek_male_1980_2012["c"],lca_male_1980_2012$kt[1
],6,65,ML_becslesek_male_1980_2012["szigma_epsilon_negyzet"],ML_becslese
k_male_1980_2012["szigma_Z_negyzet"],1980,1,1)
  duplafazis_mintahalmaz_male_65_2018[i,1]<-duplafazis_minta[1]
  duplafazis_mintahalmaz_male_65_2018[i,2]<-duplafazis_minta[2]
}
distance_male_65_2018<-duplafazis_mintahalmaz_male_65_2018[,1]-
duplafazis_mintahalmaz_male_65_2018[,2]
# -----
duplafazis_mintahalmaz_unisex_sz_2018<-matrix(nrow=5000,ncol=2)
colnames(duplafazis_mintahalmaz_unisex_sz_2018)<-
c("elso_fazisbol","dupla_fazisbol")
for(i in seq(5000)){
  duplafazis_minta<-
ujrageneralas_varhatoertekszamitas_duplafazis(lca_unisex_1980_2012$ax,lca
unisex_1980_2012$bx,ML_becslesek_unisex_1980_2012["c"],lca_unisex_1980_2
012$kt[1],6,0,ML_becslesek_unisex_1980_2012["szigma_epsilon_negyzet"],ML
_becslesek_unisex_1980_2012["szigma_Z_negyzet"],1980,1,3)
  duplafazis_mintahalmaz_unisex_sz_2018[i,1]<-duplafazis_minta[1]
  duplafazis_mintahalmaz_unisex_sz_2018[i,2]<-duplafazis_minta[2]
}
distance_unisex_sz_2018<-duplafazis_mintahalmaz_unisex_sz_2018[,1]-
duplafazis_mintahalmaz_unisex_sz_2018[,2]
# -----
duplafazis_mintahalmaz_unisex_65_2018<-matrix(nrow=5000,ncol=2)
colnames(duplafazis_mintahalmaz_unisex_65_2018)<-
c("elso_fazisbol","dupla_fazisbol")
for(i in seq(5000)){

```

```

duplafazis_minta<-
ujrageneralas_varhatoertekszamitas_duplafazis(lca_unisex_1980_2012$ax,lca
_unisex_1980_2012$bx,ML_becslesek_unisex_1980_2012["c"],lca_unisex_1980_2
012$kt[1],6,65,ML_becslesek_unisex_1980_2012["szigma_epszilon_negyzet"],M
L_becslesek_unisex_1980_2012["szigma_Z_negyzet"],1980,1,3)
  duplafazis_mintahalmaz_unisex_65_2018[i,1]<-duplafazis_minta[1]
  duplafazis_mintahalmaz_unisex_65_2018[i,2]<-duplafazis_minta[2]
}
distance_unisex_65_2018<-duplafazis_mintahalmaz_unisex_65_2018[,1]-
duplafazis_mintahalmaz_unisex_65_2018[,2]

# creating histograms:
# -----
hist(breaks=5000/35,xaxt="n",freq=FALSE,right=FALSE,x=distance_female_sz_
2018,main="" ,xlab="Várható élettartamok különbségei",ylab="Relatív
gyakoriság")
abline(v=mean(distance_female_sz_2018), lwd=2, col="blue")
abline(v=sort(distance_female_sz_2018)[5000*0.05], lwd=2, col="orange")
abline(v=sort(distance_female_sz_2018)[5000*0.95], lwd=2, col="orange")
axis(1,at=c(-
10,round(sort(distance_female_sz_2018)[5000*0.05],digits=2),round(mean(di
stance_female_sz_2018),digits=2),round(sort(distance_female_sz_2018)[5000
*0.95],digits=2),10,15),labels=c(-
10,round(sort(distance_female_sz_2018)[5000*0.05],digits=2),round(mean(di
stance_female_sz_2018),digits=2),round(sort(distance_female_sz_2018)[5000
*0.95],digits=2),10,15))
# -----
hist(breaks=5000/35,xaxt="n",freq=FALSE,right=FALSE,x=distance_female_65_
2018,main="" ,xlab="Várható élettartamok különbségei",ylab="Relatív
gyakoriság")
abline(v=mean(distance_female_65_2018), lwd=2, col="blue")
abline(v=sort(distance_female_65_2018)[5000*0.05], lwd=2, col="orange")
abline(v=sort(distance_female_65_2018)[5000*0.95], lwd=2, col="orange")
axis(1,at=c(-
4,round(sort(distance_female_65_2018)[5000*0.05],digits=2),round(mean(dis
tance_female_65_2018),digits=2),round(sort(distance_female_65_2018)[5000*
0.95],digits=2),4),labels=c(-
4,round(sort(distance_female_65_2018)[5000*0.05],digits=2),round(mean(dis
tance_female_65_2018),digits=2),round(sort(distance_female_65_2018)[5000*
0.95],digits=2),4))
# -----
hist(breaks=5000/35,xaxt="n",freq=FALSE,right=FALSE,x=distance_male_sz_20
18,main="" ,xlab="Várható élettartamok különbségei",ylab="Relatív
gyakoriság")
abline(v=mean(distance_male_sz_2018), lwd=2, col="blue")
abline(v=sort(distance_male_sz_2018)[5000*0.05], lwd=2, col="orange")
abline(v=sort(distance_male_sz_2018)[5000*0.95], lwd=2, col="orange")
axis(1,at=c(-
10,round(sort(distance_male_sz_2018)[5000*0.05],digits=2),4,round(mean(di
stance_male_sz_2018),digits=2),round(sort(distance_male_sz_2018)[5000*0.9
5],digits=2),12),labels=c(-
10,round(sort(distance_male_sz_2018)[5000*0.05],digits=2),4,round(mean(di
stance_male_sz_2018),digits=2),round(sort(distance_male_sz_2018)[5000*0.9
5],digits=2),12))
# -----
hist(breaks=5000/35,xaxt="n",freq=FALSE,right=FALSE,x=distance_male_65_20
18,main="" ,xlab="Várható élettartamok különbségei",ylab="Relatív
gyakoriság")

```

```

abline(v=mean(distance_male_65_2018), lwd=2, col="blue")
abline(v=sort(distance_male_65_2018)[5000*0.05], lwd=2, col="orange")
abline(v=sort(distance_male_65_2018)[5000*0.95], lwd=2, col="orange")
axis(1,at=c(-
3,round(sort(distance_male_65_2018)[5000*0.05],digits=2),round(mean(dista
nce_male_65_2018),digits=2),round(sort(distance_male_65_2018)[5000*0.95],
digits=2),4),labels=c(-
3,round(sort(distance_male_65_2018)[5000*0.05],digits=2),round(mean(dista
nce_male_65_2018),digits=2),round(sort(distance_male_65_2018)[5000*0.95],
digits=2),4))
# -----
hist(breaks=5000/35,xaxt="n",freq=FALSE,right=FALSE,x=distance_unisex_sz_
2018,main="" ,xlab="Várható élettartamok különbségei",ylab="Relatív
gyakoriság")
abline(v=mean(distance_unisex_sz_2018), lwd=2, col="blue")
abline(v=sort(distance_unisex_sz_2018)[5000*0.05], lwd=2, col="orange")
abline(v=sort(distance_unisex_sz_2018)[5000*0.95], lwd=2, col="orange")
axis(1,at=c(-
8,round(sort(distance_unisex_sz_2018)[5000*0.05],digits=2),4,round(mean(d
istance_unisex_sz_2018),digits=2),round(sort(distance_unisex_sz_2018)[500
0*0.95],digits=2),12),labels=c(-
8,round(sort(distance_unisex_sz_2018)[5000*0.05],digits=2),4,round(mean(d
istance_unisex_sz_2018),digits=2),round(sort(distance_unisex_sz_2018)[500
0*0.95],digits=2),12))
# -----
hist(breaks=5000/35,xaxt="n",freq=FALSE,right=FALSE,x=distance_unisex_65_
2018,main="" ,xlab="Várható élettartamok különbségei",ylab="Relatív
gyakoriság")
abline(v=mean(distance_unisex_65_2018), lwd=2, col="blue")
abline(v=sort(distance_unisex_65_2018)[5000*0.05], lwd=2, col="orange")
abline(v=sort(distance_unisex_65_2018)[5000*0.95], lwd=2, col="orange")
axis(1,at=c(-
3,round(sort(distance_unisex_65_2018)[5000*0.05],digits=2),round(mean(dis
tance_unisex_65_2018),digits=2),round(sort(distance_unisex_65_2018)[5000*
0.95],digits=2),4),labels=c(-
3,round(sort(distance_unisex_65_2018)[5000*0.05],digits=2),round(mean(dis
tance_unisex_65_2018),digits=2),round(sort(distance_unisex_65_2018)[5000*
0.95],digits=2),4))

#####
#                                     #
# ---- (SUB) SECTION: Appendix ---- #
#                                     #
#####

# calculating other statistical indicators:
# -----
install.packages("moments")
library("moments")

### calculating other statistical indicators for subsection (4.6):
### -----
min(sorted_hiszt_rep_5000_female_sz)
max(sorted_hiszt_rep_5000_female_sz)
mean(sorted_hiszt_rep_5000_female_sz)
sqrt(var(sorted_hiszt_rep_5000_female_sz))

```



```

max(sorted_hiszt_rep_5000_female_sz)-min(sorted_hiszt_rep_5000_female_sz)
sorted_hiszt_rep_5000_female_sz[5000*0.05]
sorted_hiszt_rep_5000_female_sz[5000*0.25]
sorted_hiszt_rep_5000_female_sz[5000*0.50]
sorted_hiszt_rep_5000_female_sz[5000*0.75]
sorted_hiszt_rep_5000_female_sz[5000*0.95]
skewness(sorted_hiszt_rep_5000_female_sz)
kurtosis(sorted_hiszt_rep_5000_female_sz)
### -----
min(sorted_hiszt_rep_5000_female_65)
max(sorted_hiszt_rep_5000_female_65)
mean(sorted_hiszt_rep_5000_female_65)
sqrt(var(sorted_hiszt_rep_5000_female_65))
max(sorted_hiszt_rep_5000_female_65)-min(sorted_hiszt_rep_5000_female_65)
sorted_hiszt_rep_5000_female_65[5000*0.05]
sorted_hiszt_rep_5000_female_65[5000*0.25]
sorted_hiszt_rep_5000_female_65[5000*0.50]
sorted_hiszt_rep_5000_female_65[5000*0.75]
sorted_hiszt_rep_5000_female_65[5000*0.95]
skewness(sorted_hiszt_rep_5000_female_65)
kurtosis(sorted_hiszt_rep_5000_female_65)
### -----
min(sorted_hiszt_rep_5000_male_sz)
max(sorted_hiszt_rep_5000_male_sz)
mean(sorted_hiszt_rep_5000_male_sz)
sqrt(var(sorted_hiszt_rep_5000_male_sz))
max(sorted_hiszt_rep_5000_male_sz)-min(sorted_hiszt_rep_5000_male_sz)
sorted_hiszt_rep_5000_male_sz[5000*0.05]
sorted_hiszt_rep_5000_male_sz[5000*0.25]
sorted_hiszt_rep_5000_male_sz[5000*0.50]
sorted_hiszt_rep_5000_male_sz[5000*0.75]
sorted_hiszt_rep_5000_male_sz[5000*0.95]
skewness(sorted_hiszt_rep_5000_male_sz)
kurtosis(sorted_hiszt_rep_5000_male_sz)
### -----
min(sorted_hiszt_rep_5000_male_65)
max(sorted_hiszt_rep_5000_male_65)
mean(sorted_hiszt_rep_5000_male_65)
sqrt(var(sorted_hiszt_rep_5000_male_65))
max(sorted_hiszt_rep_5000_male_65)-min(sorted_hiszt_rep_5000_male_65)
sorted_hiszt_rep_5000_male_65[5000*0.05]
sorted_hiszt_rep_5000_male_65[5000*0.25]
sorted_hiszt_rep_5000_male_65[5000*0.50]
sorted_hiszt_rep_5000_male_65[5000*0.75]
sorted_hiszt_rep_5000_male_65[5000*0.95]
skewness(sorted_hiszt_rep_5000_male_65)
kurtosis(sorted_hiszt_rep_5000_male_65)
### -----
min(sorted_hiszt_rep_5000_unisex_sz)
max(sorted_hiszt_rep_5000_unisex_sz)
mean(sorted_hiszt_rep_5000_unisex_sz)
sqrt(var(sorted_hiszt_rep_5000_unisex_sz))
max(sorted_hiszt_rep_5000_unisex_sz)-min(sorted_hiszt_rep_5000_unisex_sz)
sorted_hiszt_rep_5000_unisex_sz[5000*0.05]
sorted_hiszt_rep_5000_unisex_sz[5000*0.25]
sorted_hiszt_rep_5000_unisex_sz[5000*0.50]
sorted_hiszt_rep_5000_unisex_sz[5000*0.75]
sorted_hiszt_rep_5000_unisex_sz[5000*0.95]

```

```

skewness(sorted_hiszt_rep_5000_unisex_sz)
kurtosis(sorted_hiszt_rep_5000_unisex_sz)
### -----
min(sorted_hiszt_rep_5000_unisex_65)
max(sorted_hiszt_rep_5000_unisex_65)
mean(sorted_hiszt_rep_5000_unisex_65)
sqrt(var(sorted_hiszt_rep_5000_unisex_65))
max(sorted_hiszt_rep_5000_unisex_65)-min(sorted_hiszt_rep_5000_unisex_65)
sorted_hiszt_rep_5000_unisex_65[5000*0.05]
sorted_hiszt_rep_5000_unisex_65[5000*0.25]
sorted_hiszt_rep_5000_unisex_65[5000*0.50]
sorted_hiszt_rep_5000_unisex_65[5000*0.75]
sorted_hiszt_rep_5000_unisex_65[5000*0.95]
skewness(sorted_hiszt_rep_5000_unisex_65)
kurtosis(sorted_hiszt_rep_5000_unisex_65)

# calculating other statistical indicators:
# -----
### calculating other statistical indicators for subsection (4.6):
### -----
min(distance_female_sz_2018)
max(distance_female_sz_2018)
mean(distance_female_sz_2018)
sqrt(var(distance_female_sz_2018))
max(distance_female_sz_2018)-min(distance_female_sz_2018)
sort(distance_female_sz_2018)[5000*0.05]
sort(distance_female_sz_2018)[5000*0.25]
sort(distance_female_sz_2018)[5000*0.50]
sort(distance_female_sz_2018)[5000*0.75]
sort(distance_female_sz_2018)[5000*0.95]
skewness(distance_female_sz_2018)
kurtosis(distance_female_sz_2018)
### -----
min(distance_female_65_2018)
max(distance_female_65_2018)
mean(distance_female_65_2018)
sqrt(var(distance_female_65_2018))
max(distance_female_65_2018)-min(distance_female_65_2018)
sort(distance_female_65_2018)[5000*0.05]
sort(distance_female_65_2018)[5000*0.25]
sort(distance_female_65_2018)[5000*0.50]
sort(distance_female_65_2018)[5000*0.75]
sort(distance_female_65_2018)[5000*0.95]
skewness(distance_female_65_2018)
kurtosis(distance_female_65_2018)
### -----
min(distance_male_sz_2018)
max(distance_male_sz_2018)
mean(distance_male_sz_2018)
sqrt(var(distance_male_sz_2018))
max(distance_male_sz_2018)-min(distance_male_sz_2018)
sort(distance_male_sz_2018)[5000*0.05]
sort(distance_male_sz_2018)[5000*0.25]
sort(distance_male_sz_2018)[5000*0.50]
sort(distance_male_sz_2018)[5000*0.75]
sort(distance_male_sz_2018)[5000*0.95]
skewness(distance_male_sz_2018)
kurtosis(distance_male_sz_2018)

```

```
### -----
min(distance_male_65_2018)
max(distance_male_65_2018)
mean(distance_male_65_2018)
sqrt(var(distance_male_65_2018))
max(distance_male_65_2018)-min(distance_male_65_2018)
sort(distance_male_65_2018) [5000*0.05]
sort(distance_male_65_2018) [5000*0.25]
sort(distance_male_65_2018) [5000*0.50]
sort(distance_male_65_2018) [5000*0.75]
sort(distance_male_65_2018) [5000*0.95]
skewness(distance_male_65_2018)
kurtosis(distance_male_65_2018)
### -----
min(distance_unisex_sz_2018)
max(distance_unisex_sz_2018)
mean(distance_unisex_sz_2018)
sqrt(var(distance_unisex_sz_2018))
max(distance_unisex_sz_2018)-min(distance_unisex_sz_2018)
sort(distance_unisex_sz_2018) [5000*0.05]
sort(distance_unisex_sz_2018) [5000*0.25]
sort(distance_unisex_sz_2018) [5000*0.50]
sort(distance_unisex_sz_2018) [5000*0.75]
sort(distance_unisex_sz_2018) [5000*0.95]
skewness(distance_unisex_sz_2018)
kurtosis(distance_unisex_sz_2018)
### -----
min(distance_unisex_65_2018)
max(distance_unisex_65_2018)
mean(distance_unisex_65_2018)
sqrt(var(distance_unisex_65_2018))
max(distance_unisex_65_2018)-min(distance_unisex_65_2018)
sort(distance_unisex_65_2018) [5000*0.05]
sort(distance_unisex_65_2018) [5000*0.25]
sort(distance_unisex_65_2018) [5000*0.50]
sort(distance_unisex_65_2018) [5000*0.75]
sort(distance_unisex_65_2018) [5000*0.95]
skewness(distance_unisex_65_2018)
kurtosis(distance_unisex_65_2018)
```