

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
BUDAPESTI CORVINUS EGYETEM
KÖZGAZDASÁGTUDOMÁNYI KAR

Függőkárok tartalékolásainak összehasonlítása

— Szakdolgozat —

Témavezető:

DR. MARTINEK LÁSZLÓ

Készítette:

LŐRINCZ ORSOLYA

Belső konzulens:

DR. ARATÓ MIKLÓS

Biztosítási és pénzügyi matematika

MSc. - Aktuárius szakirány



ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM



**BUDAPESTI
CORVINUS
EGYETEM**

Budapest, 2022

NYILATKOZAT

Név: Lőrincz Orsolya

ELTE Természettudományi Kar, szak: Biztosítási és pénzügyi matematika MSc

NEPTUN azonosító: XUPEAQ

Diplomamunka címe:

Függőkárok tartalékolásainak összehasonlítása

A **diplomamunka** szerzőjeként feyelmi felelősségem tudatában kijelentem, hogy a dolgozatom önálló szellemi alkotásom, abban a hivatkozások és idézések standard szabályait következetesen alkalmaztam, mások által írt részeket a megfelelő idézés nélkül nem használtam fel.

Budapest, 2022.05.26


a hallgató aláírása

Tartalomjegyzék

1. Bevezető	4
1.1. Tartalékolás a nem-életbiztosításban	4
1.2. Áttekintő az irodalomból	9
2. Módszerek és teljesítményük metrikái	12
2.1. Módszerek	12
2.1.1. Lánclétra módszer	12
2.1.2. Mack módszer	14
2.1.3. Bootstrapet használó módszerek	16
2.2. Metrika a becslések jóságára	22
2.2.1. Átlagos négyzetes eltérés (Mean square error of prediction (MSE))	23
2.2.2. Probability integral transform (PIT)	24
3. Felhasznált adatsorok	25
3.1. Homogén csoportok	25
3.1.1. Mack	26
3.1.2. Bootstrapet használó módszerek	29
3.2. Homogén csoportok normalizálva	33
3.2.1. Bootstrapet használó módszerek	34
3.3. Egyéni káradatok	39
3.3.1. Mack módszer	39
3.3.2. Bootstrapet használó módszerek	39

1. fejezet

Bevezető

1.1. Tartalékolás a nem-életbiztosításban

Miért tartalékolnak a biztosítók?

Egy biztosító optimális működéséhez, azon belül pedig fizetőképességének megőrzéséhez rengeteg tényező teljesülése szükséges, melyek közül az egyik legfontosabb, hogy megfelelő mennyiségű tartaléka álljon rendelkezésre. Ennek meghatározása igen bonyolult aktuáriusi feladat, ugyanis sem a túl kevés, sem a túl sok tartalék nem szerencsés. Ha túl kevés a tartalék a biztosító nem tudja teljesíteni a kifizetéseit, míg ha túltartalékol, azzal az eredménye (nyeresége) csökken.

Amikor a biztosító elad egy szerződést, valamekkora kötelezettséget vállal az ügyféllel szemben, melyet a jövőben teljesítenie kell, vagyis a biztosító bevételei és kiadásai időben elcsúsznak, ennek kezeléséhez van szükség a tartalékokra. Azt azonban nem feltétlenül tudhatja előre, hogy a teljesítésre mikor kerül sor vagy éppen mekkora összegben. Ha egy teljesen „egyszerű” szerződés életét vizsgáljuk, abban szerepel először a szerződés megkötésének időpontja (tegyük fel, hogy a biztosító ettől kezdve áll kockázatban), majd a szerződés tartama során valamikor egy kár következik be, amit az ügyfél valamikor – a kárbekövetkezés időpontja után – bejelent, majd megkezdődik

a kárbecslés. Miután megtörténik a kár nagyságának megállapítása, a megfelelő összeget kifizeti a biztosító és a kárrendezés lezárul. Ezen időpontok között akár évek is eltelhetnek, ha például pereskedésre kerül a sor (vagy a kár megtörténte csak évekkel később válik világossá) és előfordulhat, hogy az „ügy” lezárása után újra megnyitják azt és az is elképzelhető, hogy eközben a szerződés tovább él (esetleg további károk következnek be).

A nem-életbiztosítási tartalékok közül nagyságukat tekintve legnagyobb jelentőséggel a függőkárok tartaléka bír [1], melyet azokra a már bekövetkezett káreseményekre képzünk meg, melyek esetén nem vagy csak részben történt meg a kárkifizetés.

1.1. Definíció (IBNR tartalék). *Már bekövetkezett, de még be nem jelentett károokra képzett tartalék, az angol Incurred But Not Reported kifejezésből alkotott mozaikszó, más néven késői károk tartaléka.*

1.2. Definíció (RBNS). *Bejelentett, de még nem rendezett károk, az angol Reported But Not Settled kifejezésből.*

Kifutási háromszög

A múltra vonatkozó kárstatisztikai adatokat táblázatba rendezhetjük az eltelt idő figyelembevételével. Ez történhet a bekövetkezés és a bejelentés között eltelt idő vagy a bekövetkezés és a kifizetés között eltelt idő segítségével (esetleg a bejelentés és kifizetés közötti idő segítségével is, de ezt most nem fogjuk alkalmazni), mely két különböző táblázathoz vezet, ám ezek kifutása megegyezik [1]. Ez azt jelenti, hogy hosszú kifutási időt vizsgálva az adott évben bekövetkezett károk megegyeznek a kifizetett károkkal, mert a biztosító előbb-utóbb minden ügyet lezár. A kárstatisztikáknak is lehetséges több fajtája, például a kárösszegek jelenthetik a bekövetkezett és bejelentett károk összegét, de azt a mennyiséget is, mely az ügyfél számára kifizetésre került. Ez a kettő nem feltétlenül egyezik meg a szerződés életútjánál már korábban említett kárbecslési folyamat miatt, azonban az egyszerűség

kedvéért ezekre a későbbiekben összefoglalóan kifizetett károkként fogunk hivatkozni.

A kárkifizetési adatokat gyakran háromszögekbe rendezik, ahol a sorok jelölik a kárbekövetkezés időpontját (általában év), míg az oszlopok az azóta eltelt időt (általában szintén év). Az 1-es oszlopban szerepelnek a bekövetkezéssel megegyező évben kifizetett összegek.

Legyen $X_{i,j}$ az i -edik évben bekövetkezett károkra $(i+j-1)$ -edik évben kifizetett összeg, ahol $1 \leq i \leq I$ (I az utolsó kárbekövetkezési év) és $1 \leq j \leq J$ (J az utolsó év, amikor történik kifizetés), az általunk vizsgált esetekben speciálisan $I = J$. Az 1.1 ábrán látható egy szematikus kifizetési háromszög, ebben a fehér háttérű X -ek ismertek, például $X_{1,2}$ jelöli, hogy az első évben bekövetkezett károkra mennyi kifizetés történt a bekövetkezést követő évben. Ezzel szemben a szürke részek egyelőre ismeretlenek, az itt szereplő értékekre szeretnénk a továbbiakban valamilyen becslést adni. Azok az értékek tartoznak egy könyvelési évhez, ahol $i + j$ megegyezik (mellékátló), tehát ha egy évet szeretnénk előrejelezni vagy egy újabb év adataival szeretnénk bővíteni a háromszöget az azt jelenti, hogy az alsó szürke mellékátlót töltjük fel értékekkel, illetve válnak ismertté az ottani adatok.

i/j	1	2	3	...	J-1	J
1	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$...	$X_{1,J-1}$	$X_{1,J}$
2	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$...	$X_{2,J-1}$	$X_{2,J}$
3	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$...	$X_{3,J-1}$	$X_{3,J}$
...
I-1	$X_{I-1,1}$	$X_{I-1,2}$	$X_{I-1,3}$...	$X_{I-1,J-1}$	$X_{I-1,J}$
I	$X_{I,1}$	$X_{I,2}$	$X_{I,3}$...	$X_{I,J-1}$	$X_{I,J}$

1.1. táblázat. Kifizetési háromszög

Ez az 1.1 táblázatban megjelenő háromszög az egyik évről a másikra történő kifizetésekből áll elő, de az adatokból készíthető kumulált kárkifizetési

háromszög is $(C_{i,j})$, melynek elemei a következő módon állnak elő:

$$C_{i,j} = \sum_{k=1}^j X_{i,k}$$

Ekkor $C_{i,j}$ jelentése: az i -edik évben bekövetkezett károokra az $(i + j - 1)$ -edik év végéig kifizetett összeg, vagyis $C_{1,2}$ jelentése ebben az esetben, hogy az első bekövetkezési évben bekövetkezett károokra mennyi kifizetés történt a bekövetkezés utáni év végéig. Tartalékolási szempontból a feladat ebben az esetben is az alsó (szürke) háromszög adatainak becslése, a kárkifizetések előrejelzése. Az 1.2 ábra mutatja egy sematikus kumulált kifutási háromszög felépítését, míg az 1.3 ábra ugyanennek a jövőbeli alakját.

i/j	1	2	3	...	J-1	J
1	$C_{1,1}$	$C_{1,2}$	$C_{1,3}$...	$C_{1,J-1}$	$C_{1,J}$
2	$C_{2,1}$	$C_{2,2}$	$C_{2,3}$...	$C_{2,J-1}$	
3	$C_{3,1}$	$C_{3,2}$	$C_{3,3}$...		
...			
I-1	$C_{I-1,1}$	$C_{I-1,2}$				
I	$C_{I,1}$					

1.2. táblázat. Kumulált kifutási háromszög

i/j	1	2	3	...	J-1	J
1						
2						$C_{2,J}$
3					$C_{3,J-1}$	$C_{3,J}$
...				
I-1			$C_{I-1,3}$...	$C_{I-1,J-1}$	$C_{I-1,J}$
I		$C_{I,2}$	$C_{I,3}$...	$C_{I,J-1}$	$C_{I,J}$

1.3. táblázat. Kumulált jövőbeli kifutási háromszög

A könnyebb érthetőség érdekében nézzünk meg egy konkrét számokból álló példát is a kumulálatlan (növekményekből álló) és kumulált kárkifizetési háromszögekre: az 1.4 és 1.5 táblázatokban azonos adatok vannak feltüntetve egyszer úgy, hogy a növekmények vannak háromszögbe rendezve, másszor pedig a kumulált adatok. A bekövetkezés éve 1988 és 1997 között változik, ennek megfelelően az utolsó év, amikor kárkifizetést figyeltünk meg 1997.

	1	2	3	4	5	6	7	8	9	10
1988	79	227	146	103	61	25	14	2	3	1
1989	77	226	187	122	77	35	6	7	2	
1990	30	262	167	109	70	18	9	2		
1991	27	172	170	95	44	15	5			
1992	35	139	118	64	23	13				
1993	19	184	76	39	22					
1994	118	132	55	37						
1995	37	124	55							
1996	44	80								
1997	30									

1.4. táblázat. Példa kárkifizetési háromszögre

A 1.4 táblázatban látszik, hogy az 1988-ban bekövetkezett károokra 79 egységet fizettek ki 1988-ban, 227-et 1989-ben és így tovább. A kumulált háromszögben (1.5 táblázat) ez már úgy jelenik meg, hogy az 1988-ban bekövetkezett károokra 306 egységet ($79 + 227$) fizettek ki 1989 végéig. Az utolsó ismert év (1997) végéig kifizetett károk összege 4039 egység (1.5 táblázat utolsó mellékátlójában szereplő értékek összege, másképp az 1.4 táblázat minden ismert elemének összege). Az 1.5 táblázatban az 1988-as évre a végső kár 661 egység.

	1	2	3	4	5	6	7	8	9	10
1988	79	306	452	555	616	641	655	657	660	661
1989	77	303	490	612	689	724	730	737	739	
1990	30	292	459	568	638	656	665	667		
1991	27	199	369	464	508	523	528			
1992	35	174	292	356	379	392				
1993	19	203	279	318	340					
1994	118	250	305	342						
1995	37	161	216							
1996	44	124								
1997	30									

1.5. táblázat. Példa kumulált kárkifizetési háromszögre

1.3. Definíció (Végső károk). Amennyiben feltételezzük, hogy J év eltelte után már nem történik kárkifizetés, az utolsó oszlop, $C_{i,J}$ ($1 \leq i \leq I$) mutatja meg nekünk a végső kárt minden egyes bekövetkezési időszakra (évre), azaz mekkora a kárkifizetés (kárszükséglet) az i -edik bekövetkezési évben.

Már volt róla említés, hogy a bekövetkezéstől mind a bejelentésig, mind a kifizetésig terjedő időt vizsgálva azonos a háromszögek kifutása, ez azt jelenti, hogy a végső kár a két felírás esetén megegyezik. Az IBNR érték a kumulált háromszögben a (becsült) végső kár és az utolsó ismert év (mellékátló) kárkifizetései különbsége.

1.2. Áttekintő az irodalomból

A nem-élet tartalékolás modellezési módszerek köre jelentősen bővült az elmúlt 40 évben és további újabbnál újabb technikák kerülnek kidolgozásra, de a bonyolultabb modell nem feltétlenül jelent pontosabb modellt, ezért nem állítható, hogy a sztochasztikus modellek pontosabb előrejelzésre vezetnek mint egyszerűbb, determinisztikus társaik. Ugyanakkor a sztochaszti-

kus modelleknek nem elhanyagolható előnye, hogy segítségükkel becslés adható a kárkifizetések eloszlására, mely jelentős elmozdulás a korábbi pontbecslésekhez képest. A kiindulás a determinisztikus modellezés, ennek leg egyszerűbb módja a lánc-létra módszer [1], mely általában kumulált kifutási háromszögekkel dolgozik, alkalmazásával pontbecslés kapható. Népszerűsége az egyszerűségében rejlik, ugyanakkor azt is tudni kell, hogy kifejezetten érzékeny a felhasznált adatok változékonyságára, azon belül is a legújabb vizsgált évekre. Ezen túl rengeteg új modell kerül kifejlesztésre, melyek eredménye lehet nagyon hasonló, de nagyon eltérő is a lánc-létra módszer eredményeitől, ezért fontos lenne tudni az előrejelzések hibáját, mint az eredmények változékonyságának egyfajta mérőszámát. Mack modelljében [2] ez bevezetésre kerül, mely módszer továbbra is eloszlásmentes, a pontbecslést tekintve eredménye azonos a lánc-létrával, de egyfajta szórást is becsül az egy számot tartalmazó előrejelzés mellé. Komoly előrelépés a becslési eljárások terén amikor már sztochasztikussá válnak a modellek, melyek lehetővé teszik a mögöttes eloszlások becslését és szimulálását, ezzel együtt lehetőség nyílik a becslések hibájának vizsgálatára is, valamint az eredmények változékonyságára vonatkozóan is hordoz magában információt. A sztochasztikus modellek között előfordulnak olyanok, melyek azzal az elhatározással készültek, hogy azonos eredményt adjanak a determinisztikus eljárásokkal [3], ennek magyarázata, hogy a tartalék meghatározása mellett más fontos szempontjai is vannak a modelleknek, mint például a modell illeszkedése, a mögöttes eloszlás feltevések teljesülése, a becsült paraméterek változékonysága [4]. A sztochasztikus modellezések során gyakori a bootstrap [5] [6] [7] technika alkalmazása, mely visszatevéses mintavétel segítségével lehetővé teszi a pontosabb előrejelzést. A dolgozat későbbi fejezetében is lesz szó arról, hogy vajon hogyan befolyásolja az eredmények alakulását az adatok nagyságrendje és van-e értelme valamilyen módon azok normalizálásának például az adott év befolyt díjaival [8]. Az irodalom részletesen tárgyal további modelleket, de ezek nem kerülnek bemutatásra a dolgozat keretein belül. Ilyen a Kalman filter [9] [10] [11] és az egyéni szintű modellezés, mely a

sztochasztikus tartalékmeghatározás egy másik lehetséges megközelítése. Az egyéni szintű vizsgálat [12] lényege, hogy olyan homogén kockázatokú adatokkal dolgozik, melyben nincs olyan szerződés vagy szerződéscsoport, ami szignifikánsan befolyásolná a kifizetések összegét, ezért elsőre azt gondolhatnánk, hogy pontosabb becsléseket eredményez, mint egy aggregált portfólió, de ez nem minden esetben teljesül [13] [14].

El lehet továbbá azon is gondolkodni, hogy melyik a legjobb modell [15], egyáltalán van-e olyan, hogy valamelyik módszer a legjobb? Nagyon valószínűtlen, hogy az első kérdésre ugyanazt a választ adjuk determinisztikus és sztochasztikus eljárások esetében is, hiszen mindegyiknek más az erőssége és az eredményeket, hibákat nagyban befolyásolják a megfogalmazott feltevések. Emellett több szempont szerint dönthetünk arról is, hogy melyik módszert tartjuk jobbnak: amelyiknek legkisebb a hibája? Amelyik a legkevesebb munkával elkészíthető? Amelyik a legkevesebb feltevéssel a legpontosabb eredményt adja?

2. fejezet

Módszerek és teljesítményük metrikái

2.1. Módszerek

2.1.1. Lánc-létra módszer

A legegyszerűbb és egyik leggyakrabban alkalmazott determinisztikus, eloszlásmentes módszer a lánc-létra módszer, mely népszerűségét többek között annak köszönheti, hogy alkalmazása során viszonylag kevés feltételezéssel élünk az IBNR tartalék kiszámításához [1].

A feltételek most kumulált kifutási háromszögre lesznek megfogalmazva, de megmutatható, hogy az eredmények nagymértékben nem térnek [4] el a kumulált és a nem kumulált háromszögre való alkalmazás esetén.

A módszer alkalmazásához feltesszük, hogy

- a kifutási háromszögben a sorok (k, l) függetlenek

$$\{C_{k,1}, \dots, C_{k,J}\} \text{ és } \{C_{l,1}, \dots, C_{l,J}\}$$

ahol $1 \leq k$ és $l \leq J$ függetlenek egymástól, ha $k \neq l$, ami azt jelenti, hogy a különböző évek kárbekövetkezései nincsenek hatással egymásra

- léteznek olyan f pozitív számok (növekedési faktorok), hogy az évek során a kárkifizetések értéke úgy alakul, hogy mindig az előző évi kárkifizetés ezen pozitív számszorosára változik. Képletekkel:

$$\exists f_1, f_2, \dots, f_{J-1} > 0 : \forall 1 \leq i \leq I, 1 \leq j \leq J$$

$$\mathbb{E}(C_{i,j} | C_{i,1}, \dots, C_{i,j-1}) = f_{j-1} C_{i,j-1}$$

Az előrejelzéshez használt formula:

$$\hat{C}_{i,I} = C_{i,I-i+1} \hat{f}_{I-i+1} \dots \hat{f}_{I-1}$$

ahol $i+j > I$. A növekedési faktorok értéke csak nagyon ritka esetben ismert, így a felső háromszög segítségével ezekre is becslést használunk, mely szerint a növekedési faktorok oszlopösszegek hányadosaként állnak elő:

$$\hat{f}_j = \frac{\sum_{i=1}^{I-j+1} C_{i,j}}{\sum_{i=1}^{I-j+1} C_{i,j-1}}$$

Ezzel az összefüggéssel az utolsó kivételével minden oszlophoz meghatározható egy növekedési faktor, így már minden paraméter adott az alsó háromszög kitöltéséhez.

A 2.1 táblázatban egy példaszámolás eredményei láthatók. A felső (fehér) háromszögben vannak a kiindulási adatok, melyek azonosak az 1.5 táblázat számaival. A táblázat alján a becsült növekedési faktorok láthatók 4 tizedesjegyre kerekítve, ezek felhasználásával lett kitöltve az alsó (szürke) háromszög a jövőbeli kárkifizetések becslésével, melyek egészekre vannak kerekítve. Az utolsó oszlop harmadik sorában szereplő 670 úgy jött ki, hogy az utolsó ismert időszak 667-es értéke először meg lett szorozva 1.0036-tal (így jött ki a harmadik sor kilencedik oszlopában szereplő 669) majd 1.0015-tel. Ez a mátrix kumulált adatokat tartalmaz, de egy egyszerű differenciálással megkaphatjuk az évenként megjelenő kárkifizetéseket is.

	1	2	3	4	5	6	7	8	9	10
1988	79	306	452	555	616	641	655	657	660	661
1989	77	303	490	612	689	724	730	737	739	740
1990	30	292	459	568	638	656	665	667	669	670
1991	27	199	369	464	508	523	528	531	533	534
1992	35	174	292	356	379	392	397	399	401	401
1993	19	203	279	318	340	353	357	359	361	361
1994	118	250	305	342	377	391	397	399	400	401
1995	37	161	216	262	290	300	304	306	307	308
1996	44	124	188	228	252	261	265	266	267	268
1997	30	130	196	239	263	273	277	278	279	280
f	4.3176	1.5159	1.2150	1.1034	1.0375	1.0134	1.0054	1.0036	1.0015	

2.1. táblázat. Példa a lánc-létra módszerre

Arról már volt szó, hogy miért olyan népszerű a lánc-létra módszer, azonban mint minden módszernek, ennek is vannak gyengeségei. Ez a módszer csupán egy értéket ad vissza eredménynek és semmit nem tudunk arról mondani, hogy ez a becslés mennyire jó, mekkora a szórása, milyen standard hibával rendelkezik és kockázatszámításra sem alkalmas. Annyit tudunk csak, hogy érzékeny az adatok változásaira, de hogy mennyire azt nem, ami nehézkessé teszi a más módszerekkel való összehasonlítást. Ebben tudott előrelépést mutatni a Mack módszer, melyben a pontbecslés mellett már legalább egy szórást is tudunk mondani a változékonyságra.

2.1.2. Mack módszer

A Mack módszer a számolás menetét tekintve nagyon hasonlít a lánc-létra módszerre [2], szintén eloszlásmentes, de itt már szórás és standard hiba is adható a tartalékbecslésekhez, mely segíti az eredmények összehasonlíthatóságát, valamint egyfajta képet adhat a becslés bizonytalanságáról. A becslés során a következő feltételezésekkel élünk:

- A bekövetkezési időszakok (évek) kumulált káralakulásai függetlenek egymástól:

$$\{C_{k,1}, \dots, C_{k,J}\}, \{C_{l,1}, \dots, C_{l,J}\}$$

ahol $1 \leq k$ és $l \leq J$ (sorok) függetlenek egymástól, ha $k \neq l$

- Léteznek pozitív növekedési faktorok úgy, hogy

$$\exists f_1, f_2, \dots, f_{J-1} > 0 : \forall 1 \leq i \leq I, 1 \leq j \leq J$$

$$\mathbb{E}(C_{i,j} | C_{i,1}, \dots, C_{i,j-1}) = f_{j-1} C_{i,j-1}$$

ekkor az előrejelzés csak a közvetlen múlttól és a növekedési faktoroktól függ.

- A növekedési faktorokhoz hasonlóan feltehetjük, hogy létezik egy paraméter a szórásra (σ_j) is, mely segítségével előállítható minden évre a háromszögbeli becslés szórása az előző időszakai értékből és ebből a paraméterből szorzás segítségével:

$$\mathbb{D}^2(C_{ij} | C_{i,1}, \dots, C_{i,j-1}) = \sigma_j^2 C_{i,j-1}$$

A növekedési faktorok becslése a lánc-létra módszerrel megegyezően a megfelelő oszlopösszegek hányadosaként áll elő:

$$\hat{f}_j = \frac{\sum_{i=1}^{I-j+1} C_{i,j}}{\sum_{i=1}^{I-j+1} C_{i,j-1}}$$

Illetve a szórás (torzítatlan) becslése:

$$\hat{\sigma}_j^2 = \frac{1}{I-j-1} \sum_{i=1}^{I-j} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^2$$

$$1 \leq j \leq I-2$$

2.1.3. Bootstrapet használó módszerek

Ebben a megközelítésben a modellezés magja egy Általánosított Lineáris Modell (Generalized Linear Model, GLM) azzal a tulajdonsággal, hogy a várható érték ($\mathbb{E}(C_{i,j}) = m_{i,j}$) multiplikatív módon áll elő [15]:

$$m_{i,j} = e^{c+\alpha_i+\beta_j} \text{ és } \alpha_1 = \beta_1 = 0$$

Így még mindig egy lánc-létra típusú struktúránk van abban az értelemben, hogy minden sorhoz és oszlophoz tartozik egy paraméter [4]. Előnye, hogy könnyen becsülhetők a paraméterek, hátránya azonban, hogy ezek a paraméterek nehezen értelmezhetők. Ez a $C_{i,j}$ értékek mögötti GLM feltételezés segít a sztochasztikus modellezés első lépésében, hogy modellt illesszünk az adatokra [16]. Ezután következik a bootstrap technikát használó rész, ahol a reziduálisok segítségével pszeudo adatokat hozunk létre és végzünk velük becsléseket a megfelelő feltételezett eloszlásokkal a jövőre nézve amellet a feltételezés mellett, hogy a $C_{i,j}$ értékek overdispersed Poisson vagy Gamma eloszlást követnek.

Overdispersed Poisson eloszlás:

Az overdispersed Poisson eloszlás abban különbözik a Poisson eloszlástól, hogy a várható értéke nem egyenlő a szórásával (van benne egy diszperziós paraméter). Erre azért van szükség, mert a megfigyelésekre nem igaz, hogy a várható érték megegyezik a szórással, de ezzel a konstans szorzóval már jobban illeszkedik a Poisson eloszlás [3] [15].

Ha $Y_{ij} \sim \text{Poisson}(\frac{\mu_i \gamma_i}{\phi})$, ahol $\phi \in R$, akkor $X_{ij} := \phi Y_{ij} \sim$ overdispersed Poisson ϕ diszperziós paraméterrel, ahol μ_i a kárbekövetkezési évre, a γ_i az eltelt évekre utal.

Gamma eloszlás:

Egy X valószínűségi változó p -edrendű λ paraméterű gamma eloszlású, ha sűrűségfüggvénye

$$f(x) = \frac{\lambda^p x^{p-1} e^{-\lambda x}}{\Gamma(p)}$$

ahol $\Gamma(p)$ a gamma-függvény, λ és p pozitív.

Bootstrap technika lépései egy példán:

A bootstrap technika lépéseinek áttekintéséhez [4] [5] [17] egy nagyon egyszerű kumulált kiindulási háromszöget használunk, amit $C_{i,j}$ jelöl, értékeit pedig a 2.2 táblázat tartalmazza.

	1	2	3	4
1	100	200	300	400
2	300	500	600	
3	200	300		
4	400			

2.2. táblázat. Bootstrap példa kiinduló kumulált háromszöge

1. A kumulált kifizetési háromszögből meghatározzuk a növekedési faktorokat:

	1	2	3
f	1.6	1.2857	1.3

2.3. táblázat. Növekedési faktorok

2. Az utolsó ismert átló értékeiből kiindulva a növekedési faktorok (f) segítségével rekurzívan egy becslést ($\hat{C}_{i,j}$) adunk a múltbeli kumulált háromszögre:

$$\hat{C}_{i,I-i+1} = C_{i,I-i+1}$$

$$\hat{C}_{i,k-1} = \hat{C}_{i,k} f_k^{-1}, \text{ ahol } 1 \leq k \leq I - i$$

	1	2	3	4
1	140	233.3	300	400
2	280	466.6	600	
3	180	300		
4	400			

2.4. táblázat. Visszabecsült kumulált értékek

3. A 2.4 táblázatból differenciálással számolható a becült (kumulátlan) kárkifizetések háromszöge, \hat{m}_{ij} :

	1	2	3	4
1	140	93.3	66.6	100
2	280	186.6	133.3	
3	180	120		
4	400			

2.5. táblázat. Visszabecsült kumulátlan értékek

4. Pearson reziduális meghatározása a kumulátlan háromszögben:

$$r_{ij}^{(P)} = \frac{X_{ij} - \hat{m}_{ij}}{\sqrt{\hat{m}_{ij}}}$$

	1	2	3	4
1	-3.3806	0.6901	4.0825	0.0000
2	1.1952	0.9759	-2.8868	
3	1.4907	-1.8257		
4	0.0000			

2.6. táblázat. Pearson reziduális értékek

5. Pearson skálaparaméter (ϕ) meghatározása: olyan hányados, melynek számlálója az imént meghatározott reziduálisok négyzetösszege, nevezője pedig a szabadságfok. A szabadságfok a megfigyelések számának $\left(\frac{I(I+1)}{2}\right)$ és a becsült paraméterek számának $(2I - 1)$ különbsége:

$$\phi = \frac{\sum_{i=1}^I \sum_{j=1}^{I-i+1} \left(r_{ij}^{(P)}\right)^2}{\frac{1}{2}I(I+1) - 2I + 1}$$

A 2.6 táblázat alapján ez a paraméter most $\phi = 44.8413$.

6. Igazított Pearson reziduálisok meghatározása:

$$r_{ij}^{adj} = \sqrt{\frac{\frac{1}{2}I(I+1)}{\frac{1}{2}I(I+1) - 2I + 1}} \times r_{ij}^{(P)}$$

	1	2	3	4
1	-6.1721	1.2599	7.4536	0.0000
2	2.1822	1.7817	-5.2705	
3	2.7217	-3.3333		
4	0.0000			

2.7. táblázat. Igazított Pearson reziduális értékek

7. Most kezdődik az iterálás: a valós adatokkal való számolásnál $N = 1000$ ismétléssel fogunk dolgozni.

- (a) Visszatevéssel mintát veszünk r_{ij}^{adj} -kből, a 2.7 táblázat értékeiből, ezzel létrehozva egy új reziduálisokból álló háromszöget:

	1	2	3	4
1	2.7217	2.7217	0.0000	2.1822
2	1.7817	1.2599	0.0000	
3	1.7817	-5.2705		
4	0.0000			

2.8. táblázat. Reziduálisok háromszöge a mintavételből

- (b) Az $r_{ij}^{(P)}$ -re felírt egyenletet X_{ij} -re rendezve egy növekményekből álló pszeudo-háromszöget kapunk a múltbeli adatokra:

	1	2	3	4
1	172.2031	119.6270	66.6667	121.8218
2	309.8142	203.8799	133.3333	
3	203.9046	62.2650		
4	400.0000			

2.9. táblázat. Növekmények pszeudo-háromszöge

- (c) Alakítsuk át a 2.9 táblázatot kumulált háromszöggé:

	1	2	3	4
1	172.2031	291.8301	358.4967	480.3185
2	309.8142	513.6942	647.0275	
3	203.9046	266.1695		
4	400.0000			

2.10. táblázat. Kumulált pszeudo-háromszög

- (d) A 2.10 táblázatra alkalmazzuk a lánc-létra módszert és először határozzuk meg a növekedési faktorokat (2.11 táblázat)...

	1	2	3
f	1.5624	1.2483	1.3398

2.11. táblázat. Kumulált pszeudo-háromszög növekedési faktorai

(e) ... majd jelezzük előre az alsó háromszögmátrixot:

	1	2	3	4
1				
2				866.8957
3			332.2556	445.1603
4		624.9655	780.1354	1045.2354

2.12. táblázat. Lánc-létra eredményei

(f) A 2.12 táblázatból differenciálással egy lépésben megkaphatjuk a lánc-létra módszer eredményeinek növekményes felírását, $\tilde{m}_{i,j}$:

	1	2	3	4
1				
2				219.8682
3			66.0860	112.9047
4		224.9655	155.1699	265.1000

2.13. táblázat. Lánc-létra növekményes eredményei

(g) Az jövőbeli háromszög minden cellájába szimulálunk egy kifizetést a választott eloszlásból (a példában overdispersed Poisson, a valós adatokon emellett Gamma is) $\tilde{m}_{i,j}$ várható értékkel (2.13 táblázat) és $\phi\tilde{m}_{i,j}$ szórással

	1	2	3	4
1				
2				163
3			1	88
4		83	248	288

2.14. táblázat. Szimulált overdispersed Poisson értékek

- (h) Ha az így kapott előrejelzéseket kárbekövetkezési évenként összegezzük, megkapjuk a becsült IBNR értékeket. Ha ehhez még pluszba hozzáadjuk a kumulált háromszög utolsó ismert átlóját a 2.2 táblázatból, akkor eljutunk az egy forgatókönyvhöz tartozó becsült végső kárhoz évenkénti bontásban.

	Utolsó ismert mellékátló	Becsült IBNR	Becsült végső kár
1	400	0	400
2	600	163	763
3	300	89	389
4	400	619	1 019

2.15. táblázat. Egy forgatókönyv eredményei

- (i) Eltároljuk az eredményeket és továbblépünk a következő iterációra.

2.2. Metrika a becslések jóságára

A modellek összehasonlítására két lehetséges módot nézünk meg, melyek nem biztosítás specifikusak, általánosságban a becsült adatok és eloszlásuk megfigyelt adatokra illetve azok eloszlására való illeszkedését, valamint annak megfelelőségét vizsgálják.

2.2.1. Átlagos négyzetes eltérés (Mean square error of prediction (MSE))

Különböző modellek összehasonlításának egyik lehetséges módja az átlagos négyzetes eltérés vizsgálata. Ez a mutató a becslési eljárás során kapott értékek átlagos négyzetes távolságát vizsgálja a megfigyeltektől [15].

Legyen C_{ij} egy kumulált kárkifizési háromszög, ebben a végső károk az utolsó oszlopban szereplő értékek ($C_{i,J}$), ezekre adnak a modellek valamilyen becslést ($\hat{C}_{i,J}$). Az összes rendelkezésre álló információ

$$D = \{C_{i,j} | i + j \leq J + 1\}$$

. Ezek segítségével az előrejelzés átlagos négyzetes hibája:

$$MSE(\hat{C}_{i,J}) = \mathbb{E}((\hat{C}_{i,J} - C_{i,J})^2 | D)$$

Példa: Tegyük fel, hogy ismerjük a 2.2 táblázat utolsó oszlopát, ezek a „valódi” végső károk és kétszer végeztük el az iterációt a bootstrap technikával. Felhasználva a 2.15 táblázat utolsó oszlopát már kiszámolható az átlagos négyzetes eltérés minden évre, ez szerepel a 2.16 táblázat utolsó oszlopában. Minél közelebb van ez a szám nullához, annál pontosabbnak mondható a becslés.

	1. forgatókönyv becslése	2. forgatókönyv becslése	„Valódi” végső kár	Átlagos négyzetes eltérés
1	400	400	400	0
2	763	765	700	4 097
3	389	427	500	8 825
4	1 019	1 028	800	15 273

2.16. táblázat. Átlagos négyzetes eltérés 2 forgatókönyv „valódi” végső kárra adott becslése alapján

MSE a Mack módszerben: Legyen D az eddig ismert kárkifizési értékek halmaza:

$$D = \{C_{ik} | i + k \leq I + 1\}$$

Az átlagos négyzetes hiba továbbra is

$$MSE(\hat{C}_{i,J}) = \mathbb{E}((\hat{C}_{i,J} - C_{i,J})^2 | D)$$

Ekkor megmutatható a $\mathbb{E}(X - a)^2 = \mathbb{D}^2(X) + (\mathbb{E}(X) - a)^2$ összefüggés felhasználásával [2], hogy

$$MSE(\hat{C}_{i,I}) = \mathbb{D}^2(C_{i,I} | D) + (\mathbb{E}(C_{i,I} | D) - \hat{C}_{i,I})^2$$

ahol $\hat{C}_{i,I}$ a $C_{i,I}$ becslése. Tehát az átlagos négyzetes eltérés szétbontható a folyamat varianciájának és a becslési hibának az összegére.

2.2.2. Probability integral transform (PIT)

Az illeszkedés vizsgálatának egy másik lehetséges megközelítése, hogy az eloszlásokat nézzük hisztogramok segítségével [15]. Tegyük fel, hogy van egy abszolút folytonos eloszlásból származó X megfigyelésünk. Ekkor a megfigyelés a saját eloszlásfüggvényébe visszahelyettesítve legyen $Y = F_X(X)$. Ha létezik F_X^{-1} és $y \in [0, 1]$, akkor

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(F_X(X) \leq y) = \mathbb{P}(X \leq F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y$$

Vagyis a $[0, 1]$ intervallumon egyenletes eloszlást kell kapnunk. A valódi adatokon végrehajtott modellezésénél az egyes forgatókönyvek által becsült teljes végső károk alkotják majd az F eloszlásfüggvényt, amit pedig ebbe behelyettesítünk az a tényleges megfigyelt végső kár értéke lesz. A $[0, 1]$ -en való egyenletes eloszlás hisztogramok segítségével ellenőrizhető, valamint a hisztogram különböző egyenletestől való eltéréseiből a becslés viselkedésére is lehet következtetni. Ha saját eloszlásfüggvénybe visszahelyettesített becslések hisztogramja U alakot vesz föl, a becslés aluszóródott, túl szűk az előrejelzés intervalluma, míg a közepen kidomborodó hisztogram túlszóródott becslésre utal, ahol az előrejelzés intervalluma túlságosan széles. Emellett még fontos azt is észben tartani, hogy a jó becslésnek csupán szükséges feltétele a $[0, 1]$ -en való egyenletes eloszlás, nem elégséges. A „szép” hisztogramra nem mondhatjuk biztosan, hogy jó becslés, az erősen torzultról viszont mondhatjuk, hogy semmiképp sem jó becslés.

3. fejezet

Felhasznált adatsorok

3.1. Homogén csoportok

A módszerek gyakorlati alkalmazásához a NAIC (National Association of Insurance Commissioners) honlapján elérhető adatsorokat használtam, itt összesen 6 homogén csoport adatai érhetők el táblázatba rendezve [17][18]. Az első oszlopban szerepelnek azonosító kódok, az egy kódhoz tartozó adatokra egy vállalat adataiként fogunk hivatkozni (például a 2003 kódú vállalat). A további oszlopokban megtalálhatók a bekövetkezési évek, az eltelt évek száma, a bekövetkezett károk és a kifizetett károk kumuláltan és az adott évhez tartozó megszolgált díjak. Ezeknek a későbbiekben az adatok egyfajta normalizálásában lesz szerepe. Az adatsorok nagyszerű jellemzője, hogy 19 év megfigyeléseit tartalmazza, így 10 éves késéseket használva felírható belőlük egy kárkifizési háromszög, amelyet aztán a fenti módszerekkel lehet négyesgésíteni, miközben az alsó háromszög „valós” adatai is ismertek, ami lehetővé teszi a módszerek tartalék eredményeinek összehasonlítását a tényleges megfigyelésekkel.

Az R-ben történő adatfeldolgozás során vállalatonként készült egy-egy kárkifizési háromszög, ahol a bekövetkezés éve 1988 és 1997 között változik, az utolsó megfigyelés 1997-ben történt, a háromszögben szereplő értékek pedig a kumulált kárkifizetések. Az alkalmazás egyik első lépése az adattisztítás

volt, ugyanis szerepeltek olyan vállalatok ezekben az adatsorokban, melyeknek az első kárbekövetkezési évben, néhol egyáltalán nem volt kárkifizetésük (a kifutási háromszög első sora csupa nullából állt), ami nem teszi lehetővé a Mack módszer alkalmazását, mert a növekedési faktorok meghatározásánál a nevezőben 0 szerepelt volna, így bizonyos vállalatok kikerültek a vizsgálatból. Ezt a kiválogatást mind a 6 homogén csoportra elvégezve a ténylegesen előrejelzésre használható vállalatok számát a 3.1 táblázat foglalja össze.

Homogén csoport	Vállalatok száma
Private passenger auto liability/medical	87
Commercial auto/truck liability/medical	80
Workers' compensation	56
Medical malpractice – claims made	12
Other liability – occurrence	93
Product liability – occurrence	14

3.1. táblázat. Felhasznált vállalatok száma homogén csoportonként

A további részekben a Commercial auto/truck liability/medical homogén csoportra különböző módszerekkel végzett előrejelzések eredményei vannak leírva.

3.1.1. Mack

Az első módszer az előrejelzésre, a korábbiakban már említett Mack módszer volt. Minden vállalatra, ami a Commercial auto/truck liability/medical homogén csoport esetén 80 vállalatot jelent, a felső háromszög mátrix alapján a módszer segítségével kiegészítésre került az alsó háromszög, így minden bekövetkezési évre kaptunk egy végső kár értéket, amelyet össze lehetett hasonlítani a megfigyelt végső kárral. Emellett bár a Mack módszer eloszlásmentes, a becslések szórására mégis tud mondani egy értéket. A 3.2 táblázatban a felső háromszög értékei láthatók, melyből a módszer kiindul,

a 3.3 táblázatban pedig a Mack módszer által kapott becslések olvashatók egészekre kerekítve.

	1	2	3	4	5	6	7	8	9	10
1988	79	306	452	555	616	641	655	657	660	661
1989	77	303	490	612	689	724	730	737	739	
1990	30	292	459	568	638	656	665	667		
1991	27	199	369	464	508	523	528			
1992	35	174	292	356	379	392				
1993	19	203	279	318	340					
1994	118	250	305	342						
1995	37	161	216							
1996	44	124								
1997	30									

3.2. táblázat. Comauto 2003 kódú vállalatának kiinduló háromszöge

	1	2	3	4	5	6	7	8	9	10
1988										
1989										740
1990									669	670
1991								531	533	534
1992							397	399	401	401
1993						353	357	359	361	361
1994					377	391	397	399	400	401
1995				262	290	300	304	306	307	308
1996			188	228	252	261	265	266	267	268
1997	130	196	239	263	273	277	278	279	280	

3.3. táblázat. Comauto 2003 kódú vállalatának előrejelzése Mack módszerrel

A 3.4 táblázat foglalja össze a Commercial auto/truck liability/medical homogén csoport 2003 kódú vállalatának adatain Mack-módszerrel végzett

előrejelzéseknek néhány jellemzőjét kárévenkénti megbontásban. Az első oszlop a módszer által becsült végső károkat mutatja, a második oszlop a becsült végső kár és a megfigyelt „valódi” végső kár között eltérést úgy, hogy a becsültből vonja ki a „valódi”-t ($\hat{C}_{i,I} - C_{i,I}$). Ez alapján azt lehet mondani, hogy a módszer inkább alulbecsülte a szükséges tartalék nagyságát, de persze előfordult olyan év is (1988-at nem számolva 3), hogy a becslés szerint több tartalékra lett volna szükség, mint ami végül a valóságban kellett. Ez különösen igaz az utolsó évre, melyben akkora pozitív eltérés szerepel, mely az oszlop összegét is pozitívvá változtatja. A harmadik oszlopban a kárévekhez tartozó szórások szerepelnek, ahol megfigyelhető, hogy minél későbbi évet nézünk annál nagyobb a szórás értéke, melynek magyarázata lehet, hogy míg 1988-ban tulajdonképp nem kellett becslést adni a végső kárra, mert már ismert volt, addig az 1997-es bekövetkezési év végső kár becsléséhez 9 másik becslésen keresztül vezet az út, melyek már önmagukban rendelkezhetnek valamekkora eltéréssel a megfigyelt értékekhez képest.

	Végső kár ($\hat{C}_{i,I}$)	Eltérés ($\hat{C}_{i,I} - C_{i,I}$)	Szórás ($\hat{\sigma}_j$)
1988	661	0.00	0.00
1989	740	1.12	0.47
1990	670	-9.59	1.17
1991	534	-2.46	2.79
1992	401	-13.59	4.09
1993	361	-6.80	6.05
1994	401	5.89	13.06
1995	308	-2.36	20.92
1996	268	-1.28	52.70
1997	280	51.65	226.99

3.4. táblázat. Comauto 2003 kódú vállalatának eredményei Mack módszerrel

3.1.2. Bootstrapet használó módszerek

Elhagyva a determinisztikus módszerek mezejét a következő sorra kerülő becslési eljárásokhoz a bootstrap technikát is felhasználtuk, mely a különböző kártartalékolási módszerekhez segít végső káreloszlást rendelni és azok becslés hibáját meghatározni. Jelen esetben a C_{ij} értékekre először azt feltételezzük, hogy overdispersed Poisson majd, hogy Gamma eloszlást követnek. A metrend itt is nagyon hasonló az előzőekhez: minden vállalatra egy kiindulási háromszögből R segítségével előrejelzések készítése a jövőbeni kárkifizésekre. A becslés során a bootstrap 1 000 különböző forgatókönyvet hoz létre, így minden vállalatra 1 000 végső kár eredményünk lesz, itt is kárévenkénti megbontásban. Az eredmények vizsgálata itt annyiban módosul, hogy a végső kárra itt nem csupán egy értéket kapunk, hanem rögtön 1 000-et minden bekövetkezési évre, ezek átlaga szerepel az első, szórása a második oszlopban összegyűjtve. A harmadik oszlop a becslések átlagos négyzetes eltérése a tényleges megfigyelt értéktől (MSE). Ennek kiszámolása úgy történik, hogy mind az 1 000 forgatókönyv esetén kapunk minden évre egy becslést a végső kárra, melyből először kivonjuk az adott évre megfigyelt „valódi” végső kárt. Ezután ezeket a különbségeket négyzetre emeljük, majd minden évre az így kapott 1000 érték átlagát vesszük. Az utolsó oszlop az 1 000 forgatókönyv 99.5%-os kvantilisét mutatja éves bontásban, ez az az összeg, melynél az 1 000 szimuláció során csupán 5 nagyobb becsült végső kár szerepelt.

Overdispersed Poisson

A 3.5 táblázatban a Commercial auto/truck liability/medical homogén csoport 2003 kódú vállalatának adataiból kiindulva kapott eredmények szerepelnek overdispersed Poisson eloszlást feltételezve, a bootstrap segítségével 1 000 forgatókönyvet szimulálva. Ebben a táblázatban is ugyanaz a tendencia mint a Mack módszer eredményeinek 3.3 táblázatánál, vagyis itt is megfigyelhető, hogy minél későbbi évet nézünk annál nagyobb az adatok szórása és ezzel együtt egyre nagyobb a különbség a becsült végső kár és a 99.5%-os kvantilis között. Az első évben továbbra is 0 az eltérés és a szórás is 0, a

végső kár pedig megegyezik a 99.5%-os kvantilissal is, mert erre az évre tulajdonképp nem történik becslés, csak felhasználásra kerül a már megfigyelt adat.

	Becsült végső károk átlaga	Becsült végső károk szórása	Átlagos négyzetes eltérés	Kvantilis (99.5%)
1988	661.00	0.00	0.00	661.00
1989	740.10	5.72	33.89	768.00
1990	670.12	7.13	148.49	702.00
1991	533.62	9.51	96.00	574.02
1992	401.95	12.15	317.72	454.03
1993	361.41	16.02	299.77	414.01
1994	402.14	27.21	790.69	487.03
1995	306.88	37.55	1 418.15	430.03
1996	269.84	52.71	2 776.57	451.01
1997	288.77	162.52	30 080.09	1 152.07

3.5. táblázat. Comauto 2003 kódú vállalatának eredményei overdispersed Poisson eloszlással

Gamma

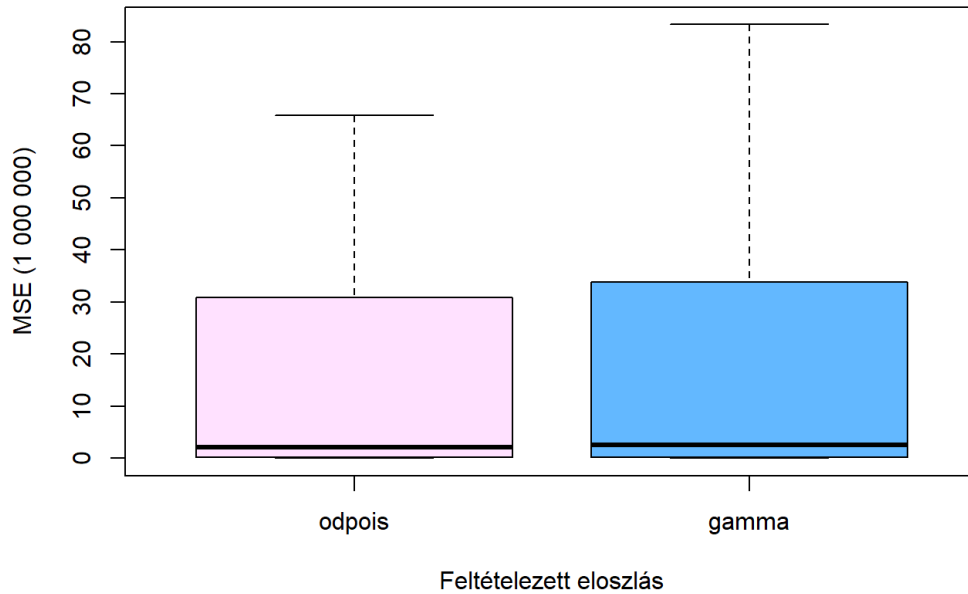
A 3.6 táblázatban vannak a Commercial auto/truck liability/medical homogén csoport 2003 kódú vállalatának adataiból kiindulva kapott eredmények Gamma eloszlást feltételezve, a bootstrap segítségével 1 000 forgatókönyvet szimulálva. Erre a konkrét vállalatra elmondható, hogy az előrejelzések nagyon hasonlóan alakultak a 3.5 táblázathoz, a becslések szórását és az átlagos négyzetes eltérés értékeket a különböző évekre vizsgálva azonban nem vonható le olyan következtetés, hogy a becslési hibák az egyik feltételezett eloszlás esetén határozottan kisebbek lennének.

	Becsült végső károk átlaga	Becsült végső károk szórása	Átlagos négyzetes eltérés	Kvantilis (99.5%)
1988	661.00	0.00	0.00	661.00
1989	740.21	5.24	28.91	765.16
1990	670.83	8.72	160.08	715.89
1991	533.44	8.87	85.19	569.99
1992	401.42	10.47	294.09	440.05
1993	361.40	15.60	286.67	420.16
1994	401.20	27.65	802.29	496.02
1995	307.64	36.38	1 327.57	421.72
1996	268.43	53.82	2 893.73	480.53
1997	281.69	149.44	25 190.97	1 023.35

3.6. táblázat. Comauto 2003 kódú vállalatának eredményei Gamma eloszlással

Ha már nem csak egy vállalatra hasonlítjuk össze a két eloszlást, hanem a vállalatok egészét nézzük, akkor átfogóbb képet kaphatunk a két különböző eloszlással végzett előrejelzés viselkedéséről. A 3.1 ábrán a két használt feltételezéssel kapott átlagos négyzetes eltérés értékek boxplotjai jelennek meg, ahol az összehasonlítás a teljes végső kárra, tehát a háromszög utolsó oszlopának összegére vonatkozik. Ez alapján ha nézzük a becslések átlagos négyzetes hibáját minden vállalatra, akkor ezek mediánja közel van a 0-hoz, de vannak ennél jóval magasabb értékek is, melyek már outlieriek. A kilógó pontok a jobb olvashatóság miatt nincsenek rajta az ábrán, mert ezek annyira összenyomnák az y -tengelyt, hogy nem lehetne semmiféle különbséget látni a két különböző feltételezésből kiinduló becslési eljárás hibája között. Így ábrázolva az látszik, hogy a Gamma eloszlás esetén kicsit magasabb a „doboz”, ami azt jelenti, hogy a becslés és ténylegesen megfigyelt „valódi” értékek közötti átlagos eltérések középső 50%-a nagyobb terjedelmű mint az overdispersed Poisson esetében.

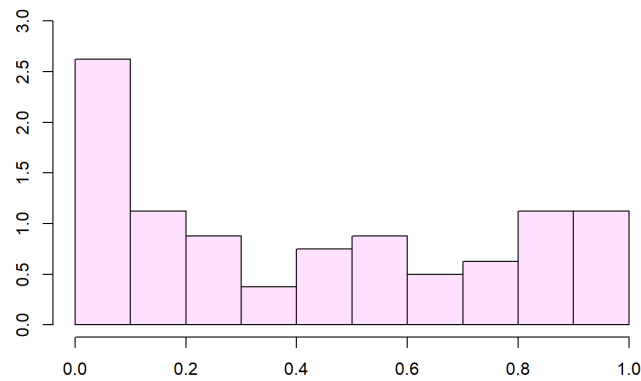
Átlagos négyzetes eltérés értékek az eredeti comauto adatsorra



3.1. ábra. Átlagos négyzetes eltérés overdispersed Poisson és Gamma eloszlásból

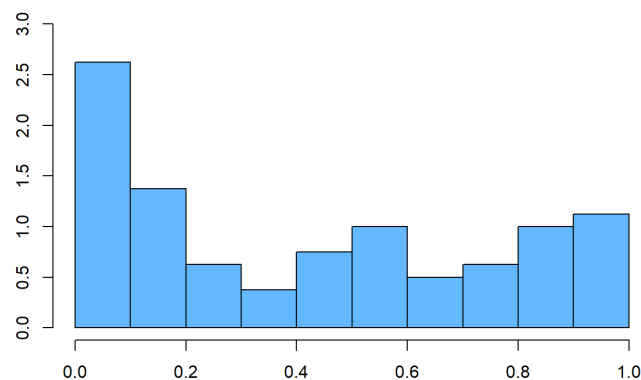
További összehasonlítási lehetőség a Probability Integral Transformation és az ahhoz tartozó hisztogram ábrája. A 3.2 és 3.3 ábrák úgy készültek, hogy a becsült végső összkárok (amik a 10 különböző kárbekövetkezési év végső kárainak összege) által meghatározott eloszlásba, behelyettesítésre kerültek „valódi” végső összkár értékek, mely minden vállalatra egy 0 és 1 közötti számot adott vissza. Ha az előrejelzés tökéletesen működne, a hisztogramok a $[0, 1]$ -en egyenletes eloszlás hisztogramjához lennének hasonlatosak. Itt ez sajnos nem mondható el teljes mértékben, de az egyenletestől való eltérés is hordoz hasznos információt. A két feltételezett eloszlás PIT ábrája nagyon hasonló, mindkét esetben egy U alakra emlékeztet, mely arra utal, hogy a becslés alulszóródott, túlságosan szűk az előrejelzés intervalluma.

Az adatok egésze Overdispersed Poisson eloszlással 1000 iterációval



3.2. ábra. PIT overdispersed Poisson eloszlásból

Az adatok egésze Gamma eloszlással 1000 iterációval



3.3. ábra. PIT Gamma eloszlásból

3.2. Homogén csoportok normalizálva

A felhasznált adatsorokban a kárkifizések mellett az éves díjak is szerepelnek, így lehetőség nyílik az adatok egyfajta normalizálására [8], mely jelen esetben azt jelenti, hogy a kárkifizéseket leosztjuk az adott éves megszolgált

díjjal és az így képzett háromszögeket vizsgáljuk. Az adatsorban előfordultak olyan vállalatok, amiknek valamelyik évben 0 volt az éves befolyt díjuk, ezek a 0-val való osztás miatt kimaradtak a vizsgálatból. Ez a normalizálással való adatátalakítás azért lehet hasznos, mert több vállalatot együtt vizsgálva a nagyobb cégek kárai dominálják az adatsort, ezzel szinte láthatatlanná téve a kisebb cégeket. Ugyanakkor a nagyobb cégek természetes módon nagyobb kármennyiséget fognak tapasztalni, ezért jogosak a normalizálás nélküli vizsgálatok is. Az eredeti terv tehát az volt, hogy az összes korábban elvégzett előrejelzést újra elvégezzük ezeken az „új” adatsorokon kivéve a Mack módszert, mert az nem ad lehetőséget a vállalatok együttes vizsgálatára, abban csupán egy átskálázás történik. Az eddigi módszerek alkalmazása esetén az a probléma merült fel, hogy az overdispersed Poisson eloszlás nagyon érzékeny a felhasznált adatok nagyságrendjére, a „kis” számokat nem tudja kezelni, ezért a normalizálás úgy lett módosítva, hogy a díjakkal való leosztás után először minden háromszög (a benne lévő összes adat) meg lett szorozva 1 000-rel majd 1 000 000-val.

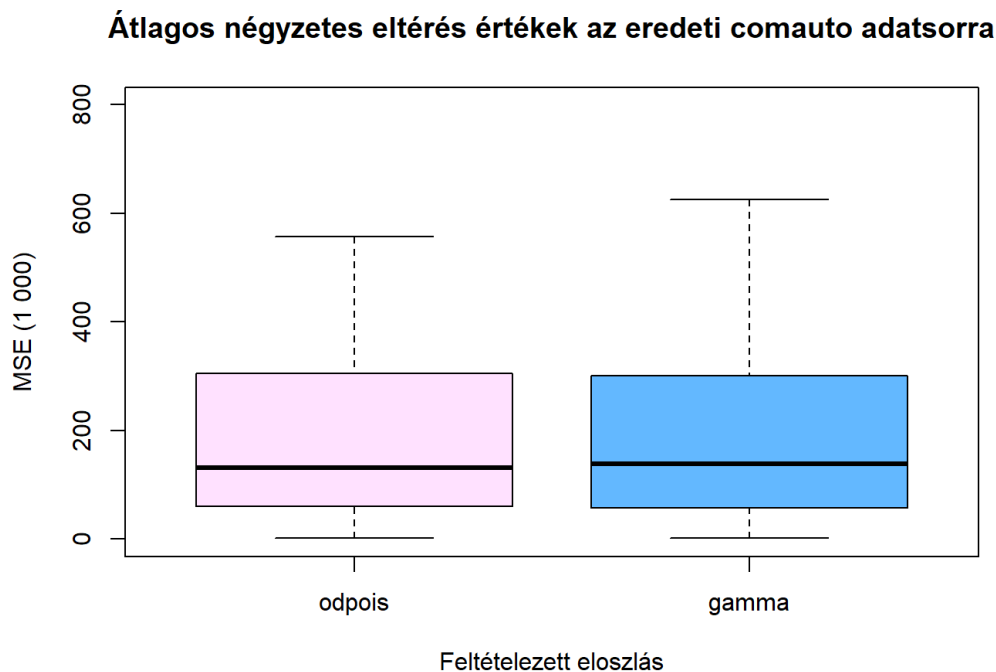
3.2.1. Bootstrapet használó módszerek

A bootstrapet használó módszereknél most is két különböző eloszlásfeltevéssel és 1 000 iterációval fogunk dolgozni, az eloszlások: overdispersed Poisson és Gamma. Ebben a fejezetben két további alfejezet szerepel amiatt, amiről pár sorral feljebb szó volt, hogy az adatok normalizálás után nem feltétlenül voltak használhatóak, ezért „fel lettek nagyítva”, eszerint vannak rendezve az alfejezetek.

Adatsorok normalizálva és 1 000-rel „felnagyítva”

Az 1 000-rel „felnagyított” normalizált adatokon az előrejelzéseket elvégezve a teljes végső kárra vonatkozó átlagos négyzetes hibák ábrázolása most is boxplot diagrammal történt (3.4 ábra), mely alakjában kicsit eltér az eredeti adatokon kapott diagramtól (3.1 ábra), ugyanis itt kisebb intervallumon mozognak az átlagos négyzetes hibák és csak minimális különbség látszik a

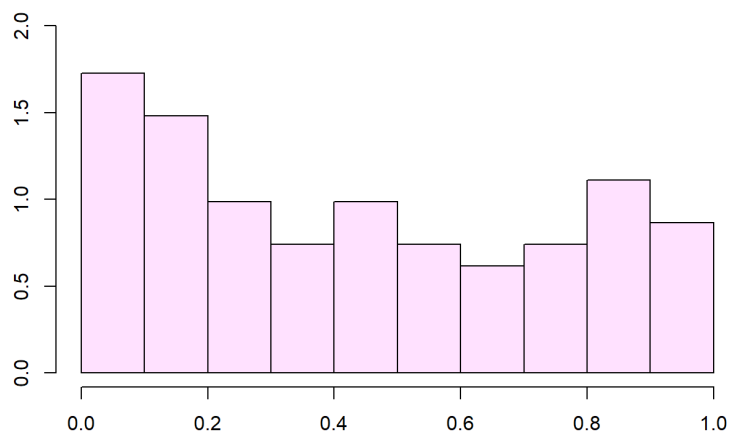
két feltételezett eloszlással kapott eredmények között. Ami még eltér a 3.1 ábrától, hogy látványosabb a medián elválása a nullától.



3.4. ábra. Átlagos négyzetes eltérés értékek a normalizált és 1 000-rel „fel-nagyított” adatokra

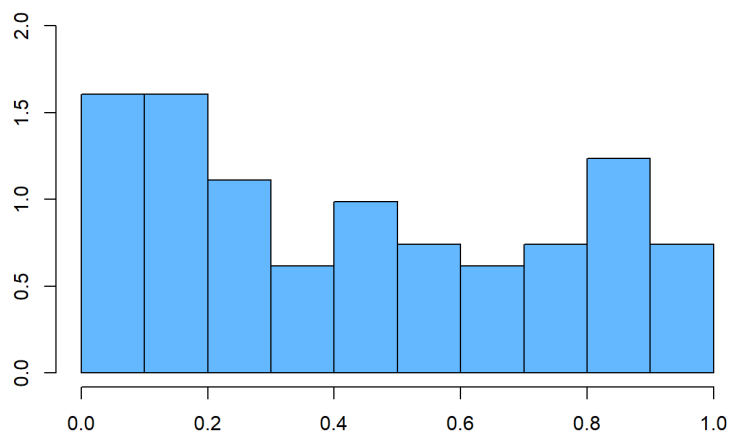
Az adatok egészén végzett előrejelzések által meghatározott eloszlásba a „valódi” értékeket behelyettesítve a két hisztogram a 3.5 és 3.6 ábrákon látható. Az alak itt sem felel meg teljesen a $[0, 1]$ -en egyenletes eloszlásnak, de talán közelebb van hozzá mint a 3.2 és 3.3 ábrákon. A torzulásokra nem lehet egyértelműen rámondani, hogy U alakot vesz fel a hisztogram, de inkább az U alakhoz van közelebb, mint a középen domborúhoz.

Az adatok egésze Overdispersed Poisson eloszlással 1000 iterációval



3.5. ábra. PIT overdispersed Poisson eloszlásból (ezres „nagyítással”)

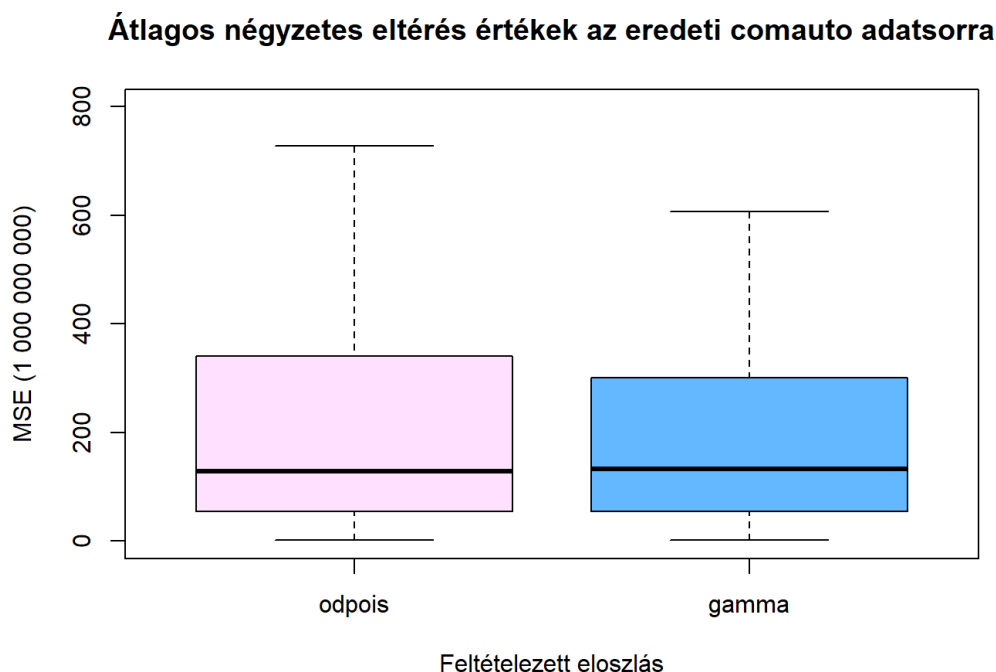
Az adatok egésze Gamma eloszlással 1000 iterációval



3.6. ábra. PIT Gamma eloszlásból (ezres „nagyítással”)

Adatsorok normalizálva és 1 000 000-val „felnagyítva”

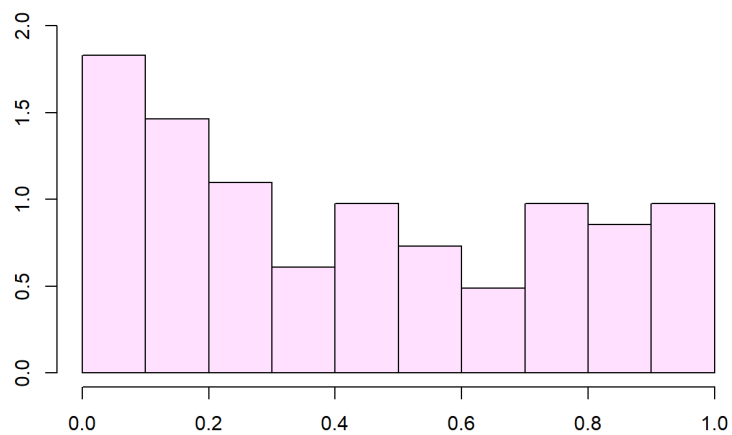
A normalizálás, majd 1 000 000-val történő felnagyítása után várható módon az átlagos négyzetes hibák nagyságrendekkel nagyobb értékeket vettek fel, ez látszik a 3.7 ábrán. A két feltételezett eloszlás boxplot diagramja között itt sincs hatalmas eltérés, ami más az eredeti adatsoron (3.1 ábrán) kapott eredményekhez képest, hogy itt az overdispersed Poisson feltételezett eloszláshoz tartozó „doboz” a magasabb.



3.7. ábra. Átlagos négyzetes eltérés értékek a normalizált és 1 000 000-val „felnagyított” adatokra

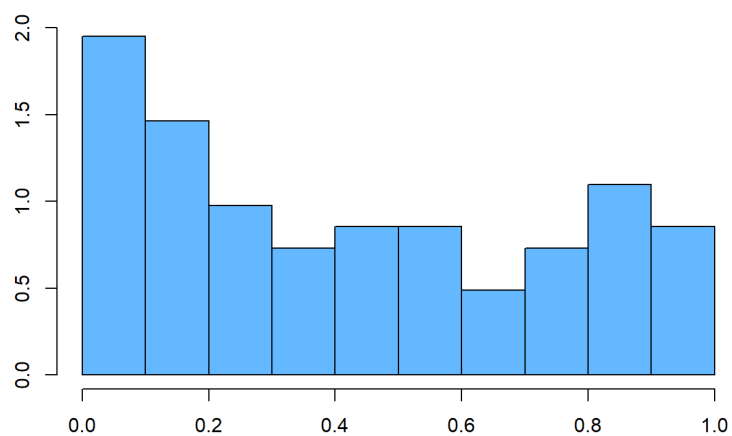
A PIT ábrák (3.8 és 3.9) sem hoztak látványos eltérést az 1 000-rel „nagyított” adatokon kapott (3.5 és 3.6) ábrákhoz képest. Az eloszlás inkább a 0 felé koncentrálódik, míg közepén egyfajta bemélyedés van.

Az adatok egésze Overdispersed Poisson eloszlással 1000 iterációval



3.8. ábra. PIT overdispersed Poisson eloszlásból (millió „nagyítással”)

Az adatok egésze Gamma eloszlással 1000 iterációval



3.9. ábra. PIT Gamma eloszlásból (millió „nagyítással”)

3.3. Egyéni káradatok

Egy másik előrejelzésre használt adathalmaz az egyéni káreseményeket tartalmazó magyar adatsor, melyben összesen 5 oszlop található [19]. Az első oszlop egy azonosítószám, a második oszlopban szerepelnek a kárbekövetkezési időszakok 1-től 6-ig. A harmadik és negyedik oszlop a bejelentési és kifizetési időszakot tartalmazza szintén 1-től 6-ig, az utolsó oszlopban pedig a kárkifizetések vannak. Ezen adatok alapján két különböző háromszöget lehetett készíteni azzal a megkülönböztetéssel, hogy a kárbekövetkezéshez képest a kárkifizetésig vagy a kárbejelentésig eltelt időt vesszük figyelembe. Mivel a teljes „négyzög” adatai ismertek, itt is lehetőség nyílik a becült és „valódi” kárkifizetések összehasonlítására. A két kumulált háromszög értékei kezdetben eltérnek, de az utolsó oszlopra összeérnek, tehát kifutásuk megegyezik, mert a biztosító előbb-utóbb minden ügyet lezár. Erre a két háromszögre elvégezve a Mack-módszert ugyanazt az eredményt kapjuk.

3.3.1. Mack módszer

	Végső kár ($\hat{C}_{i,I}$)	Eltérés ($\hat{C}_{i,I} - C_{i,I}$)	Szórás ($\hat{\sigma}_j$)
1	2 276 089 908	0	0
2	1 609 796 910	-3 961 140	2 220 577
3	1 780 737 067	-10 534 397	3 815 345
4	2 077 959 280	-7 463 360	6 050 534
5	3 769 100 871	91 306 555	27 942 888
6	1 839 149 499	-339 739 594	240 670 467

3.7. táblázat. Mack módszer eredményei az egyéni adatokra

3.3.2. Bootstrapet használó módszerek

A bootstrapet használó módszerek az eddigieknek megfelelően overdispersed Poisson és Gamma eloszlással külön-külön lettek elvégezve a kifizetésig eltelt

idő szerint rendezett és a bejelentésig eltelt idő szerint rendezett háromszögekre, mindkét esetben 1 000 iterációval.

Háromszög a kifizetésig eltelt idő szerint rendezve

Először a kifizetésig eltelt idő szerint rendezett háromszög kumulált adataiból kiindulva készültek előrejelzések, ezek eredményeit foglalják össze a 3.8 és 3.9 táblázatok. Itt is elmondható, hogy a két feltételezett eloszlással kapott előrejelzésekben nincs nagy eltérés, így az átlagos négyzetes hibák nagyságában sem.

Overdispersed Poisson

	Becsült végső károk átlaga	Becsült végső károk szórása	Átlagos négyzetes eltérés	Kvantilis (99.5%)
1	2 276 089 908	0	0	2 276 089 908
2	1 610 190 712	15 170 229	242 631 629 334 881	1 689 237 996
3	1 781 812 785	37 628 340	1 503 942 693 746 499	1 990 021 883
4	2 081 317 127	56 639 248	3 221 651 609 180 260	2 283 418 853
5	3 776 780 556	135 021 945	28 010 970 510 393 712	4 209 842 582
6	1 855 780 685	169 218 458	133 005 295 089 577 248	2 362 771 890

3.8. táblázat. Eredmény a kifizetésig eltelt idő szerint rendezett háromszögre overdispersed Poisson eloszlással és 1 000 iterációval

Gamma

	Becsült végső károk átlaga	Becsült végső károk szórása	Átlagos négyzetes eltérés	Kvantilis (99.5%)
1	2 276 089 908	0	0	2 276 089 908
2	1 609 982 643	16 199 653	276 420 025 933 612	1 689 494 019
3	1 780 454 464	32 968 309	1 202 829 961 231 364	1 915 764 339
4	2 079 605 618	58 076 495	3 403 344 104 212 557	2 310 870 735
5	3 784 093 114	140 667 011	31 066 855 267 986 552	4 252 806 442
6	1 848 933 102	170 657 685	137 965 877 544 827 744	2 397 205 851

3.9. táblázat. Eredmény a kifizetésig eltelt idő szerint rendezett háromszögre Gamma eloszlással és 1 000 iterációval

Háromszög a bejelentésig eltelt idő szerint rendezve

Másodszor a bejelentésig eltelt idő szerint rendezett háromszög kumulált adataiból kiindulva készültek előrejelzések, melyek kárévenként megbontva a 3.10 és 3.11 táblázatokban látszik a további eredményekkel együtt.

Overdispersed Poisson

	Becsült végső károk átlaga	Becsült végső károk szórása	Átlagos négyzetes eltérés	Kvantilis (99.5%)
1	2 276 089 908	0	0	2 276 089 908
2	1 610 153 120	13 220 742	187 608 758 047 348	1 687 178 968
3	1 779 862 398	34 193 769	1 298 211 412 895 702	1 922 717 976
4	2 082 081 994	60 576 694	3 677 026 293 240 053	2 289 008 809
5	3 774 072 544	132 576 648	26 828 488 271 474 352	4 205 994 832
6	1 840 528 009	171 692 282	143 936 984 899 148 704	2 342 039 897

3.10. táblázat. Eredmény a bejelentésig eltelt idő szerint rendezett háromszögre overdispersed Poisson eloszlással és 1 000 iterációval

Gamma

	Becsült végső károk átlaga	Becsült végső károk szórása	Átlagos négyzetes eltérés	Kvantilis (99.5%)
1	2 276 089 908	0	0	2 276 089 908
2	1 610 672 392	18 198 349	340 370 024 590 088	1 713 925 020
3	1 780 018 797	36 390 865	1 449 593 262 051 599	1 963 850 398
4	2 078 473 046	57 297 628	3 328 032 063 021 562	2 300 680 615
5	3 773 948 217	137 264 353	28 068 233 815 596 516	4 250 437 735
6	1 852 712 350	167 390 400	134 382 794 187 954 016	2 364 100 092

3.11. táblázat. Eredmény a bejelentésig eltelt idő szerint rendezett háromszögre Gamma eloszlással és 1 000 iterációval

4. fejezet

Konklúzió

A dolgozat fejezeteiben először röviden áttekintettük különböző tartalékolási módszerek irodalomban való előfordulását, majd bemutatásra került a Mack módszer és a bootstrap algoritmus, ezek valós adatokon való alkalmazása és az általuk becsült végső kárelőrejelzések minőségének elemzése. A felhasznált adatsorok közül az egyik 6 homogén csoportból állt, ahol a homogén csoportokat vállalatok bekövetkezési és kifutási évenként csoportosított kárösszegei alkották. A másik, előzőtől függetlenül felhasznált magyar adatsor egyéni káradatokat tartalmazott. Mindkét adatsor kiváló tulajdonsága, hogy tartalmaz annyi adatot, hogy annak egy része alapján előrejelzést lehessen készíteni, majd a becsléseket össze lehessen hasonlítani az adatok másik részével, ezzel ellenőrizve, hogy mennyire feleltek meg a valóságnak az általunk kiszámolt kárkifizetések. Mind az amerikai, mind a magyar adatsorokra elvégeztük az előrejelzéseket Mack módszerrel, majd sztochasztikusan overdispersed Poisson és Gamma eloszlást feltételezve a bootstrap eljárást alkalmazva. Ezután már csak a Commercial auto/truck liability/medical homogén csoportra és a sztochasztikus előrejelzésekre megnéztük változnak-e az eredmények, ha a kárkifizetések normalizálva vannak az adott kár bekövetkezési év díjaival. Azért lehet érdemes normalizált adatokból kiindulva is modellezni, mert ellenkező esetben az adatokat dominálják a nagyobb cégek kárösszegei és a kisebb vállalatok láthatatlanná válnak. A normalizálás el-

len szól ugyanakkor, hogy a nagyobb cégek jogosan szembesülnek nagyobb kárösszegekkel és összpontosul náluk a károk nagy része. Ebből is két különböző esetet vizsgáltunk, ugyanis az overdispersed Poisson eloszlás diszkrét jellegéből adódóan érzékeny az adatok nagyságrendjére, ezért az egyik esetben a normalizálás után 1 000-rel, míg a másik esetben 1 000 000-val „nagyítottuk fel” az adatokat. Az eredményeket a Mack módszer esetén csupán a becült és „valódi” végső károk közti különbség, valamint a szórás ($\hat{\sigma}$) alapján vizsgáltuk. A sztochasztikus esetben megnéztük az egy-egy vállalatra 1 000 iterációval kapott végső károk átlagát, szórását és 99.5%-os kvantilisét, valamint a becült végső károk „valódi” végső károktól vett átlagos négyzetes eltérését kárévenkénti bontásban mind az overdispersed Poisson, mind a Gamma eloszlást feltételezve. Az átlagos négyzetes eltéréseket csoport szinten boxplot diagrammal vizsgáltuk, ahol csak a normalizálás előtti adatokban látszott némi eltérés a két eloszlás eredményei között, a normalizálás után, ez már szinte észrevehetetlenné vált. A sztochasztikus modelleknél, ahol empirikus eloszlásfüggvényt tudunk készíteni a végső kárra, ott az eloszlások ellenőrzésére a PIT ábrákat használtuk, amihez a becslések által meghatározott eloszlásba helyettesítettük vissza a megfigyelt adatokat és az így kapott eloszlás hisztogramjának hasonlóságát vizsgáltuk a $[0, 1]$ -en egyenletes eloszlás hisztogramjához. Akár normalizálás előtt vagy után, akár overdispersed Poisson vagy Gamma eloszlással jeleztünk előre, a hisztogram mindig a 0-felé tolódott, a két szélén magasabb, míg közepén inkább alacsonyabb értékeket vett föl. Ez arra utal, hogy a becslési eredmény alulszóródott, túlságosan szűk az előrejelzés intervalluma. Az adatok normalizálásával annyi változás történt, hogy a hisztogramok képe kicsivel közelebb került az egyenletes eloszlás hisztogramjának képéhez. Tehát a sztochasztikus modellekkel becsléseket végezve megállapíthatjuk, hogy biztosítási szerződések egy csoportjának mennyire bizonytalan a végső károkra való előrejelzése. Azonban ezek a modell feltételezések maradéktalan teljesülése esetén lennének csupán igazak, így ezekkel óvatosan kell bánnunk, hiszen megbízhatóságuknak megvannak a korlátai, ahogyan azt a PIT hisztogramoknál is láttuk. A nor-

malizálást és csoport szintű összehasonlítást már nem végeztük el az egyéni káradatakra, mert ott csupán 2 háromszögből dolgoztunk, ami nem teszi lehetővé az olyan összehasonlításokat, amiket a homogén csoportoknál leírtunk.

Irodalomjegyzék

- [1] M. Arató, „Nem-élet biztosítási matematika,” *ELTE Eötvös Kiadó*, 2001.
- [2] T. Mack, „Distribution-free calculation of the standard error of chain ladder reserve estimates,” *ASTIN Bulletin: The Journal of the IAA*, vol. 23, no. 2, pp. 213–225, 1993.
- [3] P. England and R. Verrall, „Analytic and bootstrap estimates of prediction errors in claims reserving,” *Insurance: mathematics and economics*, vol. 25, no. 3, pp. 281–293, 1999.
- [4] P. D. England and R. J. Verrall, „Stochastic claims reserving in general insurance,” *British Actuarial Journal*, vol. 8, no. 3, pp. 443–518, 2002.
- [5] R. Kaas, M. Goovaerts, J. Dhaene, and M. Denuit, *Modern actuarial risk theory: using R*, vol. 128. Springer Science & Business Media, 2008.
- [6] R. Scarth, S. Jain, and R. Roberto, „A practitioner’s introduction to stochastic reserving,” 2020.
- [7] K. Peremans, P. Segaert, S. Van Aelst, and T. Verdonck, „Robust bootstrap procedures for the chain-ladder method,” *Scandinavian Actuarial Journal*, vol. 2017, no. 10, pp. 870–897, 2017.
- [8] P. Shi, „A multivariate analysis of intercompany loss triangles,” *Journal of Risk and Insurance*, vol. 84, no. 2, pp. 717–737, 2017.

- [9] N. Chukhrova and A. Johannssen, „Kalman filter learning algorithms and state space representations for stochastic claims reserving,” *Risks*, vol. 9, no. 6, p. 112, 2021.
- [10] R. Hendrych and T. Cipra, „Applying state space models to stochastic claims reserving,” *ASTIN Bulletin: The Journal of the IAA*, vol. 51, no. 1, pp. 267–301, 2021.
- [11] V. Elpidorou, C. Margraf, M. D. Martínez-Miranda, and B. Nielsen, „A likelihood approach to bornhuetter–ferguson analysis,” *Risks*, vol. 7, no. 4, p. 119, 2019.
- [12] M. Rathouský, „Micro-level stochastic claims reserving,” 2019.
- [13] K. Antonio and R. Plat, „Micro-level stochastic loss reserving for general insurance,” *Scandinavian Actuarial Journal*, vol. 2014, no. 7, pp. 649–669, 2014.
- [14] M. Maciak, O. Okhrin, and M. Pešta, „Infinitely stochastic micro reserving,” *Insurance: Mathematics and Economics*, vol. 100, pp. 30–58, 2021.
- [15] L. Martinek, „Experience rating and stochastic reserving in general insurance,” 2019.
- [16] M. R. Shapland, „Using the odp bootstrap model: A practitioner’s guide,” Casualty Actuarial Society Arlington, 2016.
- [17] L. Martinek, „Analysis of stochastic reserving models by means of naic claims data,” *Risks*, vol. 7, no. 2, p. 62, 2019.
- [18] G. G. Meyers and P. Shi, „Loss reserving data pulled from naic schedule p,” *URL: <https://www.casact.org/publications-research/research/research-resources/loss-reserving-data-pulled-naic-schedule-p>*, vol. 5, 2011. Letöltés időpontja: 2021.12.13.
- [19] <http://amiklos.web.elte.hu/stochreserve/> Letöltés időpontja 2022.01.26.

Ábrák jegyzéke

3.1. Átlagos négyzetes eltérés overdispersed Poisson és Gamma eloszlásból	32
3.2. PIT overdispersed Poisson eloszlásból	33
3.3. PIT Gamma eloszlásból	33
3.4. Átlagos négyzetes eltérés értékek a normalizált és 1 000-rel „felnagyított” adatokra	35
3.5. PIT overdispersed Poisson eloszlásból (ezres „nagyítással”) . .	36
3.6. PIT Gamma eloszlásból (ezres „nagyítással”)	36
3.7. Átlagos négyzetes eltérés értékek a normalizált és 1 000 000-val „felnagyított” adatokra	37
3.8. PIT overdispersed Poisson eloszlásból (milliós „nagyítással”) .	38
3.9. PIT Gamma eloszlásból (milliós „nagyítással”)	38

Táblázatok jegyzéke

1.1. Kifutási háromszög	6
1.2. Kumulált kifutási háromszög	7
1.3. Kumulált jövőbeli kifutási háromszög	7
1.4. Példa kárkifizetési háromszögre	8
1.5. Példa kumulált kárkifizetési háromszögre	9
2.1. Példa a lánc-létra módszerre	14
2.2. Bootstrap példa kiinduló kumulált háromszöge	17
2.3. Növekedési faktorok	17
2.4. Visszabecsült kumulált értékek	18
2.5. Visszabecsült kumulálatlan értékek	18
2.6. Pearson reziduális értékek	18
2.7. Igazított Pearson reziduális értékek	19
2.8. Reziduálisok háromszöge a mintavételből	20
2.9. Növekmények pszeudo-háromszöge	20
2.10. Kumulált pszeudo-háromszög	20
2.11. Kumulált pszeudo-háromszög növekedési faktorai	21
2.12. Lánc-létra eredményei	21
2.13. Lánc-létra növekményes eredményei	21
2.14. Szimulált overdispersed Poisson értékek	22
2.15. Egy forgatókönyv eredményei	22
2.16. Átlagos négyzetes eltérés 2 forgatókönyv „valódi” végső kárra adott becslése alapján	23

3.1. Felhasznált vállalatok száma homogén csoportonként	26
3.2. Comauto 2003 kódú vállalatának kiinduló háromszöge	27
3.3. Comauto 2003 kódú vállalatának előrejelzése Mack módszerrel	27
3.4. Comauto 2003 kódú vállalatának eredményei Mack módszerrel	28
3.5. Comauto 2003 kódú vállalatának eredményei overdispersed Poisson eloszlással	30
3.6. Comauto 2003 kódú vállalatának eredményei Gamma eloszlással	31
3.7. Mack módszer eredményei az egyéni adatokra	39
3.8. Eredmény a kifizetésig eltelt idő szerint rendezett háromszögre overdispersed Poisson eloszlással és 1 000 iterációval	40
3.9. Eredmény a kifizetésig eltelt idő szerint rendezett háromszögre Gamma eloszlással és 1 000 iterációval	41
3.10. Eredmény a bejelentésig eltelt idő szerint rendezett háromszögre overdispersed Poisson eloszlással és 1 000 iterációval	41
3.11. Eredmény a bejelentésig eltelt idő szerint rendezett háromszögre Gamma eloszlással és 1 000 iterációval	42