Stohastički okvir metode ulančanih ljestvica

Topić, Tončica

SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO-MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Tončica Topić

Stohastički okvir metode ulančanih ljestvica

Završni rad

Zagreb, 2022.

SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO-MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Poslijediplomski specijalistički studij aktuarske matematike

Tončica Topić

Stohastički okvir metode ulančanih ljestvica

Završni rad

Voditelj završnog rada:
Izv. prof. dr. sc. Nikola Sandrić

Zagreb, 2022.

Ovaj završni rad obranjen je dana ____________________ pred ispitnim povjerenstvom u sastavu:

1. ______________________________ , predsjednik

2. ______________________________ , član

3. ______________________________ , član

Povjerenstvo je rad ocijenilo ocjenom ______________________ .

Potpisi članova povjerenstva:

1. ______________________________

2. ______________________________

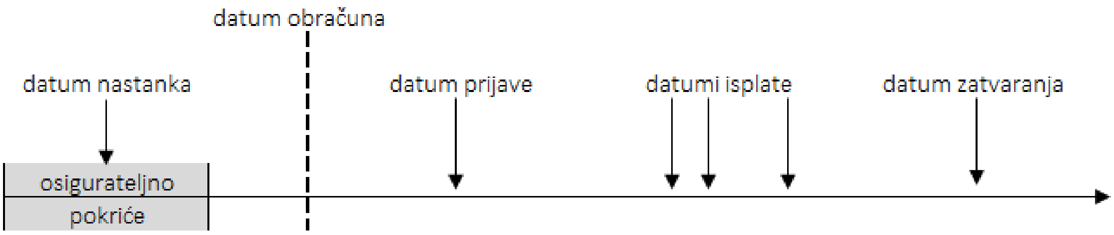
3. ______________________________

Sadržaj

1 Uvod

Stohastičko modeliranje pričuve šteta neživotnih osiguranja nepresušna je aktuarska tema kojoj se posvećuje sve više pažnje. Pod njezinim okriljem nastao je i ovaj rad, s ciljem razumijevanja osnovih postavki određenih stohastičkih modela i mogućnostima primjene. Kako bi dali uvod, sada ćemo svaki od tri segmenta koji su sastavni dio teme oživjeti detaljnijim opisima. Krenimo od kraja.

Neživotna osiguranja. U osigurateljnom žargonu, osiguranja se dijele na životna i neživotna. Neživotna osiguranja se najkraće definiraju kao sva osiguranja koja nisu životna. Dalo bi se tu na dugo i široko nabrajati razlike pa da ne duljimo navest ćemo nekoliko vrsta neživotnih osiguranja: osiguranje od nezgode, zdravstveno osiguranje, osiguranje motornih vozila, osiguranja imovine, osiguranja od odgovornosti, osiguranje brodica, osiguranje kredita itd. Polica neživotnog osiguranja je ugovor između ugovaratelja osiguranja i osiguratelja. Po tom ugovoru osiguratelj je dužan isplatiti štetu u slučaju da nastane osigurani slučaj dok se ugovaratelj obvezuje platiti premiju osiguranja. Polica u pravilu traje godinu dana i u konačnici može i ne mora imati štetu. Kako bi u priču uveli drugi segment, pričuvu šteta, pogledajmo sliku 1. Na toj vremenskoj skali prikazan je život jedne police neživotnog osiguranja sa pripadnom štetom. Glavnu ulogu ovdje imaju datumske komponente: šteta je nastala kroz period pokrića (inače ne bi bio osigurani slučaj), prijavljena je nakon što je polica istekla, a isplaćena u nekoliko navrata dosta nakon datuma prijave. Osiguratelj ima zakonsku obvezu štetu priznati (kroz financijske izvještaje) u trenutku njezinog nastanka - dužan je obračunati pričuvu šteta.



Slika 1: Vremenska crta života jedne police sa štetom

Pričuva šteta. Ukupna obveza potrebna za pokriće svih nastalih neisplaćenih šteta zove se pričuva šteta. Nekoliko je razloga zbog čega osiguratelj, u trenutku obračuna, nije u mogućnosti precizno znati kolike su njegove buduće obveze po štetama koje su nastale do tog datuma:

- Nerijetko se događa vremenski odmak između nastanka i prijave štete (kako je i prikazano na slici 1)

- Nakon prijave štete može doći do promjene iznosa potrebnog za isplatu.
- Događa se i preotvaranje već zatvorenih šteta.

S druge strane, rekli bi da je očito zbog čega pričuva šteta treba biti što preciznije izračunata: u slučaju da je podcijenjena osiguratelj neće biti u mogućnosti isplatiti nastale štete, dok precijenjena pričuva znači da se nepotrebno drže sredstva koja mogu biti iskorištena u druge svrhe. Dodatno, obzirom da pričuva šteta čini veliki dio budućih obveza društva, male promjene u izračunima znače velike apsolutne novčane utjecaje.

Kako izračunati pričuvu šteta?

Stohastičke metode. Dugo vremena oblikovanje pričuve šteta shvaćano je kao deterministički proces opisan jednostavnim algoritmima. Tek u drugoj polovici 20. stoljeća počelo se osvještavati pitanje koliko su precizni dobiveni rezultati. Na takva pitanja možemo dati odgovor jedino ako se smjestimo u stohastički okvir. Klasični statistički proces bi bio prvo definirati model, zatim procijeniti pričuvu i konačno kvantificirati pouzdanost dobivenih rezultata. U aktuarskom okruženju možemo reći da je smjer obratan - najčešće se algortimi i rezultati pokušavaju opisati modelima kako bi se procijenila njihova varijabilnost. Tim putem ćemo i mi ići.

Metoda ulančanih ljestvica vjerojatno je najpopularnija metoda rezerviranja u praksi. Ona u svojim počecima primjene nije bila ništa više od algoritma. Kako je jednostavna za implementaciju i često daje prilično precizne rezultate, počela se opisivati različitim modelima u potrazi za pripadnom varijabilnošću. U ovom radu obraditi ćemo dva stohastička modela koja reproduciraju rezultate algoritma metode ulančanih ljestvica:

- Model ulančanih ljestvica bez pretpostavljene distribucije
- Raspršeni Poissonov model

Oba modela okrunit ćemo bootstrap metodom.

Rad je koncipiran u dva dijela: teorijski i primijenjeni. Teorijski dio otvaramo opisom samog algoritma metode ulančanih ljestvica kojeg potom postepeno stavljamo u stohastički okvir. Na taj način izgradit ćemo prvi model koji, kao što samo ime kaže, ne počiva na pretpostavci o distribuciji. Zatim mu suprostavljamo model koji od početka sadrži pretpostavku o raspršenoj Poissonovoj distribuciji, ali daje identične rezultate kao i prvi, u smislu očekivanih šteta. On počiva na puno široj matematičkoj teoriji generaliziranih linearnih modela, međutim mi se ograničavamo samo na Poissonovu distribuciju iz familije eksponencijalnih. Iako oba modela produciraju pričuvu šteta jednaku algoritmu metode ulančanih ljestvica, prilično su različiti u stohastičkom smislu. To će se očitovati kroz druge i više momente. Za kraj teorijskog dijela obradit ćemo bootstrap metodu. U drugom dijelu predstavljamo podatke na kojima potom primijenjujemo prethodno opisane modele, uz testiranje pripadnih pretpostavki i analizu dobivenih rezultata.

2 Podaci i pripadne oznake

Prije nego krenemo obrađivati modele bitno je shvatiti formu podataka na koju se primjenjuju. Cilj ovog odjeljka je dati pregled osnovnih i najčešćih aktuarskih koraka pri grupiranju podataka a sve u svrhu procjene budućih obveza po štetama koje su već nastale kako je opisano u uvodu. Također, navest ćemo notaciju kao i neke opće pretpostavke koje čine kostur rada.

Osiguravajuća društva obično raspolažu sa prilično opsežnim povijesnim podacima o štetama. Podaci najčešće sadrže informacije o osiguraniku, oštećenom, uzroku štete, vrsti rizika, iznosu štete, zatim datumske komponente koje štetu stavljaju u vremenski okvir te dodatne specifičnosti koje pružaju detaljniji uvid (je li šteta sudska, je li reosigurana, postoji li regres po šteti itd). Aktuari redovito analiziraju povijesne podatke te ih koriste kao temelj pri modeliranju očekivanih budućih kretanja.

Grupiranje podataka najčešće započinje formiranjem homogenih skupina rizika kao što su osiguranje od nezgode, osiguranje motornih vozila, osiguranje od požara, potresa itd. Ponekad postoji potreba da se iz dobivenih skupina dodatno odvajaju ekstremne, sudske ili neke druge štete za koje se vjeruje da odstupaju ili će odstupati u budućnosti. Takvo profinjavanje segmentacije ima smisla dok god je promatrana grupa statistički stabilna sa dovoljnim brojem podataka. To pak ukazuje na iterativnu prirodu procesa grupiranja podataka.

Nakon segmentacije portfelja slijedi slaganje podataka unutar svakog segmenta u tzv. razvojne trokute. Trokuti mogu sadržavati podatke o iznosu ili broju šteta (i razne druge pokazatelje dobivene iz njih). Mi ćemo se od sada do kraja ograničiti na trokute iznosa šteta iako sve vrijedi i za trokute broja šteta. Neka $X_{i,j}$ označava iznos šteta plaćenih u trenutku i, j pri čemu je

- i = godina nastanka štete; $i \in \{0, ...I\}$
- j = godina isplate od trenutka nastanka i; $j \in \{0, ..., J\}$.

Interpretacija je sljedeća: isplata se dogodila u računovodstvenoj godini $i + j$. Kumulativne isplate u godini nastanka i i razvojnom periodu j označavamo s $C_{i,j}$ i za njih vrijedi

$$C_{i,j} = \sum_{k=0}^{j} X_{i,k}$$

U aktuarskom svijetu opisani podaci čine razvojne trokute prikazane niže u Tablici 1. Klasična interpretacija takve forme je sljedeća: u trenutku I tablica je podjeljena na tzv gornji trokut koji se sastoji od opaženih podataka do tog trenutka odnosno $\{X_{i,j} : i + j \leq I\}$ i donjeg trokuta budućih događaja $\{X_{i,j} : i + j > I\}$ koje pokušavamo procjeniti raznim metodama uz pomoć gornjeg trokuta. Možemo

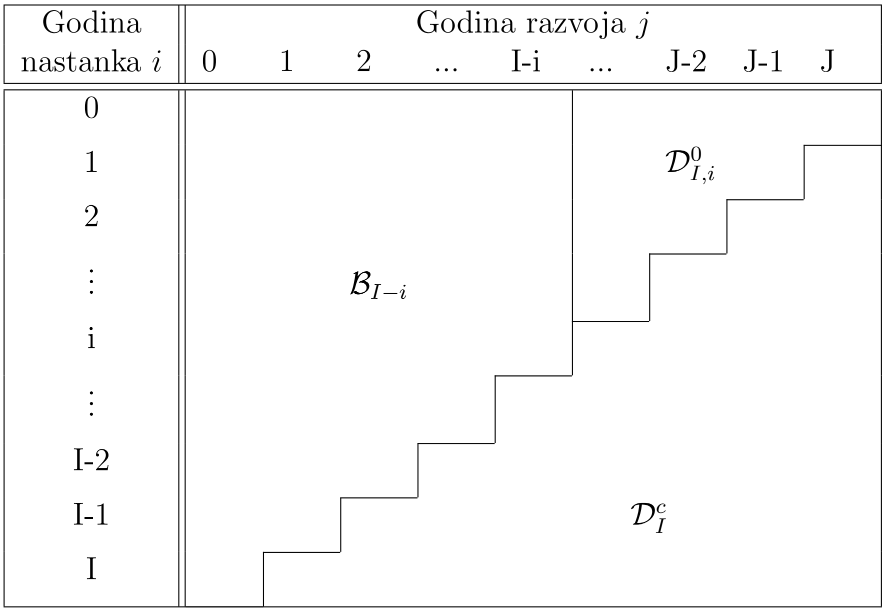
Godina nastanka i	Godina razvoja j								
	0	1	2	3	...	j	...	J-1	J
0									
1									
2	gornji trokut opaženih $X_{i,j}$								
$\vdots$	$i+j \leq I$								
i									
$\vdots$									
I-2									
I-1				donji trokut procijenjenih $X_{i,j}$					
I						$i+j > I$			

Tablica 1: Inkrementalni trokut razvoja šteta

reći da imamo trokut opaženih i trokut procijenjenih vrijednosti.

Slijedi još nekoliko oznaka za određene dijelove trokuta koje će se koristiti do kraja rada. Ovdje ćemo ih i nacrtati kako bi se kasnije, kada se uvedu prilikom izgradnje modela, lakše razumijelo na što se odnose.

- Sa $\mathcal{D}_I = \{X_{i,j} : i+j \leq I, 0 \leq j \leq J\}$ označavamo gornji trokut opaženih vrijednosti prikazanih u Tablici 1
- Sa $\mathcal{D}_I^c = \{X_{i,j} : i+j > I, i \leq I, j \leq J\}$ označavamo donji trokut procijenjenih vrijednosti prikazanih u Tablici 1
- Sa $\mathcal{B}_k = \{X_{i,j} : i+j \leq I, 0 \leq j \leq k, k \leq J\}$ označavamo podskup od $\mathcal{D}_I$ kako je označeno u Tablici 2
- Sa $\mathcal{D}_{I,i}^0 = \{X_{k,j} \in \mathcal{D}_I : j > I-i\}$ označavamo podskup od $\mathcal{D}_I$ kako je označeno u Tablici 2



Tablica 2: Podskupovi trokuta razvoja šteta

Generalna pretpostavka koja vrijedi u ovom radu je:

$$I = J$$

tj. vrijedi $X_{i,j} = 0$ za sve $j > J$.

Za kraj uvodnog dijela precizirati ćemo što je zapravo cilj ovakve vrste analize podataka odnosno aktuarskih izračuna. Već je rečeno da je cilj odrediti buduće obveze društva u smislu pričuve šteta. U kontekstu opisanih trokuta, to bi značilo da se procijeni donji trokut točnije tzv konačni razvoj šteta koji nije ništa drugo nego $C_{i,J}$ po godinama nastanka $i \in \{0, 1, ..., I\}$. Uz nominalne vrijednosti konačnih šteta željeli bi znati i koliko su nam dobiveni rezultati pouzdani.
Do kraja rada pokušati ćemo doći do tog cilja na način da

1. **izračunamo konačne štete** $C_{i,J}$ - to možemo pomoću samog algoritma
2. **kvantificiramo pouzdanost dobivenih** $C_{i,J}$ - za to nam trebaju stohastički modeli

3 Model ulančanih ljestvica

Ovo poglavlje otvaramo pregledom algoritma metode ulančanih ljestvica, nakon čega ga postepeno stavljamo u stohastički okvir. Konačni cilj je kvantificirati pouzdanost dobivenih rezultata. Na putu do tog cilja, prirodno će se oblikovati podloga za bootstrap metodu.
Napomena vezana za korištenu terminologiju: metoda ulančanih ljestvica odnosi se na algoritam odnosno determinističku metodu, a model ulančanih ljestvica na stohastički model.

3.1 Algoritam metode ulančanih ljestvica

Algoritam metode ulančanih ljestvica primijenjuje se na trokute kumulativnih podataka. Ograničili smo se na trokute iznosa šteta $C_{i,j}$. Kao što je rečeno u prethodnom poglavlju gdje su opisani podaci, prvi cilj je dobiti konačni razvoj šteta $C_{i,J}$ za svaku godinu nastanka $0 \leq i \leq I$.
Algoritam metode ulančanih ljestvica trokut razvija rekurzivno pomoću tzv. razvojnih faktora $\widehat{f}_j, j = 0, ..., J-1$ na način da su konačne štete, označit ćemo ih sa $\widehat{C_{i,J}}^{CL}$ (CL kao oznaka za eng. chain ladder), jednake

$$\widehat{C_{i,J}}^{CL} = C_{i,J-1}\widehat{f}_{J-1} = ... = C_{i,I-i}\widehat{f}_{I-i} \cdots \widehat{f}_{J-1} \tag{1}$$

pri čemu su razvojni faktori procijenjeni s

$$\widehat{f}_j = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}} = \sum_{i=0}^{I-j-1} \frac{C_{i,j}}{\sum_{k=0}^{I-j-1} C_{k,j}} \frac{C_{i,j+1}}{C_{i,j}} \tag{2}$$

Ako s

$$F_{i,j+1} = \frac{C_{i,j+1}}{C_{i,j}} \tag{3}$$

označimo individualne razvoje faktore po godinama nastanka i, onda $\widehat{f}_j$ možemo interpretirati kao težinski prosjek individualnih razvoj faktora. Ovime je opisana deterministička metoda ulančanih ljestvica odnosno pripadni algoritam.

Stohastički okvir metode ulančanih ljestvica. Sada ćemo postaviti okvir unutar kojeg ćemo u nastavku graditi stohastički model.

Algoritam krećemo stohastički oživljavati sa temeljne dvije pretpostavke koje zapravo prešutno postoje u njegovoj pozadini.

Pretpostavke modela 3.1. *(model ulančanih ljestvica bez pretpostavljene distribucije)*

P1. *Kumulativne štete $C_{i,j}$ za različite godine nastanka su nezavisne*

P2. *Postoje razvojni faktori $f_0, ..., f_{J-1} > 0$ takvi da za svaki $0 \leq i \leq I$ i za sve $1 \leq j \leq J$ vrijedi*

$$E[C_{i,j}|C_{i,0}, ..., C_{i,j-1}] = E[C_{i,j}|C_{i,j-1}] = C_{i,j-1}f_{j-1} \tag{4}$$

Model kojeg ćemo razvijati u ovom poglavlju u niti jednom dijelu ne sadrži pretpostavku o distribuciji podataka pa se zbog toga naziva model ulančanih ljestvica bez pretpostavljene distribucije. Za početak primijetimo da u 3.1 imamo pretpostavke samo na prve momente koji su nam zapravo dovoljni za procjenu uvjetno očekivanih budućih šteta (a time i opis samog algoritma). Međutim iz navedenog ne možemo ništa zaključiti o pouzdanosti dobivenih rezultata. Za sada ćemo se kratko zadržati na 3.1 kako bi opisali osnovne stohastičke postavke, nakon čega nastavljamo graditi model s ciljem kvantificiranja pouzdanosti. Prva pretpostavka zajednička je skoro svim metodama rezerviranja, a njome se najviše želi eliminirati utjecaj pojedinih financijskih (kalendarskih) godina. Vrijedi sljedeći rezultat:

Lema 3.2. *Pod Pretpostavkama modela 3.1 za sve $1 \leq i \leq I$ vrijedi*

$$E[C_{i,J}|\mathcal{D}_I] = E[C_{i,J}|C_{i,I-i}] = C_{i,I-i}f_{I-i} \cdots f_{J-1} \tag{5}$$

Lema 3.2 daje zapravo formulu za procjenu konačnih šteta $C_{i,J}$ uz poznatu informaciju o gornjem trokutu $\mathcal{D}_I$. Ako razvojne faktore f_i zamijenimo s $\widehat{f}_i$ definiranim formulom (2) iz prethdno opisanog algoritma dolazimo upravo do procjene konačnog razvoja šteta metodom ulančanih ljestvica

$$\widehat{C_{i,j}}^{CL} = \widehat{E}[C_{i,j}|\mathcal{D}_I] = C_{i,I-i}\widehat{f}_{I-i} \cdots \widehat{f}_{j-1} \tag{6}$$

Slijedi nekoliko rezultata koji proizlaze iz ovakve definicije modela.

Lema 3.3. *Pod Pretpostavkama modela 3.1 vrijedi*

(a) uz dano $\mathcal{B}_j$, $\widehat{f}_j$ je nepristran procjenitelj za f_j: $E[\widehat{f}_j|\mathcal{B}_j] = f_j$

(b) $\widehat{f}_j$ je (bezuvjetno) nepristran procjenitelj za f_j: $E[\widehat{f}_j] = f_j$

(c) $E[\widehat{f}_0...\widehat{f}_j] = E[\widehat{f}_0]...E[\widehat{f}_j]$, odnosno $\widehat{f}_0, ..., \widehat{f}_{J-1}$ su nekorelirani

(d) uz dano $C_{i,I-i}$, $\widehat{C_{i,J}}^{CL}$ je nepristran procjenitelj za $E[C_{i,J}|\mathcal{D}_I] = E[C_{i,J}|C_{i,I-i}]$, odnosno vrijedi $E[\widehat{C_{i,J}}^{CL}|C_{i,I-i}] = E[C_{i,J}|\mathcal{D}_I]$

(e) $\widehat{C_{i,J}}^{CL}$ *je (bezuvjetno) nepristran procjenitelj za* $E[C_{i,J}]$*, odnosno vrijedi* $E[\widehat{C_{i,J}}^{CL}] = E[C_{i,J}]$

Osvrt na Lemu 3.3:

- Iako su $\widehat{f_j}$ nepristrani procjenitelji za f_j i to bez pretpostavke o distribuciji, može se postaviti pitanje što je s ostalim nepristranim procjeniteljima i kolika im je pouzdanost. To je tema koja će se u sljedećem poglavlju obraditi.
- Rezultat nekoreliranosti procjenjenih faktora $\widehat{f_j}$ nije možda na prvu intuitivan zbog toga što susjedni razvojni faktori ovise o istim brojkama - jednom su u naziviniku drugom u brojniku. Općenito, nekoreliranost ne povlači nezavisnost. Uskoro ćemo pokazati da takvi faktori zapravo i nisu nezavisni.

3.2 Stohastički model ulančanih ljestvica

U uvodnom dijelu se dalo naslutiti u kojem smjeru će ići stohastiči model. Kao nastavak Pretpostavkama modela 3.1 slijedi nadogradnja istih tako što se u priču uključuju drugi momenti. Oni su ujedno i put ka kvantificiranju varijabilnosti modela.

Pretpostavke modela 3.4. *(model ulančanih ljestvica bez pretpostavljene distribucije)*

P1. *Kumulativne štete* $C_{i,j}$ *za različite godine nastanka su nezavisne*

P2. $(C_{i,j})_{j\geq 0}$ *je Markovljev lanac* [1]*. Postoje razvojni faktori* $f_0, ..., f_{J-1} > 0$ *i parametri varijance* $\sigma_0^2, ..., \sigma_{J-1}^2 > 0$ *takvi da za svaki* $0 \leq i \leq I$ *i za sve* $1 \leq j \leq J$ *vrijedi*

$$E[C_{i,j}|C_{i,j-1}] = f_{j-1}C_{i,j-1} \tag{7}$$

$$Var(C_{i,j}|C_{i,j-1}) = \sigma_{j-1}^2 C_{i,j-1} \tag{8}$$

Primijetimo da i dalje nemamo pretpostavku o distribuciji kumulativnih šteta, samo pretpostavke na prva dva momenta (i prešutnu pretpostavku da su $C_{i,j} > 0$ inače (8) nebi imala smisla).

[1] Markovljev lanac s diskretnim vremenom i prebrojivim skupom stanja S je slučajni proces $X = (X_n : n \geq 0)$ definiran na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$ za koji vrijedi sljedeće Markovljevo svojstvo:

$$\mathbb{P}(X_{n+1} = j|X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0) = \mathbb{P}(X_{n+1} = j|X_n = i)$$

za svaki $n \geq 0$ i za sve $i_0, ..., i_{n-1}, i, j \in S$ za koje su obje uvjetne vjerojatnosti dobro definirane.

Prisjetimo se procjenitelja $\widehat{f_j}$ za razvojne faktore f_j s početka odjeljka:

$$\widehat{f_j} = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}} = \sum_{i=0}^{I-j-1} \frac{C_{i,j}}{\sum_{k=0}^{I-j-1} C_{k,j}} \frac{C_{i,j+1}}{C_{i,j}}. \tag{9}$$

Vidjeli smo da za njih vrijedi da su uz dano $\mathcal{B}_j$, uvjetno ali i bezuvjetno nepristrani za f_j i da su međusobno nekorelirani za sve $j = 0, ..., J-1$. Lako se pokaže da su i individualni razvojni faktori $F_{i,j}$ definirani u (3) uvjetno na $C_{i,j}$ nepristrani procjenitelji za f_j. Sljedeći rezultat je već najavljen a opravdava izbor $\widehat{f_j}$ kao procjenitelja za f_j.

Lema 3.5. *Pod pretpostavkama modela 3.4 procjenitelj $\widehat{f_j}$ je $\mathcal{B}_{j+1} -$ izmjeriv nepristran procjenitelj za f_j, koji ima najmanju uvjetnu varijancu od svih nepristranih linearnih kombinacija nepristranih procjenitelja $(F_{i,j+1})_{0\leq i\leq I-j-1}$ za f_j uvjetno na $\mathcal{B}_j$, odnosno vrijedi*

$$Var(\widehat{f_j}|\mathcal{B}_j) = \min_{\sum_{\alpha_i \in \mathbb{R}} \alpha_i = 1} Var(\sum_{i=0}^{I-j-1} \alpha_i F_{i,j+1}|\mathcal{B}_j) \tag{10}$$

Uvjetna varijanca of $\widehat{f_j}$ dana je s

$$Var(\widehat{f_j}|\mathcal{B}_j) = \frac{\sigma_j^2}{\sum_{i=0}^{I-j-1} C_{i,j}} \tag{11}$$

Dokaz ove leme može se pronaći u [7] Wüthrich M. V. and Merz M. (2008). Bitno je naglasiti da do ovog rezultata ne bi mogli doći bez pretpostavke o postojanju faktora $\sigma_0^2, ..., \sigma_{J-1}^2 > 0$ takvih da vrijedi $Var(C_{i,j}|C_{i,j-1}) = \sigma_{j-1}^2 C_{i,j-1}$.

Ako pretpostavimo da je procjenitelj za parametar σ_j^2 dan kao težinski prosjek odstupanja individualnih razvojnih faktora od procijenjenog $\widehat{f_j}$ odnosno

$$\widehat{\sigma}_j^2 = \frac{1}{I-j-1} \sum_{i=0}^{I-j-1} C_{i,j} (\frac{C_{i,j+1}}{C_{i,j}} - \widehat{f_j})^2 \tag{12}$$

vrijedi sljedeći rezultat:

Lema 3.6. *Pod Pretpostavkama modela 3.4 imamo*

(a) uz dano $\mathcal{B}_j$, $\widehat{\sigma}_j^2$ je nepristran procjenitelj za σ_j^2, odnosno vrijedi $E[\widehat{\sigma}_j^2|\mathcal{B}_j] = \sigma_j^2$

(b) $\widehat{\sigma}_j^2$ je (bezuvjetno) nepristran procjenitelj za σ_j^2, odnosno vrijedi $E[\widehat{\sigma}_j^2] = \sigma_j^2$

Do sada smo pokazali kako procijeniti očekivani konačni razvoj šteta $C_{i,J}$ uz poznatu informaciju o gornjem trokutu $\mathcal{D}_I$, (vidi (6))

$$\widehat{C_{i,J}}^{CL} = \widehat{E}[C_{i,J}|\mathcal{D}_I] = C_{i,I-i}\widehat{f}_{I-i}...\widehat{f}_{J-1}$$

Također pokazali smo kako izgledaju procjene parametara f_j i σ_j^2 kao i njihovu osnovnu karakteristiku a to je nepristranost. Naš cilj je kvantificiranje varijabilnosti $\widehat{C_{i,J}}^{CL}$. Mjera kojom ćemo to opisati je tzv. srednjekvadratna greška procjenitelja (eng. msep - mean square error of prediction) i definirana je na sljedeći način:

Definicija 3.7. *Uvjetna srednjekvadratna greška procjenitelja* $\widehat{C_{i,J}}^{CL}$ *za* $E[C_{i,J}|\mathcal{D}_I]$ *dana je s*

$$msep_{C_{i,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{CL}\right) = E\left[\left(\widehat{C_{i,J}}^{CL} - C_{i,J}\right)^2 |\mathcal{D}_I\right], i \in \{1, ..., I\} \tag{13}$$

Obzirom da je $\widehat{C_{i,J}}^{CL}$ $\mathcal{D}_I$-izmjeriv raspisom (13) dobijemo

$$msep_{C_{i,J}|\mathcal{D}_I}\big(\widehat{C_{i,J}}^{CL}\big) = Var(C_{i,J}|\mathcal{D}_I) + \big(\widehat{C_{i,J}}^{CL} - E[C_{i,J}|\mathcal{D}_I]\big)^2 \tag{14}$$

Prvi izraz s desne strane (14) zove se **uvjetna varijanca procesa** koja, kao što joj ime kaže, opisuje varijabilnost samog stohastičkog modela odnosno neizvjesnosti budućih događaja. Drugi izraz u (14) zove se **uvjetna pogreška procjene parametara** i govori nam koliko su naše procjene parametara na kojima se zasniva model pouzdane. Iz definicije je jasno da za mjeru varijabilnosti trebamo dati procjenu i za uvjetnu varijancu procesa i za pogrešku procjene parametara (obzirom da f_j i σ_j^2 procjenjujemo s $\widehat{f}_j$ odnosno $\widehat{\sigma}_j^2$).

3.2.1 Uvjetna varijanca procesa

Prvo ćemo vidjeti što možemo reći o uvjetnoj varijanci procesa.

$$\begin{aligned}
Var\left(C_{i,J}|\mathcal{D}_I\right) &= Var\left(C_{i,J}|C_{i,I-i}\right) \\
&= E\left[Var\left(C_{i,J}|C_{i,J-1}\right)|C_{i,I-i}\right] + Var\left(E\left[C_{i,J}|C_{i,J-1}\right]|C_{i,I-i}\right) \\
&= \sigma_{J-1}^2 E\left[C_{i,J-1}|C_{i,I-i}\right] + f_{J-1}^2 Var\left(C_{i,J-1}|C_{i,I-i}\right) \\
&= \sigma_{J-1}^2 C_{i,I-i}\prod_{m=I-i}^{J-2} f_m + f_{J-1}^2 Var\left(C_{i,J-1}|C_{i,I-i}\right)
\end{aligned} \tag{15}$$

Dobili smo rekurzivnu formulu za uvjetnu varijancu procesa za godinu nastanka $i > 0$. Ako nastavimo iterirati formulu, dobijemo

$$
\begin{aligned}
Var(C_{i,J}|C_{i,I-i}) &= C_{i,I-i} \sum_{j=I-i}^{J-1} \prod_{n=j+1}^{J-1} f_n^2 \sigma_j^2 \prod_{m=I-i}^{j-1} f_m \\
&= \sum_{j=I-i}^{J-1} \prod_{n=j+1}^{J-1} f_n^2 \sigma_j^2 E[C_{i,J}|C_{i,I-i}] \\
&= (E[C_{i,J}|C_{i,I-i}])^2 \sum_{j=I-i}^{J-1} \frac{\sigma_j^2/f_j^2}{E[C_{i,j}|C_{i,I-i}]}
\end{aligned}
\tag{16}
$$

čime smo dokazali sljedeću Lemu

Lema 3.8. *Pod pretpostavkama modela 3.4, uvjetna varijanca procesa konačnih šteta $\widehat{C_{i,J}}^{CL}$ po godini nastanka $i \in \{1, ...I\}$ dana je s*

$$
Var(C_{i,J}|\mathcal{D}_I) = (E[C_{i,J}|C_{i,I-i}])^2 \sum_{j=I-i}^{J-1} \frac{\sigma_j^2/f_j^2}{E[C_{i,j}|C_{i,I-i}]} \tag{17}
$$

Uvjetnu varijancu procesa procjenjujemo s

$$
\widehat{Var}(C_{i,J}|\mathcal{D}_I) = \left(\widehat{C_{i,J}}^{CL}\right)^2 \sum_{j=I-i}^{J-1} \frac{\widehat{\sigma}_j^2/\widehat{f}_j^2}{\widehat{C_{i,j}}^{CL}}. \tag{18}
$$

Kako su godine nastanka međusobno nezavisne vrijedi da je varijanca zbroja godina nastanka jednaka zbroju varijanci. Stoga je procjenitelj uvjetne varijance procesa za sve godine nastanka zajedno dan s

$$
\widehat{Var}\left(\sum_{i=1}^{I} C_{i,J}|\mathcal{D}_I\right) = \sum_{i=1}^{I} \widehat{Var}(C_{i,J}|\mathcal{D}_I). \tag{19}
$$

3.2.2 Uvjetna pogreška procjene parametara

Preostaje nam analizirati uvjetnu pogrešku procjene parametara. Za godinu nastanka i ona je dana s

$$
\begin{aligned}
\left(\widehat{C_{i,J}}^{CL} - E[C_{i,J}|\mathcal{D}_I]\right)^2 &= C_{i,I-i}^2 \left(\widehat{f}_{I-i} \cdots \widehat{f}_{J-1} - f_{I-i} \cdots f_{J-1}\right)^2 \\
&= C_{i,I-i}^2 \left(\prod_{j=I-i}^{J-1} \widehat{f}_j^2 + \prod_{j=I-i}^{J-1} f_j^2 - 2 \prod_{j=I-i}^{J-1} \widehat{f}_j f_j \right)
\end{aligned}
\tag{20}
$$

Primijetimo da su procijenjeni faktori $\widehat{f}_{I-i}, ..., \widehat{f}_{J-1}$ poznati u trenutku I, ali *pravi* faktori $f_{I-i}, ..., f_{J-1}$ nisu, stoga se izraz (20) ne može izračunati. Kako bi dali procjenu istog analitizati ćemo na koji način procijenjeni faktori $\widehat{f}_j$ fluktuiraju oko *pravih* faktora f_j. Metode kojima se to postiže su razne: od Bayesovih preko metoda uzorkovanja kao što su bootstrap i Monte Carlo metode do zatvorenih analitičkih metoda. Do kraja ovog poglavlja baviti ćemo se zatvorenim analitičkim formama, međutim, kao što smo već spomenuli, prirodno će se postaviti podloga za bootstrap metodu. Iako u nastavku nećemo uzorkovati u pravom smislu te riječi (kao što ćemo to činiti kod bootstrap metode) koristiti ćemo taj izraz.
Vratimo se na izraz (20) i fiksirajmo godinu $i \in \{1, \dots, I\}$. Vidimo da je najveći izazov odrediti volatilnost kvadrata procijenjenih faktora. Zadnji izraz oslanja se na nepristranost i nekoreliranost. Stoga, trebamo uzorkovati sljedeći umnožak kvadrata procijenjenih faktora

$$\widehat{f}_{I-i}^2 \cdots \widehat{f}_{J-1}^2.$$

Obradit ćemo dva pristupa uzorkovanja u zatvorenim analitičkim formama: bezuvjetni i uvjetni.

Pristup 1 - Bezuvjetno uzorkovanje uz $\mathcal{D}_{I,i}^0$

U ovom pristupu računamo sljedeće očekivanje

$$E[\widehat{f}_{I-i}^2 \cdots \widehat{f}_{J-1}^2 | \mathcal{B}_{I-i}] \tag{21}$$

Kako vrijedi $\mathcal{D}_{I,i}^0 \cap \mathcal{B}_{I-i} = \emptyset$, (21) ne ovisi o opaženim vrijednostima u gornjem trokutu $\mathcal{D}_{I,i}^0$ što znači da one ne utječu na određivanje pogreške procjene parametara pa se zato ovaj pristup zove bezuvjetno uzorkovanje.
Kako smo već spomenuli, zbog nekoreliranosti i nepristranosti (Lema (3.3)), vrijedi

$$\begin{aligned} E&\left[\left(\widehat{C_{i,J}}^{CL} - E[C_{i,J}|\mathcal{D}_I]\right)^2 |\mathcal{B}_{I-i}\right] \\ &= C_{i,I-i}^2 E\left[\prod_{j=I-i}^{J-1} \widehat{f}_j^2 + \prod_{j=I-i}^{J-1} f_j^2 - 2\prod_{j=I-i}^{J-1} \widehat{f}_j f_j \middle| \mathcal{B}_{I-i}\right] \\ &= C_{i,I-i}^2 \left(E\left[\prod_{j=I-i}^{J-1} \widehat{f}_j^2 \middle| \mathcal{B}_{I-i}\right] - \prod_{j=I-i}^{J-1} f_j^2\right) \end{aligned} \tag{22}$$

Dakle, kako bi procijenili grešku procjene parametara metodom bezuvjetnog uzorkovanja trebali bi izračunati (21). To bi bilo jednostavnije kada bi procijenjeni parametri $\widehat{f}_j$ bili nezavisni. Lema 3.3 i sljedeća Lema nam ukazuju da su oni nekorelirani, ali ne i nezavisni.

Lema 3.9. *Pod Pretpostavkama modela 3.4 kvadrati dva uzastopna procjenitelja $\widehat{f}_{j-1}$ i $\widehat{f}_j$ uz uvjet da je poznato $\mathcal{B}_{j-i}$ su negativno korelirani, odnosno vrijedi*

$$Cov\left(\widehat{f}_{j-1}^2, \widehat{f}_j^2 | \mathcal{B}_{j-i}\right) < 0$$

za $1 \leq j \leq J-1$.

S tog aspekta bezuvjetno uzorkovanje nije obečavajući pristup u okvirima zatvorene analitičke forme. Međutim obradit ćemo je kroz bootstrap metodu.

Pristup 2 - Uvjetno uzorkovanje uz $\mathcal{D}_{I,i}^0$

Ovim pristupom pokušati ćemo izračunati sljedeća očekivanja

$$E\left[\widehat{f}_{I-i}^2 | \mathcal{B}_{I-i}\right] \ldots E\left[\widehat{f}_{J-1}^2 | \mathcal{B}_{J-1}\right] \tag{23}$$

Kako vrijedi $\mathcal{D}_{I,i}^0 \cap \mathcal{B}_j \neq \emptyset$ za $j > I-i$, opažene vrijednosti u gronjem trokutu $\mathcal{D}_{I,i}^0$ imaju direktan utjecaj na procjenu (23). Zbog toga se ovaj pristup naziva uvjetno uzorkovanje uz $\mathcal{D}_{I,i}^0$.
Kako bi omogućili ovu vrstu uzorkovanja, još jednom (posljedni put za model ulančanih ljestvica) nadograđujemo odnosno postrožujemo pretpostavke uvodeći vremenske nizove.

Pretpostavke modela 3.10. *(model ulančanih ljestvica bez pretpostavljene distribucije)*

- ***P1.*** *Kumulativne štete $C_{i,j}$ za različite godine nastanka su nezavisne*

- ***P2.*** *Postoje konstante $f_j > 0$, $\sigma_j > 0$ i slučajne varijable $\varepsilon_{i,j+1}$ takve da za svaki $0 \leq i \leq I$ i za sve $0 \leq j \leq J-1$ vrijedi*

$$C_{i,j+1} = f_j C_{i,j} + \sigma_j \sqrt{C_{i,j}} \varepsilon_{i,j+1} \tag{24}$$

  gdje su $\varepsilon_{i,j+1}$ uz dano $\mathcal{B}_0$ uvjetno nezavisni i $E\left[\varepsilon_{i,j+1} | \mathcal{B}_0\right] = 0$ i $E\left[\varepsilon_{i,j+1}^2 | \mathcal{B}_0\right] = 1$ te vrijedi $P\left[C_{i,j+1} > 0 | \mathcal{B}_0\right] = 1$ za sve $0 \leq i \leq I$ i za sve $0 \leq j \leq J-1$

Osvrt na Pretpostavke modela 3.10

- Slučajne varijable $\varepsilon_{i,j+1}$ su definirane uvjetno na $\mathcal{B}_0$ kako bi se osiguralo da su kumulativne štete $C_{i,j+1}$ pozitivne, $P[\cdot|\mathcal{B}_0] - g.s.$. Ovo implicitno povlači da svi izračuni sadrže pretpostavku uvjetne vjerojatnosti $P[\cdot|\mathcal{B}_0]$, koju ćemo zbog jednostavnosti označavati s P.

- Može se pokazati da Pretpostavke modela 3.10 impliciraju Pretpostavke modela 3.4.

Dakle potrebno je uzorkovati vrijednosti $\widehat{f}_{I-i}, \ldots, \widehat{f}_{J-1}$ uz dani gornji trokut $\mathcal{D}_I$. Ne zaboravimo da nas je do ovdje doveo cilj da izmjerimo odstupanja $\widehat{f}_j$ od f_j. Kako bi to postigli, postupat ćemo na sljedeći način. Uz poznat gornji trokut $\mathcal{D}_I$ generirat ćemo "nove" vrijednosti $\widetilde{C}_{i,j+1}$ za $i \in \{0, ..., I\}$ i $j \in \{0, ..., J-1\}$ koristeći formulu

$$\widetilde{C}_{i,j+1} = f_j C_{i,j} + \sigma_j \sqrt{C_{i,j}} \widetilde{\varepsilon}_{i,j+1} \tag{25}$$

gdje je $\sigma_j > 0$ i $\widetilde{\varepsilon}_{i,j+1}$, $\varepsilon_{i,j+1}$ nezavisne jednako distribuirane kopije uz dano $\mathcal{B}_0$. Korištena je drugačija notacija kako bi se naglasilo da su $\widetilde{C}_{i,j+1}$ slučajne varijable a $C_{i,j}$ determinističke vrijednosti uz dano $\mathcal{D}_I$.
U duhu (23) i prethodno navedenog, vrijednosti $\widehat{f}_j$ uzorkovat ćemo na način da ćemo samo uzorkovati vrijednosti za godinu $j+1$, dok će godina j biti fiksna. Za nove procijenjene faktore $\widehat{f}_j$ vrijedi (uz Pretpostavke modela 3.10):

$$\widehat{f}_j = \frac{\sum_{i=0}^{I-j-1} \widetilde{C}_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}} = f_j + \frac{\sigma_j}{S_j^{[I-j-1]}} \sum_{i=0}^{I-j-1} \sqrt{C_{i,j}} \widetilde{\varepsilon}_{i,j+1} \tag{26}$$

pri čemu je

$$S_j^{[I-j-1]} = \sum_{i=0}^{I-j-1} C_{i,j} \tag{27}$$

Sa $P^*_{\mathcal{D}_I}$ definiramo vjerojatnosnu mjeru na sljedeći način: uz poznato $\mathcal{D}_I$, svaki razvojni period $j \in \{0, ..., J-1\}$ mjerimo uvjetno na informacije poznate do dog trenutka, odnosno uvjetno na $\mathcal{B}_j$. Tako smo zapravo i krenuli sa uvjetnim uzorkovanjem, pogledati (23). Procijenjeni razvojni faktori $\widehat{f}_j$ metodom uzorkovanja na gore opisan način imaju istu distribuciju kao originalno izračunati razvojni faktori (9), uz uvjet $\mathcal{B}_j$, odnosno vjerojatnosnu mjeru $P^*_{\mathcal{D}_I}$. Treba imati na umu da su za razliku od opaženih vrijednosti $C_{i,j} : i + j \leq I$, $\widetilde{C}_{i,j} : i + j \leq I$ kao i uzorkovani procjenitelji $\widehat{f}_j$ slučajne varijable uz dano $\mathcal{D}_I$. Ako uz sve navedeno uzmemo u obzir da su slučajne varijable $\widetilde{\varepsilon}_{i,j+1}$ nezavisne uz dano $\mathcal{B}_0 \subset \mathcal{B}_j \subset \mathcal{D}_I$, dobijemo sljedeći rezultat:

(1) uzorkovani procjenitelji $\widehat{f}_0, ..., \widehat{f}_{J-1}$ su nezavisni uz vjerojatnosnu mjeru $P^*_{\mathcal{D}_I}$

(2) $E^*_{\mathcal{D}_I}\left[\widehat{f}_j\right] = f_j$ za $0 \leq j \leq J-1$

(3) $E^*_{\mathcal{D}_I}\left[\widehat{f}_j^2\right] = f_j^2 + \sigma_j^2 / S_j^{[I-j-1]}$ za $0 \leq j \leq J-1$

Sada smo spremni vidjeti na koji način ćemo u pristupu 2 procijeniti uvjetnu pogrešku procjene parametara

$$\begin{aligned}
&C_{i,I-i}^2 E_{\mathcal{D}_I}^*\left[\left(\widehat{f}_{I-i}\cdots\widehat{f}_{J-1}-f_{I-i}\cdots f_{J-1}\right)^2\right]\\
&\quad= C_{i,I-i}^2 Var_{\mathcal{P}_{\mathcal{D}_I^*}}\left(\widehat{f}_{I-i}\cdots\widehat{f}_{J-1}\right)\\
&\quad= C_{i,I-i}^2\left(\prod_{j=I-i}^{J-1} E_{\mathcal{D}_I}^*\left[\left(\widehat{f}_j\right)^2\right]-\prod_{j=I-i}^{J-1} f_j^2\right)\\
&\quad= C_{i,I-i}^2\left(\prod_{j=I-i}^{J-1}\left(f_j^2+\frac{\sigma_j^2}{S_j^{[I-j-1]}}\right)-\prod_{j=I-i}^{J-1} f_j^2\right)
\end{aligned}\tag{28}$$

Ako parametre $\sigma_{I-i}^2,...,\sigma_{J-1}^2$ i $f_{I-i},...,f_{J-1}$ zamjenimo sa pripadnim procjenama dobijemo sljedeću procjenu uvjetne pogreške procjene parametara

$$\begin{aligned}
\widehat{Var}\left(\widehat{C_{i,J}}^{CL}|\mathcal{D}_I\right)&=\widehat{E}_{\mathcal{D}_I}^*\left[\left(\widehat{C_{i,J}}^{CL}-E\left[C_{i,J}|\mathcal{D}_I\right]\right)^2\right]\\
&=C_{i,I-i}^2\left(\prod_{j=I-i}^{J-1}\left(\widehat{f}_j^2+\frac{\widehat{\sigma}_j^2}{S_j^{[I-j-1]}}\right)-\prod_{j=I-i}^{J-1}\widehat{f}_j^2\right)
\end{aligned}\tag{29}$$

Lema 3.11. *(procjenitelj msep za jednu godinu nastanka)*
Pod Pretpostavkama modela 3.10 dobili smo sljedeći procjenitelj za uvjetnu srednjekvadratnu grešku procjenitelja (msep) konačnih šteta $\widehat{C_{i,J}}^{CL}$ *za pojedinu godinu nastanka* $i\in 1,...,I$

$$\begin{aligned}
\widehat{msep}_{C_{i,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{CL}\right)&=\widehat{E}\left[\left(\widehat{C_{i,J}}^{CL}-C_{i,J}\right)^2|\mathcal{D}_I\right]\\
&=\left(\widehat{C_{i,J}}^{CL}\right)^2\sum_{j=I-i}^{J-1}\frac{\sigma_j^2/f_j^2}{\widehat{C_{i,J}}^{CL}}\\
&+C_{i,I-i}^2\left(\prod_{j=I-i}^{J-1}\left(\widehat{f}_j^2+\frac{\widehat{\sigma}_j^2}{S_j^{[I-j-1]}}\right)-\prod_{j=I-i}^{J-1}\widehat{f}_j^2\right)
\end{aligned}\tag{30}$$

Iz Pretpostavki modela 3.10 (kao i prethodnih varijanti) znamo da su kumulativne štete $C_{i,J}$ za različite godine nastanka nezavisne. Međutim to ne vrijedi i za njihove procjenitelje $\widehat{C_{i,J}}^{CL}$ obzirom da se zasnivaju na istim razvojnim faktorima. Zbog toga trebamo biti pažljivi kod procjene varijabilnosti konačnih šteta za sve godine nastanka zajedno. Navedeno nam neće utjecati na procjenu uvjetne varijance procesa jer se ona odnosi samo na $C_{i,J}$, kako smo i raspisali u (19).

Za uvjetnu pogrešku procjene parametara, međuovisnosti $\widehat{C_{i,J}}^{CL}$ različitih godina nastanka se trebaju uzeti u obzir i to na način kako je navedeno u sljedećoj lemi:

Lema 3.12. *(procjenitelj msep za sve godine nastanka zajedno)*
Pod Pretpostavkama modela 3.10 dobili smo sljedeći procjenitelj za uvjetnu srednjekvadratnu grešku procjenitelja (msep) konačnih šteta za sve godine nastanka zajedno

$$
\begin{aligned}
\widehat{msep}_{\sum_i C_{i,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{CL}\right) &= \widehat{E}\left[\left(\sum_{i=1}^{I}\widehat{C_{i,J}}^{CL} - \sum_{i=1}^{I} C_{i,J}\right)^2 \Bigg| \mathcal{D}_I\right] \\
&= \sum_{i=1}^{I}\widehat{msep}_{C_{i,J}|\mathcal{D}_I}\left(\widehat{C_{i,J}}^{CL}\right) \\
&\quad + 2\sum_{1\le i<k\le I} C_{i,I-i}\widehat{C_{k,I-i}}^{CL}\left(\prod_{j=I-i}^{J-1}\left(\widehat{f}_j^2 + \frac{\widehat{\sigma}_j^2}{S_j^{[I-j-1]}}\right) - \prod_{j=I-i}^{J-1}\widehat{f}_j^2\right)
\end{aligned}
\tag{31}
$$

4 Raspršeni Poissonov model

Za razliku od modela kojeg smo opisali u prethodnom poglavlju i to bez pretpostavke o distribuciji, u ovom poglavlju imat ćemo eksplicitnu pretpostavku o distribuciji podataka - Poissonova [2]. To baca potpuno drugačije svjetlo na modeliranje budućih obveza, a razlog zbog kojeg ćemo analizirati takav model je taj što on daje identične pričuve kao i prethodno promatrani - ulančanih ljestvica. Dakle, pokazat ćemo da su očekivane vrijednosti oba modela iste, a ono gdje će se oni razlikovati jesu viši momenti. Što se tiče izgleda poglavlja, oni su također različiti. U prošlom dijelu smo postepeno od algoritma do stohastičkog okvira preko razvoja modela i pretpostavki došli do cilja - mjera varijabilnost rezultata. U ovom poglavlju imamo drugačiji pristup - kroz uvod kratko predstavljamo generalizirane linearne modele nakon čega se ograničavamo samo na jednu vrstu distribucije iz eksponencijalne familije - Poissonovu.

Kao što je već rečeno, a nije loše ponoviti prije nego krenemo graditi model, naš cilj je na osnovu povijesnih podataka procijeniti buduće štete uz pripadnu mjeru pouzdanosti dobivenih rezultata. Prethodni model građen je ne kumulativnim štetama $C_{i,j}$. U ovom poglavlju promatrati ćemo isključivo inkrementalne štete $X_{i,j}$. U poglavlju 2 dan je detaljan pregled oznaka.

4.1 Generalizirani linearni modeli

Ako sa $x_{i,j}$ označimo podatke za koje želimo utvrditi na koji način utječu na varijable $\widehat{x}_{i,j}$, tada $(x_{i,j}, \widehat{x}_{i,j})$ predstavlja niz parova pri čemu je $\widehat{x}_{i,j}$ realizacija slučajne varijable $X_{i,j}$ čija razdioba ovisi upravo o kovarijatama $x_{i,j}$. Osnovna ideja generaliziranih linearnih modela je da postoji veza između očekivanja varijable odaziva i kovarijata na sljedeći način

$$E\left[X_{i,j}\right] = g^{-1}\left(\boldsymbol{\Gamma_{i,j}}\mathbf{x}\right) \tag{32}$$

pri čemu je

- $\mathbf{x}$ vektor kovarijata
- g^{-1} inverz funkcije g
- $\eta_{i,j} = \boldsymbol{\Gamma_{i,j}}\mathbf{x}$ tzv. linearni prevoditelj

[2]Neka je $\lambda \in \mathbb{R}$, $\lambda > 0$. Kažemo da slučajna varijabla X ima Poissonovu distribuciju ako joj je distribucija dana s

$$P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad k \in \mathbb{N} \cup \{0\}$$

Broj λ zovemo parametar Poissonove slučajne varijable

- $X_{i,j}$ ima unaprijed određenu razdiobu iz eksponencijalne familije

Zadnja točka odnosno pretpostavka da varijabla odaziva mora imati razdiobu iz neke od eksponencijalnih familija je iz teorijskih razloga. Međutim takva restrikcija je prihvatljiva jer ove familije uključuju najčešće korištene razdiobe. U nastavku definicija eksponencijalnih familija razdioba.

Definicija 4.1. *Kažemo da slučajna varijabla $X_{i,j}$ pripada nekoj ekponencijalnoj familiji ako joj gustoća (neprekidna ili diskretna) ima oblik*

$$f(x;\theta_{i,j},\phi_{i,j},\omega_{i,j}) = a\left(x, \frac{\phi_{i,j}}{\omega_{i,j}}\right) exp\left\{\frac{x\theta_{i,j} - b(\theta_{i,j})}{\phi_{i,j}/\omega_{i,j}}\right\} \tag{33}$$

pri čemu je funkcija b uvijek dva puta neprekidno diferencijabilna t.d. je b' invertibilna. Primijetimo da familija ima tri parametra: $\theta_{i,j}$ tzv. ***prirodni parametar****, $\phi_{i,j}$ tzv* ***parametar disperzije*** *i $\omega_{i,j}$ je poznata konstanta a predstavlja faktor težine.*

Vrijede sljedeći rezultati:

Lema 4.2. *Ako slučajna varijabla $X_{i,j}$ pripada ekponencijalnoj familiji distribucija tada vrijedi:*

$$E\left[X_{i,j}\right] = b'(\theta_{i,j}) \tag{34}$$

$$Var\left[X_{i,j}\right] = \frac{\phi_{i,j}}{\omega_{i,j}} b''(\theta_{i,j}) \tag{35}$$

Prvo primijetimo da očekivanje ne ovisi o parametru disperzije $\phi_{i,j}$ već samo o prirodnom parametru $\theta_{i,j}$, dok varijanca ovisi o oba parametra. Stavljajući $\mu_{i,j} = b'(\theta_{i,j})$ uvodimo novi parametar, tzv **parametar srednje vrijednosti**. Obzirom da je b' neprekidno diferencijabilna i invertibilna vrijedi da je **funkcija varijance** $\mu_{i,j} \to V(\mu_{i,j}) = b''(\theta_{i,j}) = b''(b^{-1}(\mu_{i,j}))$ dobro definirana. Kako bi naglasili utjecaj očekivanja na varijancu izrazimo (35) u obliku

$$Var\left[X_{i,j}\right] = \frac{\phi_{i,j}}{\omega_{i,j}} V(\mu_{i,j})$$

Da zaokružimo uvodni dio:

Generalizirani linearni model pretpostavlja da varijable odaziva $X_{i,j}$ opažamo na nezavisan način za različite vrijednosti $x_{i,j}$. Model pri tome ima sljedeće karakteristike:

- Kovarijate $x_{i,j}$ na linearan način utječu na razdiobu od $X_{i,j}$ preko tzv. linearnog prevoditelja $\eta_{i,j} = \boldsymbol{\Gamma_{i,j}}\mathbf{x}$

- Razdioba slučajne varijable $X_{i,j}$ za dane kovarijate pripada uvijek istoj eksponencijalnoj familiji razdioba

- Slijedi da je očekivanje od $X_{i,j}$ glatka i invertibilna funkcija linearnog prediktora $\eta_{i,j}$ oblika $b' \circ h$ za funkciju h, tj.

$$\theta_{i,j} = b'^{-1}(g^{-1}(\eta_{i,j})) = h(\eta_{i,j})$$
$$\mu_{i,j} = E\left[X_{i,j}\right] = b'(\theta_{i,j}) = b'(h(\eta_{i,j})).$$

  Vrijedi da je $g = h^{-1} \circ b'^{-1}$ i zovemo ju **funkcija veze**.

4.2 Raspršeni Poissonov model

Primijenimo do sada rečeno na podatke koje želimo modelirati i rezultate koji su nam od interesa. Dakle poznat nam je gornji trokut $\mathcal{D}_I$ iz kojeg, uz pretpostavku o distribuciji $X_{i,j}$, želimo procijeniti konačni razvoj šteta. Za poznate realizacije iz gornjeg trokuta $\mathcal{D}_I$ očekivali bismo da vrijedi $E\left[X_{i,j}\right] = x_{i,j}$. To znači da bi za procjenu parametara svih razdioba inkrementalnih šteta trebali na neki način procijeniti $(I+1)(J+1)$ nepoznatih parametara. Ako pak uvedemo sljedeću multiplikativnu strukturu:

$$x_{i,j} = \mu_i \gamma_j \tag{36}$$

smanjujemo broj nepoznatih parametara na $I + J + 2$. Ovakva struktura točno definira kako se informacije sadržane u gornjem trokutu prenose na donji trokut.

Pretpostavke modela 4.3. *(raspršeni Poissonov model)*

P1. *Inkrementalne štete $X_{i,j}$ različitih godina nastanka i, i različitih perioda razvoja j su nezavisne*

P2. *Postoje parametri $\mu_0, ..., \mu_I > 0$, $\gamma_0, ..., \gamma_J > 0$ i $\phi > 0$ takvi da inkrementalne štete $X_{i,j}$ pripadaju raspršenoj Poissonovoj familiji razdioba i vrijedi*

$$E\left[X_{i,j}\right] = x_{i,j} = \mu_i \gamma_j \tag{37}$$
$$Var\left[X_{i,j}\right] = \phi x_{i,j} = \phi \mu_i \gamma_j \tag{38}$$
$$\mu_0 = 1 \tag{39}$$

Primijetimo da ovako definirani model pretpostavlja da su inkrementalne štete pozitivne što nije uvijek slučaj. Za trokute gdje to ne vrijedi, Poissonov model nije primjenjiv, barem ne u ovom obliku (postoje metode kojima se to premosti, ali o njima ovdje neće biti riječi). Usporedbe radi, model ulančanih ljestvica primjenjiv je na trokute gdje nisu svi inkrementalni podaci pozitivni dok god kumulativni podaci ostaju pozitivni.

Opišimo pojedine djelove Pretpostavki modela 4.3 rječnikom eksponencijalne familije razdioba:

- $\theta_{i,j} = ln(x_{i,j}) = ln(\mu_i\gamma_j) = ln(\mu_i) + ln(\gamma_j)$
- $b(\theta_{i,j}) = x_{i,j} = \mu_i\gamma_j \to b(\theta_{i,j}) = exp\{\theta_{i,j}\}$
- $\phi_{i,j}/\omega_{i,j} = \phi$

Iz uvodnog dijela ovog poglavlja u kojem smo opisali generalizirane linarne modele znamo da vrijedi sljedeća jednakost

$$x_{i,j} = E\left[X_{i,j}\right] = b'(\theta_{i,j}) = exp\{\theta_{i,j}\} = exp\{ln(\mu_i) + ln(\gamma_j)\} = \mu_i\gamma_j \tag{40}$$

što znači da je za Poissonov model funkcija veze jednaka $g(x_{i,j}) = ln(x_{i,j})$.
Iz jednakosti (40) jasno je da $\eta_{i,j} = ln(\mu_i) + ln(\gamma_j)$. Označimo s

$$\mathbf{b} = (ln(\mu_1), ..., ln(\mu_I), ln(\gamma_0), ..., ln(\gamma_J)) \tag{41}$$

$$\Gamma_{0,j} = (0, ..., 0, 0, 0, ..., 0, e_{I+j+1}, 0, ..., 0) \tag{42}$$

$$\Gamma_{i,j} = (0, ..., 0, e_i, 0, ..., 0, e_{I+j+1}, 0, ..., 0) \tag{43}$$

za sve $1 \leq i \leq I$ i $0 \leq j \leq J$ gdje su $e_i = 1$ i $e_{I+j+1} = 1$ jedinične vrijednosti na i-toj i $(I + j + 1)$-toj poziciji. Sada $\eta_{i,j}$ možemo zapisati u obliku

$$\eta_{i,j} = \Gamma_{i,j}\mathbf{b}$$

Dakle naš cilj je procijeniti $I + J + 1$ nepoznatih parametara μ_i i γ_j, odnosno vektor $\mathbf{b}$. Za procjenu vektora $\mathbf{b}$ koristi se metoda maksimalne vjerodostojnosti uz poznat skup podataka gornjeg trokuta $\mathcal{D}_\mathcal{I} = \{X_{i,j}; i + j \leq I\}$. Što znači da se maksimizira funkcija

$$l_{\mathcal{D}_\mathcal{I}}(\mathbf{b}) = ln\left(\prod_{i+j\leq I} f(X_{i,j}; \mu_i\gamma_j, \phi)\right) = \sum_{i+j\leq I} l(X_{i,j}; \mu_i\gamma_j, \phi) \tag{44}$$

na način da se $I + J + 1$ parcijalna derivacija po nepoznatim parametrima μ_i i γ_j izjednačava sa 0. Rezultat su procijenjeni parametri $\widehat{\mu_i}$ i $\widehat{\gamma_j}$ odnosno pripadni vektor

$$\widehat{\mathbf{b}} = (\widehat{ln(\mu_1)}, ..., \widehat{ln(\mu_I)}, \widehat{ln(\gamma_0)}, ..., \widehat{ln(\gamma_J)}) \tag{45}$$

Lema 4.4. *(MLE za raspršeni Poissonov model) Procjenitelj maksimalne vjerodostojnosti za Model 4.3 dan je s*

$$\widehat{X_{i,j}}^{POI} = \widehat{x_{i,j}} = \widehat{E}\left[X_{i,j}|\mathcal{D}_I\right] = \widehat{\mu_i}\widehat{\gamma_j} \tag{46}$$

$$\widehat{C_{i,J}}^{POI} = \widehat{E}\left[C_{i,J}|\mathcal{D}_I\right] = C_{i,I-i} + \sum_{j=I-i+1}^{J} \widehat{X_{i,j}}^{POI} \tag{47}$$

za $i + j > I$.

POI oznaka predstavlja rezultate raspršenog Poissonovog modela. Raspisom parcijalnih derivacija (44) možemo vidjeti da je za procjenu $\widehat{x_{i,j}}$ odnosno $\widehat{\mu_i}$ i $\widehat{\gamma_j}$ ne trebamo parametar raspršenja ϕ (dok god je konstantan kako smo i pretpostavili u Modelu 4.3). To znači da je procjenitelj maksimalne vjerodostojnosti za Poissonov i raspršeni Poissonov model isti. Također, ovdje nećemo raspisivati sustave jednadžbi za procjenu $\widehat{\mu_i}$ i $\widehat{\gamma_j}$, već je pretpostavka da postoji pozitivno rješenje za iste.
Prije nego krenemo na vrijabilnost rezultata dobivenih ovim modelom, još jednom naglašavamo, što ćemo kasnije i pokazati primjerima, da procjenitelj metode ulančanih ljestvica dan s (6) i procjenitelj dobiven u Poissonovom modelu 4.3 daju istu procjenu pričuva: $\widehat{C_{i,J}}^{CL} = \widehat{C_{i,J}}^{POI}$.

4.2.1 Uvjetna srednjekvadratna greška procjenitelja

Preostaje nam kvantificirati koliko dobro $\widehat{C_{i,J}}^{POI}$ procjenjuje $C_{i,J}$. Kao i u prethodnom poglavlju, mjera koju ćemo koristiti je uvjetna srednjekvadratna greška procjenitelja za koju uz Pretpostavke modela 4.3 vrijedi

$$\begin{aligned}
& msep_{\sum_i C_{i,J}|\mathcal{D}_I}\left(\sum_{i=1}^{I}\widehat{C_{i,J}}^{POI}\right) \\
&= E\left[\left(\sum_{i=1}^{I}\widehat{C_{i,J}}^{POI} - \sum_{i=1}^{I} C_{i,J}\right)^2 \middle| \mathcal{D}_I\right] \\
&= E\left[\left(\sum_{i+j>I}\widehat{X_{i,J}}^{POI} - \sum_{i+j>I} X_{i,J}\right)^2 \middle| \mathcal{D}_I\right] \\
&= Var\left(\sum_{i+j>I} X_{i,j}\middle|\mathcal{D}_I\right) + \left(\sum_{i+j>I}\left(\widehat{X_{i,j}}^{POI} - E\left[X_{i,j}|\mathcal{D}_I\right]\right)\right)^2 \\
&= Var\left(\sum_{i+j>I} X_{i,j}\right) + \left(\sum_{i+j>I}\left(\widehat{X_{i,j}}^{POI} - E\left[X_{i,j}\right]\right)\right)^2 \\
&= \sum_{i+j>I}\phi V(x_{i,j}) + \left(\sum_{i+j>I}\left(\widehat{X_{i,j}}^{POI} - E\left[X_{i,j}\right]\right)\right)^2
\end{aligned} \tag{48}$$

U predzadnjoj i zadnjoj jednakosti korištena je činjenica da su $X_{i,j}$ nezavisni od $\mathcal{D}_I$ za $i + j > I$ te da su $X_{i,j}$ međusobno nezavisni. Vidimo da smo ovim raspisom dali procjenu uvjetne varijance procesa, dok procjena pogrešeke parametara opet treba daljnji raspis. Primijetimo da za bezuvjetnu srednjekvadratnu grešku procjenitelja vrijedi:

$$
\begin{aligned}
& msep_{\sum_i C_{i,J}}\left(\sum_{i=1}^{I} \widehat{C_{i,J}}^{POI}\right) \\
& = E\left[msep_{\sum_i C_{i,J}|\mathcal{D}_I}\left(\sum_{i=1}^{I} \widehat{C_{i,J}}^{POI}\right)\right] \\
& = \sum_{i+j>I} \phi V(x_{i,j}) + E\left[\left(\sum_{i+j>I}\left(\widehat{X_{i,j}}^{POI} - E\left[X_{i,j}\right]\right)\right)^2\right]
\end{aligned} \tag{49}
$$

Iako generalno $\widehat{X_{m,n}}^{POI}$ dobiven metodom maksimalne vjerodostojnosti nije nepristran procjenitelj za $E\left[X_{i,j}\right]$ tu pristanost sada ne uzimamo u obzir zbog toga što je najčešće zanemariva u odnosu na srednjekvadratnu pogrešku. Stoga, ako očekivani izraz procijenimo sa $Var(\widehat{X_{i,j}}^{POI})$ odnosno $Cov(\widehat{X_{i,j}}^{POI}, \widehat{X_{m,n}}^{POI})$ dobivamo procjenitelj oblika

$$
E\left[\left(\sum_{i+j>I}\left(\widehat{X_{i,j}}^{POI} - E\left[X_{i,j}\right]\right)\right)^2\right] \approx \sum_{i+j>I, n+m>I} x_{i,j} x_{n,m} \Gamma_{i,j} Cov(\widehat{b}, \widehat{b}) \Gamma'_{n,m} \tag{50}
$$

Ostaje procijeniti $Cov(\widehat{b}, \widehat{b})$. Taj dio nećemo raspisivati, samo ćemo dati konačan rezultat: kovarijanca procjenitelja maksimalne vjerodostojnosti $\widehat{b}$ procjenjuje se tzv. Fisherovom matricom $H(\widehat{b})$ odnosno vrijedi

$$
\widehat{Cov}(\widehat{b}, \widehat{b}) = H(\widehat{b})^{-1} = \left(\left(\sum_{i+j\leq I} W(\widehat{x}_{i,j}) \Gamma_{i,j}^{(k)} \Gamma_{i,j}^{(l)}\right)_{k.l=1,\ldots,I+J+1}\right)^{-1} \tag{51}
$$

pri čemu je uz Pretpostavke modela 4.3

$$
W(\widehat{x}_{i,j}) = \phi^{-1} \frac{\widehat{x}_{i,j}^2}{V(\widehat{x}_{i,j})} \tag{52}
$$

Uz sve navedeno procjena varijabilnosti raspršenog Poissonovog modela dana je s

Lema 4.5. *(procjenitelj msep za sve godine nastanka)*
Procjena uvjetne srednjekvadratne greške procjenitelja msep za konačne štete $\sum_i \widehat{C_{i,J}}^{POI}$

dana je s

$$\widehat{msep}_{\sum_i C_{i,J}}\left(\sum_{i=1}^{I}\widehat{C_{i,J}}^{POI}\right) = \sum_{i+j>I}\phi V(\widehat{x}_{i,j}) + \sum_{i+j>I,n+m>I}\widehat{x}_{i,j}\widehat{x}_{n,m}\Gamma_{i,j}H(\widehat{b})^{-1}\Gamma_{n,m}^{'} \tag{53}$$

Vidimo da nam za procjenu varijabilnosti rezultata treba parametar raspršenja ϕ (dok nam za procjenu μ i γ nije trebao). U tu svrhu promotrimo Pearsonove reziduale

$$R_{i,j}^{(P)}(x_{i,j}) = \frac{X_{i,j} - x_{i,j}}{\sqrt{V(x_{i,j})}} = \frac{X_{i,j} - x_{i,j}}{\sqrt{x_{i,j}}} \tag{54}$$

i primijetimo da je

$$E\left[\left(R_{i,j}^{(P)}(x_{i,j})\right)^2\right] = \phi$$

Ako sa $\widehat{R}_{i,j}^{(P)}(x_{i,j})$ označimo procijenjene Pearsonove reziduale

$$\widehat{R}_{i,j}^{(P)}(\widehat{x}_{i,j}) = \frac{X_{i,j} - \widehat{x}_{i,j}}{\sqrt{V(\widehat{x}_{i,j})}} = \frac{X_{i,j} - \widehat{x}_{i,j}}{\sqrt{\widehat{x}_{i,j}}}$$

tada ϕ možemo procijeniti na sljedeći način

$$\widehat{\phi} = \frac{\sum_{i+j\leq I}\left(\widehat{R}_{i,j}^{(P)}\right)^2}{N-p} \tag{55}$$

pri čemu je N broj realizacija u gornjem trokutu $\mathcal{D}_I$ a p broj procijenjenih parametara $I + J + 1$.

5 Bootstrap

U prethodna dva poglavlja fokusirali smo se na procjenu očekivanih kočanih šteta $C_{i,J},\ i \in \{0, ..., J\}$ i msep kao pripadnu mjeru varijabilnosti. U prijevodu to bi značilo da smo procijenili prvi i drugi moment budućih obveza. Vidjeli smo i da procjena za msep u zatvorenoj analitičkoj formi nije jednoznačno određena pa smo se koristili različitim metodama: lažnim uzorkovanjem u modelu ulančanih ljestvica odnosno očekivanim vrijednostima pojedinih izraza u raspršenom Poissonovom modelu. Međutim i dalje nemamo distribuciju konačnih šteta za koju ne trebamo ni naglašavati od kakve vrijednosti bi bila.
Bootstrap metoda je snažan alat za dobivanje informacije o distribuciji na osnovu poznatog skupa podataka.

5.1 Osnovna ideja

Neka je $z_1, ..., z_n$ realizacija slučajnog uzorka $Z_1, ..., Z_n$ iz nepoznate distribucije F

$$Z_1, ..., Z_n \quad n.j.d \sim F \tag{56}$$

Promatrajmo parametar θ iz distribucije F te pretpostavimo da postoji funkcija g takva da je njegov procjenitelj jednak $\widehat{\theta}_n = g(Z_1, ..., Z_n)$. Kao što smo u uvodnom dijelu rekli, često nas zanima distribucija procjenitelja $\widehat{\theta}_n$ ali ju je rijetko moguće dobiti. Kada bi F bila poznata mogli bismo odrediti uzoračku distribuciju od $\widehat{\theta}_n$ tako da ponavljamo uzorkovanje određen broj puta. Budući da je F nepoznata uzorkovati ćemo iz empirijske distribucije $\widehat{F}_n$ dobivene na temelju realizacija $z_1, ..., z_n$

$$Z_1^*, ..., Z_n^* \quad n.j.d \sim \widehat{F}_n \tag{57}$$

Simulirani vektor $(Z_1^*, ..., Z_n^*)$ zove se bootstrap uzorak i pomoću njega računamo novu procjenu parametra

$$\widehat{\theta}_n^* = g(Z_1^*, ..., Z_n^*) \tag{58}$$

Ponavljajući ovaj postupak određen broj puta dobije se empirijska distribucija F_n^* za $\widehat{\theta}_n^*$.

Bitno je uočiti i imati na umu prilikom korištenja neparametarske bootstrap metode da ona pretpostavlja kako realizacije $Z_1, ..., Z_n$ sadrže sve informacije na osnovu kojih se može izgraditi prikladan model. Što znači da nismo u mogućnosti generirati niti jednu novu informaciju različitu od već opaženih. Isto tako sve postojeće informacije odrazit će se na projicirane veličine.

5.2 Model ulančanih ljestvica bez pretpostavljene distribucije

U poglavlju 3 konačni cilj bio je kvantificiranje varijabilnosti modela. Prilikom procjene varijabilnosti mjerom uvjetne srednjekvadratne greške procjenitelja $\widehat{C_{i,J}}^{CL}$ (msep) došli smo do problema koji se ne može direktno analitički riješiti. Uzrok tome je da su u trenutku I procijenjeni faktori $\widehat{f}_{I-i}, ..., \widehat{f}_{J-1}$ poznati, ali *pravi* faktori $f_{I-i}, ..., f_{J-1}$ nisu. Kako bi procijenili koliko faktori $\widehat{f}_j$ fluktuiraju oko f_j koristila se vrsta uzorkovanja i to kroz dva pristupa uvjetni i bezuvjetni. U ovom poglavlju ćemo spomenute pristupe dodatno analizirati i to u svrhu dobivanja distribucije konačnih šteta.

Temelji za bootstrap metodu postavljeni su već kroz Pretpostavke modela 3.10 pa će i ovo poglavlje počivati na njima. Za početak potrebno je naći prikladne reziduale koje ćemo koristiti za konstrukciju empirijske distribucije $\widehat{F}_n$. Pojednostavljeno, trebamo standardizirati poznate podatke u gornjem trokutu $\mathcal{D}_I$ kako bi iz tog skupa mogli uzorkovati. Kandidati za takve reziduale nalaze se u formuli za individualne razvojne faktore

$$F_{i,j+1} = \frac{C_{i,j+1}}{C_{i,j}} = f_j + \sigma_j C_{i,j}^{-1/2} \varepsilon_{i,j+1} \tag{59}$$

i to su upravo $\varepsilon_{i,j}$. Promotrimo dakle sljedeće reziduale za $i + j \leq I$ i $j \geq 1$

$$\tilde{\varepsilon}_{i,j} = \frac{F_{i,j} - \widehat{f}_{j-1}}{\widehat{\sigma}_{j-1} C_{i,j-1}^{-1/2}} \tag{60}$$

Može se pokazati da ovako definirani reziduali imaju sljedeća svojstva

- $E\left[\tilde{\varepsilon}_{i,j} | \mathcal{B}_{j-1}\right] = 0$
- $Var\left(\tilde{\varepsilon}_{i,j} | \mathcal{B}_{j-1}\right) < 1$

Druga točka znači da empirijska distribucija ima premalu varijancu te će posljedično i srednjekvadratna greška procjenitelja biti podcijenjena. Zbog toga je potrebno skalirati reziduale što možemo napraviti na sljedeći način

$$Z_{i,j} = \left(\frac{I-j}{I-j-1}\right)^{1/2} \frac{F_{i,j} - \widehat{f}_{j-1}}{\widehat{\sigma}_{j-1} C_{i,j-1}^{-1/2}} \tag{61}$$

pri čemu su $\widehat{f}_j$ i $\widehat{\sigma}_j^2$ dani s (9) odnosno (12).
Reziduali $\{Z_{i,j} : i + j \leq I\}$ definiraju bootstrap distribuciju $\widehat{F}_{\mathcal{D}_I}$ iz koje zapravo uzorkujemo nezavisne jednako distribuirane

$$Z^*_{i,j} \sim \widehat{F}_{\mathcal{D}_I} \tag{62}$$

i na taj način dobijemo bootstrap uzorak $\{Z^*_{i,j} : i+j \leq I\}$. Sljedeći korak dobivanja razvojnih faktora odnosno konačnih šteta iz bootstrap uzorka nije jednoznačno određen a ovisi o tome hoćemo li koristiti uvjetnu ili bezuvjetnu transformaciju (metode odgovaraju onima opisanim u poglavlju 3).

Pristup 1 - Bezuvjetna pogreška procjene parametara

Ovim pristupom stvoriti ćemo potpuno nove trokute na način da ćemo uzorkovati $\mathcal{D}_I$ uz poznato $\mathcal{B}_0$. To je nešto različit pristup od onog opisanog u poglavlju 3 gdje se uzorkuje samo $\mathcal{D}^0_I$, ali je lakši za implementirati.
Stoga definiramo početne iznose $C^*_{i,0} = C_{i,0}$, a sve sljedeće $j \geq 1$ na način

$$C^*_{i,j} = \widehat{f}_{j-1} C^*_{i,j-1} + \widehat{\sigma}_{j-1} \sqrt{C^*_{i,j-1}} Z^*_{i,j} \tag{63}$$

Iz toga proizlazi da su bootstrap razvojni faktori dani s

$$\begin{aligned} \widehat{f}^*_j = \frac{\sum_{i=0}^{I-j-1} C^*_{i,j+1}}{\sum_{i=0}^{I-j-1} C^*_{i,j}} &= \sum_{i=0}^{I-j-1} \frac{C^*_{i,j}}{\sum_{k=0}^{I-j-1} C^*_{k,j}} F^*_{i,j+1} \\ &= \sum_{i=0}^{I-j-1} \frac{C^*_{i,j}}{\sum_{k=0}^{I-j-1} C^*_{k,j}} \left(\widehat{f}_j + \frac{\widehat{\sigma}_j}{\sqrt{C^*_{i,j}}} Z^*_{i,j+1} \right) \end{aligned} \tag{64}$$

Pristup 2 - Uvjetna pogreška procjene parametara

Ovim pak pristupom proizvodimo nove podatke samo za sljedeći vremenski trenutak promatranih nizova. To znači da stalno djelujemo uvjetno na $\mathcal{D}_I$. Iz te perspektive razvojni faktori dani su s

$$\widehat{f}^*_j = \sum_{i=0}^{I-j-1} \frac{C_{i,j}}{\sum_{k=0}^{I-j-1} C_{k,j}} F^*_{i,j+1} = \sum_{i=0}^{I-j-1} \frac{C_{i,j}}{\sum_{k=0}^{I-j-1} C_{k,j}} \left(\widehat{f}_j + \frac{\widehat{\sigma}_j}{\sqrt{C_{i,j}}} Z^*_{i,j+1} \right) \tag{65}$$

U oba pristupa konačni razvoj šteta dobijemo na način:

$$\widehat{C_{i,J}}^* = C_{i,I-i} \prod_{j=I-i}^{J-1} \widehat{f}^*_j \tag{66}$$

Razlika je u procjeni $\widehat{f}^*_j$.

5.3 Raspršeni Poissonov model

U poglavlju 4 analizirali smo raspršeni Poissonov model sa sljedećim pretpostavkama

$$E[X_{i,j}] = x_{i,j} = \mu_i \gamma_j$$
$$Var[X_{i,j}] = \phi x_{i,j} = \phi \mu_i \gamma_j$$

Parametre μ_i i γ_j procijenili smo metodom maksimalne vjerodostojnosti. Procjena za srednjekvadratnu pogrešku točnije za procjenu pogreške parametara dana je uz pomoć Fisherove matrice. Sada ćemo, kao i u prethodnom poglavlju, iskoritit bootstrap metodu kako bi sagledali procjenu pogreške iz drugog kuta.
Za početak potrebno je pronaći jednako distribuirane reziduale koji će konstruirati empirijsku distribuciju. Prisjetimo se Pearsonovih reziduala koje smo na kraju poglavlja 4 korstili za procjenu ϕ

$$R_{i,j}^{(P)}(x_{i,j}) = \frac{X_{i,j} - x_{i,j}}{\sqrt{V(x_{i,j})}} = \frac{X_{i,j} - x_{i,j}}{\sqrt{x_{i,j}}}$$

Promatrane reziduale potrebno je standardizirati odnsono podijeliti sa pripadnom varijancom ϕ. Dodatno, može se pokazati da je takvom uzorku varijanca premala u istom smislu kao u prethodnom poglavlju - dobiveni bootstrap rezultati ne bi bili usporedivi sa analitičkim modelom. Prema tome, potrebno ih je dodatno skalirati

$$Z_{i,j} = \frac{X_{i,j} - \widehat{x}_{i,j}}{\sqrt{\phi \widehat{x}_{i,j}}} \left(\frac{N}{N-p} \right)^{1/2} \quad za \quad i + j \leq I \tag{67}$$

Ovako definirani reziduali $\{Z_{i,j} : i + j \leq I\}$ čine bootstrap distribuciju $\widehat{F}_{\mathcal{D}_I}$. Iz iste uzorkujemo nezavisne jednako distribuirane reziduale

$$Z_{i,j}^* \sim \widehat{F}_{\mathcal{D}_I}$$

pomoću kojih dobijemo bootstrap vrijednosti

$$X_{i,j}^* = \widehat{x}_{i,j} + \sqrt{\phi \widehat{x}_{i,j}} Z_{i,j}^* \tag{68}$$

Bootstrap uzorak $X_{i,j}^*$ zapravo čini gornji trokut $\mathcal{D}_I^* = \{X_{i,j}^* : i + j \leq I\}$. Za svaki takav uzorkovani trokut ponavljamo postupak iz poglavlja 4 - procjena μ_i^* i γ_j^* pomoću metode maksimalne vjerodostojnosti iz bootstrap uzorka $X_{i,j}^*$, procjena $x_{i,j}^*$ pomoću μ_i^* i γ_j^* što nas zapravo vodi do krajnjeg rezultata a to je pričuva šteta $\widehat{X_{i,j}^*}^{POI}$ za $i + j > I$. Ponavljajući ovaj postupak dobijemo bootstrap distribuciju pričuve šteta, uvjetno na $\mathcal{D}_I$.

6 Analiza rezultata

U prethodnom dijelu rada detaljno su opisane stohastičke metode ali s teorijskog aspekta. Sada nam je cilj razrađene metode primijeniti na stvarne podatke te interpretirati dobivene rezultate i međusobno ih usporediti. Bitno je naglasiti da svrha ovog dijela nije pronaći najprikladniji model već razumijevanje pretpostavki modela i pravilna interpretacija dobivenih stohastičkih veličina.

Za početak nekoliko napomena o samim podacima. Svi prethodno opisani modeli ovise o kvaliteti podataka, što znači da će se sve pravilnosti odnosno nepravilnosti u podacima direktno odraziti na rezultate. Isto tako, sve što se ne nalazi u početnom trokutu model očekivano ne može ni predvidjeti. Dakle, prije upuštanja u modeliranje, potrebno je razumjeti što nam podaci imaju za reći. Osnovna područja koja se trebaju razmotriti su sljedeća:

- Izvor. Odakle podaci dolaze. Kako se njima upravlja i koliko su se procesi mijenjali tijekom vremena.
- Kvaliteta. Koliko su određeni podaci točni; tko ih unosi, koliko se mogu mijenjati, koliko su se mijenjali tijekom vremena. Tko se može njima koristiti i koliko su zaštićeni.
- Kvantiteta. Koliko podataka je potrebno za pouzdane rezultate.
- Forma. U kojoj formi ih koristiti kako bi bili prikladni za model i rezultate kojima se teži.

Prirodni nastavak osluškivanja podataka su outlieri. Pojavljuju se uslijed velikih i katastrofalnih šteta kao što su prirodne katastrofe ili događaji uzrokovani ljudskim faktorom. Pošto imaju tendenciju uvelike utjecati na rezultate treba razmotriti je li taj utjecaj opravdan ili ne. Generalni prijedlog bi bio modelirati podatke sa i bez ouliera kako bi se moglo kvantificirati koliko utječu na rezultat. Svako izbacivanje ili ostavljanje outliera trebalo bi se argumentirano dokumentirati.

Sve metode obrađene u prethodnim primjenjujemo na isti skup podataka kojeg ćemo od sada zvati trokut podataka a njegove inkrementalne vrijednosti prikazane su na slici 2. Osnovne karakteristike odabranog trokuta su sljedeće:

- I dalje vrijedi pretpostavka sa početke rada da se trokut sastoji od jednakog broja godina nastanka i godina razvoja: I = J,
- Obuhvaća razdoblje od 10 godina ($I = J = 0, \dots, 9$),
- Trokut sadrži iznose isplaćenih šteta,

- Nema negativnih inkrementalnih iznosa – to nam je bitno za raspršeni Poissonov model.

Prve dvije točke pretpostavljaju kako je desetogodišnji period dovoljan za potpuni razvoj šteta. To pak znači da se nećemo baviti modeliranjem tzv repa koji bi nam omogućio razvoj i nakon godine J.

U prvom dijelu rada vidjeli smo da svi modeli počivaju na određenim pretpostavkama. Stoga bi bilo razumno prije primjene samog modela na odabrani skup podataka testirati jesu li one zadovoljene. U slučaju da nisu, potrebno je revidirati model/podatke ovisno o dijelu koji nije zadovoljen ili razmotriti alternativni model. U nastavku ćemo za model ulančanih ljestvica i raspršeni Poissonov model prvo testirati pretpostavke a zatim analizirati rezultate dok će nam za bootstrap metodu preostati samo analiza rezultata. Testiranje pretpostavki provodimo kroz:

- Grafičku analizu reziduala
- Specifične testove

Prije nego krenemo nekoliko riječi o grafičkoj analizi reziduala. Ona se sastoji od prikaza reziduala po sve tri vremenske komponente: godini razvoja, godini nastanka i kalendarskoj godini. Uz raziduale, na grafovima su prikazane vrijednosti očekivanja i standardne devijacije. Takav pregled trebao bi nam ukazivati na dvije opasne nepravilnosti: postoji li trend u rezidualima i je li narušena pretpostavka o slučajnom raspršenju reziduala.

Godina nastanka	Godina razvoja									
	0	1	2	3	4	5	6	7	8	9
0	12,570,592	10,441,853	3,221,742	1,311,709	373,453	237,101	108,213	71,571	7,084	1,134
1	13,934,166	10,147,894	2,580,896	1,384,772	482,771	133,834	151,449	59,555	12,104	
2	13,684,926	10,768,899	2,629,258	1,801,456	379,570	111,550	81,316	86,027		
3	14,137,997	10,952,614	2,458,880	1,497,201	265,377	242,745	38,835			
4	13,927,680	11,333,184	2,277,728	1,209,790	266,048	190,371				
5	12,980,952	9,364,395	2,426,357	934,456	317,141					
6	12,146,686	9,707,372	2,699,208	1,221,736						
7	11,634,832	8,337,317	2,079,906							
8	12,220,827	8,232,173								
9	12,421,424									

Slika 2: Trokut inkrementalnih iznosa isplaćenih šteta

6.1 Model ulančanih ljestvica bez pretpostavljene distribucije

Kako je i rečeno, analizu otvaramo grafičkim pregledom reziduala. Obzirom da se metoda ulančanih ljestvica bazira na razvojnim faktorima, prikladno je promatrati reziduale u obliku kako je definirano formulom (60). To su isti oni koji će se koristiti i u bootstrap simulacijama, do na skaliranje, a prikazani su na slici 3. Kod godina razvoja očekivano vlada stabilnost jer smo po toj vremenskoj dimenziji krojili i standardizirali reziduale. Kad dobivene reziduale presložimo po godinama nastanka veća odstupanja pojavljuju se kod novijih godina nastanka koje počivaju tek na nekoliko podataka pa nas zbog toga ne zabrinjavaju. Kalendarske godine najviše nepravilnosti bilježe kod starijih godina, međutim ono što je bitno da ne ukazuju na nikakav trend niti pristranosti u raspršenju.

Slijedi testiranje temeljne dvije pretpostavke Modela ulančanih ljestvica bez pretpostavljene distribucije (Pretpostavke modela 3.1):

P1. Kumulativne štete $C_{i,j}$ za različite godine nastanka su nezavisne

P2. Postoje razvojni faktori $f_0, ..., f_{J-1} > 0$ takvi da za svaki $0 \leq i \leq I$ i za sve $0 \leq j \leq J$ vrijedi

$$E[C_{i,j}|C_{i,0}, ..., C_{i,j-1}] = E[C_{i,j}|C_{i,j-1}] = C_{i,j-1}f_{j-1}$$

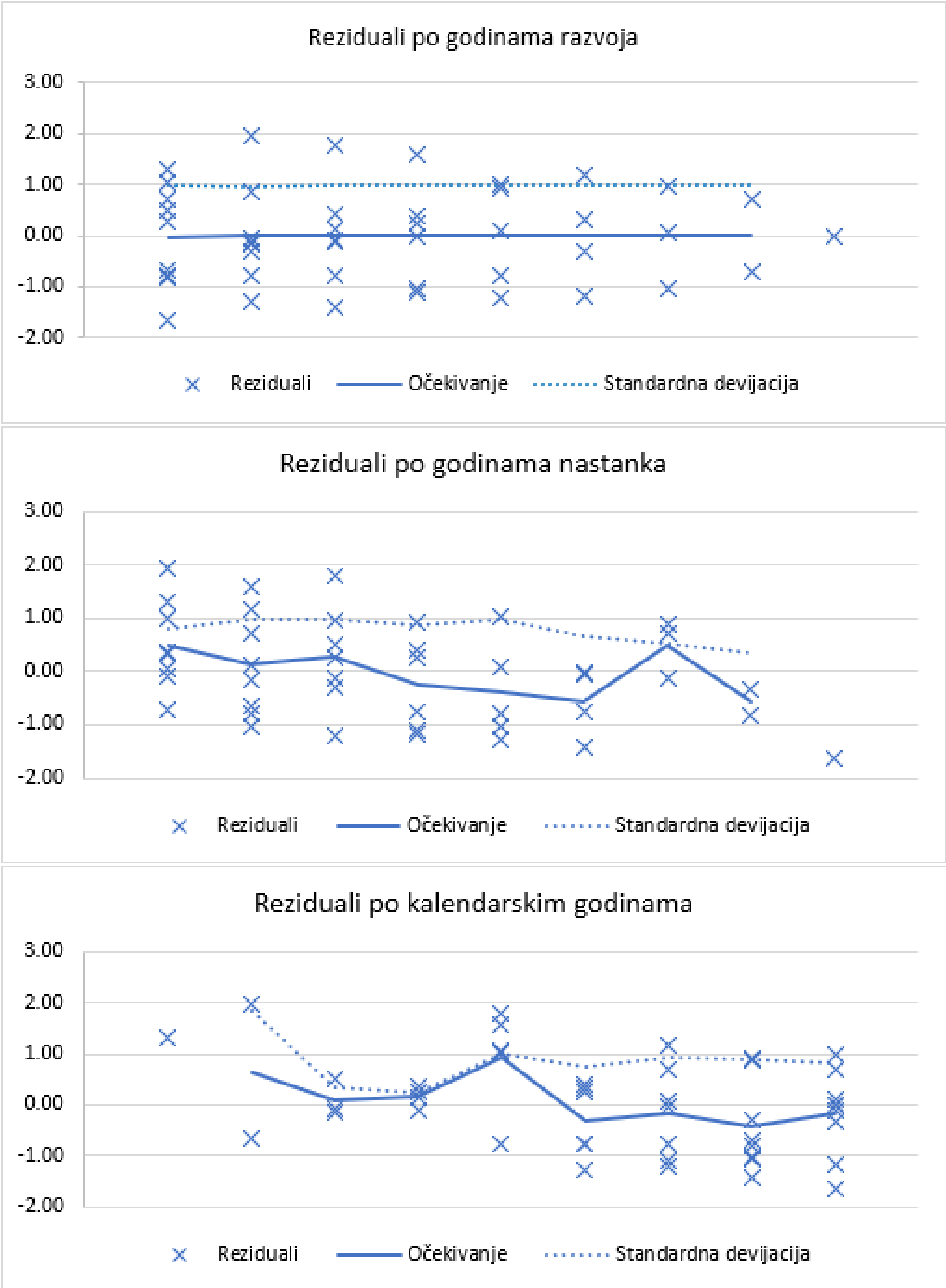
Obzirom da su oba testa koja ćemo provesti preuzeta iz [5] Mack T. (1993), tamo se mogu pronaći detaljnija objašnjenja.

Nezavisnost šteta po godinama nastanka Glavni uzrok narušavanja nezavisnosti među godinama nastanka jest postojanje trenda unutar kalendarskih godina. Takav trend može nastati zbog:

- Promjene u uvjetima i glavnim značajkama proizvoda
- Promjene u brzini prijave, obrade ili rješavanja šteta
- Promjene u pravnom okruženju u kojem Društvo djeluje
- Izuzetno velike ili male štete

Ideja u pozadini testa je sljedeća

1. Za svaki razvojni period $k = 0, \ldots, 8$ označiti individualne razvojne faktore koji su veći odnosno manji od medijana tog perioda.



Slika 3: Reziduali modela ulančanih ljestvica

2. Promotriti tako označene razvojne faktore po dijagonalama koje predstavljaju kalendarske godine. U slučaju da ne postoji trend, svaka dijagonala bi trebala imati približno jednak broj većih i manjih faktora. Kako bi razumjeli zašto, pomotrimo ekstremni slučaj da su svi razvojni faktori jedne kalendarske godine veći od pripadnih medijana. To bi gotovo sigurno indiciralo na jedan od gore navednih uzroka narušavanja nezavisnosti npr promjena pravnog okruženja u smislu povećnja minimalne odštete za određenu vrstu šteta. U skladu s tim se sve pričuve aktivnih šteta trebaju povećati što zapravo znači da nam je narušena nezavisnost među godinama nastanka.

Test je proveden prema [5] Mack T. (1993) i dobivene su sljedeće vrijednosti

$$\begin{gathered} Testna\ statistika\ Z = 13 \\ 95\%\ pouzdani\ interval = [9.00, 16.50] \end{gathered}$$

Na temelju toga zaključujemo da uz značajnost od 5% ne odbacujemo nul hipotezu odnosno ne postoji značajan utjecaj kalendarskih godina.

Nekoreliranost razvojnih faktora Rezultat Leme 3.3 pod c) konstatira nekoreliranost procijenjenih razvojnih faktora $\widehat{f}_0, ..., \widehat{f}_{J-1}$. U osvrtu na lemu komentirali smo kako je to na prvu začuđujući rezultat obzirom da susjedni faktori ovise o istim brojkama. Kasnije smo iznijeli rezultat (Lema 3.9) koji dokazuje kako su susjedni faktori, iako nekorelirani, zavisni.

Thomas Mack iskoristio je posljedicu nekoreliranosti kako bi razvio test u svrhu ispitivanja prikladnosti modela. Primijenio je zapravo neparametarski test poznat kao Spearmanov koeficijent korelacije, samo što ga je nadogradio tako da se ne promatraju susjedni razvojni faktori zasebno već trokut kao cjelina.

Kao i kod prethodnog testa, pratili smo [5] Mack T. (1993) te dobili sljedeće vrijednosti:

$$\begin{gathered} Testna\ statistika\ T = -0.086 \\ 50\%\ pouzdani\ interval = [-0.127, 0.127] \end{gathered}$$

Na temelju toga zaključujemo da uz značajnost od 50% ne odbacujemo nul hipotezu odnosno ne postoji korelacija među razvojnim faktorima.

Ovdje treba naglasiti da je korištena 50%-tna pouzdanost uz objašnjenje da je test prilično aproksimativne prirode, a mi bi željeli detektirati korelacije i u manjim mjerama zbog čega su postrožene granice pozdanog intervala.

Rezultati Analizu rezultata počinjemo pregledom osnovnih sastavnica modela. Na slici 4 nalazi se gornji trokut kumulativnih šteta $\mathcal{D}_I$ na osnovu kojeg su procijenjene štete donjeg trokuta $\mathcal{D}_I^c$ prema formuli (6).

Godina nastanka	Godina razvoja										CL pričuva šteta
	0	1	2	3	4	5	6	7	8	9	
0	12,570,592	23,012,445	26,234,187	27,545,897	27,919,350	28,156,452	28,264,664	28,336,235	28,343,319	28,344,453	0
1	13,934,166	24,082,059	26,662,955	28,047,728	28,530,499	28,664,333	28,815,781	28,875,336	28,887,440	28,888,596	1,155
2	13,684,926	24,453,825	27,083,084	28,884,540	29,264,110	29,375,661	29,456,977	29,543,003	29,552,912	29,554,094	11,091
3	14,137,997	25,090,611	27,549,491	29,046,691	29,312,068	29,554,813	29,593,648	29,667,908	29,677,859	29,679,046	85,398
4	13,927,680	25,260,864	27,538,592	28,748,381	29,014,430	29,204,801	29,300,630	29,374,155	29,384,007	29,385,182	180,382
5	12,980,952	22,345,347	24,771,704	25,706,160	26,023,301	26,188,720	26,274,652	26,340,584	26,349,419	26,350,473	327,172
6	12,146,686	21,854,058	24,553,266	25,775,002	26,094,830	26,260,703	26,346,872	26,412,985	26,421,844	26,422,901	647,898
7	11,634,832	19,972,149	22,052,055	23,171,575	23,459,098	23,608,217	23,685,682	23,745,118	23,753,082	23,754,032	1,701,977
8	12,220,827	20,453,000	22,692,511	23,844,545	24,140,418	24,293,868	24,373,583	24,434,745	24,442,940	24,443,918	3,990,918
9	12,421,424	21,881,235	24,277,132	25,509,613	25,826,147	25,990,312	26,075,594	26,141,026	26,149,794	26,150,840	13,729,416
Ukupno											20,675,406
$\widehat{f}_j$	1.7616	1.1095	1.0508	1.0124	1.0064	1.0033	1.0025	1.0003	1.0000		
$\widehat{\sigma}_j$	187.34	74.90	45.91	16.10	11.25	9.05	2.31	0.64	0.18		

Slika 4: Kumulativni trokut isplaćenih šteta sa procijenjenim parametrima modela ulančanih ljestvica

U zadnjem stupcu izračunata je očekivana pričuva šteta kao razlika $\widehat{C_{i,J}}^{CL}$ i $C_{i,I-i}$, $0 < i < I$. Za najstariju godinu nastanka ona iznosi 0 jer je pretpostavka da su se štete za tu godinu potpuno razvile ($I = J$). Pričuva raste od starijih prema novijim godinama nastanka jer su novije godine manje razvijene, što je i očekivano.

U prvom retku ispod trokuta nalaze se razvojni faktori korišteni za procjenu donjeg trokuta, definirani formulom (9). Karakterizira ih nekoreliranost ali ne i nezavisnost, što smo prethodno djelomično i testirali. Ako pogledamo dobivene vrijednosti razvojnih faktora možemo primijetiti da one padaju što bi se moglo proglasiti generalnim pravilom obzirom da kumulativne štete rastu ali padajućim trendom (do na iznimne događaje). U drugom retku ispod trokuta nalaze se procijenjeni parametri varijance $\widehat{\sigma}_j$ definirani formulom (12). σ_8 procijenjen je formulom

$$\widehat{\sigma}_8^2 = min\{\widehat{\sigma}_7^4/\widehat{\sigma}_6^2; \widehat{\sigma}_6^2; \widehat{\sigma}_7^2\}$$

Silazni trend parametara varijance može se interpretirati na sljedeći način: težinski prosjek odstupanja individualnih razvojnih faktora od procijenjenog $\widehat{f}_j$ pada u apsolutnom iznosu - što je u skladu sa očekivanom stabilizacijom kumulativnog iznosa šteta. Da ponovimo - prikazani parametri varijance su uvjetno i bezuvjetno nepristrani procjenitelji za σ_j^2, pod Pretpostavkama modela 3.10.

Ono čemu cijelo vrijeme težimo jest procjena varijabilnosti dobivenih rezultata. Prije nego krenemo kratka napomena: na slici 4 su sve komponente koje se koriste pri procjeni varijabilnosti: $C_{i,j}$, $\widehat{C_{i,J}}^{CL}$, $\widehat{f}_j$ i $\widehat{\sigma}_j^2$.

Godina nastanka	$\widehat{C_{i,J}^{CL}}$	Pričuva šteta	$\widehat{msep}_{C_{i,J}\mid D_I}(\widehat{C_{i,J}^{CL}})^{1/2}$		$\widehat{Var}(C_{i,J}\mid D_I)^{1/2}$		$\widehat{Var}(\widehat{C_{i,J}^{CL}}\mid D_I)^{1/2}$	
0	28,344,453	0	0	0.0%	0	0%	0	0%
1	28,888,596	1,155	1,356	117.3%	954	50%	963	50%
2	29,554,094	11,091	4,499	40.6%	3,609	64%	2,686	36%
3	29,679,046	85,398	15,223	17.8%	13,061	74%	7,820	26%
4	29,385,182	180,382	56,959	31.6%	50,762	79%	25,835	21%
5	26,350,473	327,172	82,347	25.2%	75,116	83%	33,744	17%
6	26,422,901	647,898	121,267	18.7%	111,846	85%	46,864	15%
7	23,754,032	1,701,977	260,231	15.3%	245,158	89%	87,280	11%
8	24,443,918	3,990,918	466,587	11.7%	441,589	90%	150,676	10%
9	26,150,840	13,729,416	960,847	7.0%	911,740	90%	303,244	10%
Ukupno	272,973,533	20,675,406	1,158,558	5.6%	1,052,277	82%	484,738	18%

Slika 5: Rezultati modela ulančanih ljestvica bez pretpostavljene distribucije

Na slici 5 prikazani su redom: konačne štete $\widehat{C_{i,J}}^{CL}$, pričuva koju smo već komentirali a prikazana je i na slici 4. Zatim uvjetna srednjekvadratna greška procjenitelja dobivena formulom (30) uz pripadni postotak udjela u pričuvi šteta, uvjetna varijanca procesa (18) i uvjetna pogreška procjene parametara (29), obje prikazane i kao udio u ukupnoj srednjekvadratnoj grešci. Sve navedene formule odnose se samo za pojedinačne godine nastanka. Kako smo već i komentirali u teorijskom dijelu, za sve godine nastanka zajedno uvjetna varijanca procesa jednaka je zbroju pojedinačnih (zbog pretpostavke o nezavisnosti godina nastanka), dok isto ne vrijedi za uvjetnu pogrešku procjene parametara. Ukupna varijabilnost izračunata je po formuli (31).

Pogledajmo prvo apsolutne vrijednosti koje nam opisuju varijabilnost. U sva tri pokazatelja vidimo da od starih prema novijim godinama iznos varijabilnosti raste što bi i očekivali obzirom da bi trebala pratiti volumen samih pričuva. Također, iznos varijabilnosti za sve godine zajedno veći je od svih pojedinačnih godina. S druge strane omjer varijabilnosti i pričuve u pravilu pada od starijih prema novijim godinama što bi se interpretiralo na način da je kod starijih godina nastanka ostao jedan ili nekoliko perioda isplata te se sva neizvjesnost očituje kroz pokazatelje varijabilnosti, dok se u novijim godinama te isplate mogu premjestiti iz jedne godine razvoja u drugu. Iznimka je 3 godina nastnka kojoj su relativni pokazatelji varijabilnosti manji od susjednih godina nastanka. Međutim to su apsolutno male razlike stoga je ta nepravilnost u ovom slučaju zanemariva. Struktura srednjekvadratne greške kreće od jednakog omjera varijabilnosti procesa i pogreške procjene parametara te se penje do 90% udjela u korist varijabilnosti procesa. To znači da u modelu ulančanih ljestvica većina varijabilnosti novijih godina proizlazi iz neizvjesnosti budućih dagađaja u smislu varijabilnosti samog procesa. Sada ćemo vidjeti kako stvari stoje kod Poissonovog modela.

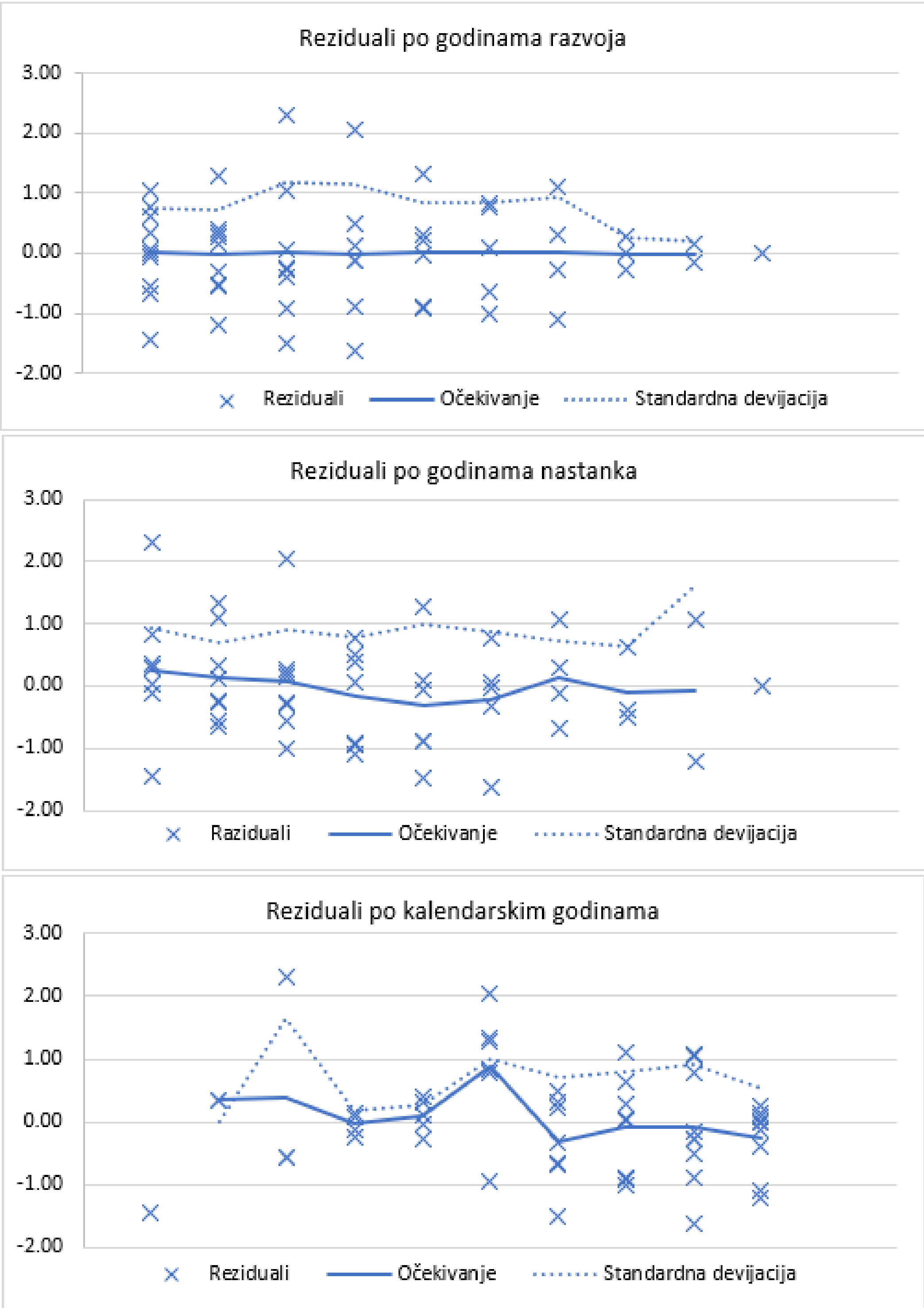
6.2 Raspršeni Poissonov model

Na slici 6 prikazani su reziduali po sve tri vremenske komponente a dobiveni formulom:

$$Z_{i,j} = \frac{X_{i,j} - \widehat{x}_{i,j}}{\sqrt{\phi \widehat{x}_{i,j}}}$$

Primijetimo da je to ista formula kao (67) odnosno isti reziduali koji će se koristiti za bootstrap simulacije, do na skaliranje. Godine razvoja pokazuju blagu nestabilnost, pogotovo u zadnja dva razdoblja ali na tom dijelu su pokazatelji osjetljiviji obzirom da ovise o nekoliko podatka. Sada je pravi trenutak da prokomentiramo parametar raspršenja ϕ.

Za potrebe ovog rada pretpostavili smo da je konstatan iako teoretski model dozvoljava varirajući parametar raspršenja po svim (i, j). Ako sada ponovno pogledamo graf rezidala po npr. godinama razvoja moglo bi se reći da je prisutna blaga heteroskedastičnost jer su 2 i 3 godina raspršenije od ostatka, što bi



Slika 6: Reziduali raspršenog Poissonovog modela

upućivalo da je za njih možda potreban različit parametar raspršenja. Ipak, treba biti oprezan kod uvođenja više parametara raspršenja jer posljedično može doći do preparametrizacije modela.

Po godinama nastanka uočavamo pravilnije ponašanje nego kod modela ulančanih ljestvica iz čega bi se dalo naslutiti da je ovo prikladniji model za promatrane podatke, iako razlike nisu toliko značajne da bi se jedan odbacio. Kalendarske godine na ukazuju na nikakav trend, što potvrđuje testove iz prethodnog modela.

Na sličan način kao i prethodne, obradit ćemo rezultate Poissonovog modela, do na njegove specifičnosti.

Tako se na slici 7 nalazi gornji trokut inkrementalnih vrijednosti šteta sa pripadnom procjenom donjeg trokuta koristeći Poissonov model: prema pretpostavkama modela 4.3. postoje parametri μ_i i γ_j takvi da $X_{i,j}$ pripada raspršenoj Poissonovoj distribuciji i vrijedi $X_{i,j} = \mu_i \gamma_j$ za $0 \leq i \leq I$ i $0 \leq j \leq J$. U zadnjem stupcu nalaze se procijenjene vrijednosti $\widehat{\mu}_i$, dok se ispod trokuta nalaze $\widehat{\gamma}_i$. Prisjetimo se da model ulančanih ljestvica koristi kumulativne štete dok Poissonov model operira na inkrementalnim podacima.

Poissonov model daje identične rezultate kao i model ulančanih ljestvica u smislu očekivanih konačnih šteta pa tako i pričuva. Ta tvrdnja se može i dokazati a slijedi iz definicija modela. Ono gdje očekujemo da će se modeli razlikovati jesu drugi i viši momenti. U tu svrhu pogledajmo varijabilnosti proizašle iz Poissonovog modela. Kretanja apsolutnih iznosa varijabilnosti kao i relativnih u skladu su sa opisanim pravilnostima kod DFM metode. Ono gdje se rezultati značajnije razlikuju je:

- Udio srednjekvadratne greške procjenitelja u pričuvi za starije godine. Uzrok tome je različita definicija modela. Iako relativno izgledaju veliko, na ukupnom su to zanemarive razlike.

- Udio varijance procesa i procjene pogreške parametara u ukupnoj srednjekvadratnoj grešci. To proizlazi prvenstveno iz različitih definicija modela zbog čega te veličine nisu ni usporedive. No zanimljivo je vidjeti kako je udio varijabilnosti samog procesa od 50% narastao do 82% u 7 godini nakon čega pada na 43% za razliku od prethodnog modela.

U obje navedene točke tjecaj ima i izbor konstantnog parametra raspršenja umjesto varirajućeg. Iako to otvara prostor za daljnje analize, ovdje se nećemo upuštati u iste.

Kao što je već spomenuto u poglavlju 4, za potrebe procjene μ_i i γ_j nije potrebna informacija o parametru raspršenja ϕ dok za procjenu varijabilnosti je. On je dobiven prema formuli (55) i iznosi

$$\phi = 28.606$$

Godina nastanka	Godina razvoja 0	1	2	3	4	5	6	7	8	9	POI pričuva šteta	$\widehat{\mu}_i$
0	12,570,592	10,441,853	3,221,742	1,311,709	373,453	237,101	108,213	71,571	7,084	1,134	0	*1.00*
1	13,934,166	10,147,894	2,580,896	1,384,772	482,771	133,834	151,449	59,555	12,104	1,155	1,155	*1.02*
2	13,684,926	10,768,899	2,629,258	1,801,456	379,570	111,550	81,316	86,027	9,908	1,182	11,091	*1.04*
3	14,137,997	10,952,614	2,458,880	1,497,201	265,377	242,745	38,835	74,261	9,950	1,187	85,398	*1.05*
4	13,927,680	11,333,184	2,277,728	1,209,790	266,048	190,371	95,829	73,525	9,852	1,175	180,382	*1.04*
5	12,980,952	9,364,395	2,426,357	934,456	317,141	165,419	85,932	65,932	8,834	1,054	327,172	*0.93*
6	12,146,686	9,707,372	2,699,208	1,221,736	319,827	165,873	86,169	66,113	8,859	1,057	647,898	*0.93*
7	11,634,832	8,337,317	2,079,906	1,119,520	287,523	149,119	77,465	59,436	7,964	950	1,701,977	*0.84*
8	12,220,827	8,232,173	2,239,511	1,152,034	295,873	153,450	79,715	61,162	8,195	978	3,990,918	*0.86*
9	12,421,424	9,459,811	2,395,897	1,232,481	316,534	164,165	85,281	65,433	8,767	1,046	13,729,416	*0.92*
Ukupno											20,675,406	
$\widehat{\gamma}_j$	*13,463,372*	*10,253,329*	*2,596,872*	*1,335,865*	*343,086*	*177,936*	*92,435*	*70,921*	*9,503*	*1,134*		

Slika 7: Inkrementalni trokut isplaćenih šteta sa procijenjenim parametrima raspršenog Poissonovog modela

Godina nastanka	$\widehat{C_{i,J}^{POI}}$	Pričuva šteta	$\widehat{msep}_{C_{i,J}\|D_I}(\widehat{C_{i,J}^{POI}})^{1/2}$		$\widehat{Var}(C_{i,J}\|D_I)^{1/2}$		$\widehat{Var}\left(\widehat{C_{i,J}^{POI}}\middle\|D_I\right)^{1/2}$	
0	28,344,453	0	0	0.0%	0	*0%*	0	*0%*
1	28,888,596	1,155	8,170	707.0%	5,749	*50%*	5,804	*50%*
2	29,554,094	11,091	22,340	201.4%	17,812	*64%*	13,484	*36%*
3	29,679,046	85,398	57,983	67.9%	49,426	*73%*	30,317	*27%*
4	29,385,182	180,382	82,365	45.7%	71,834	*76%*	40,298	*24%*
5	26,350,473	327,172	107,834	33.0%	96,743	*80%*	47,633	*20%*
6	26,422,901	647,898	150,400	23.2%	136,140	*82%*	63,923	*18%*
7	23,754,032	1,701,977	244,132	14.3%	220,653	*82%*	104,465	*18%*
8	24,443,918	3,990,918	391,261	9.8%	337,885	*75%*	197,280	*25%*
9	26,150,840	13,729,416	958,147	7.0%	626,698	*43%*	724,773	*57%*
Ukupno	272,973,533	20,675,406	1,142,276	5.5%	769,058	*45%*	844,597	*55%*

Slika 8: Rezultati rasprešnog Poissonovog modela

6.3 Bootstrap - Model ulančanih ljestvica bez pretpostavljene distribucije

Bootstrap metoda potpuno se oslanja na reziduale, a one koji se koriste za uzorkovanje u modelu ulančanih ljestvica analizirali smo u poglavlju 6.1, do na skaliranje. Sada ćemo vidjeti kako su se informacije sadržane u gornjem trokutu $\mathcal{D}_I$ odrazile na rezultate kroz proces većeg broja simulacija.

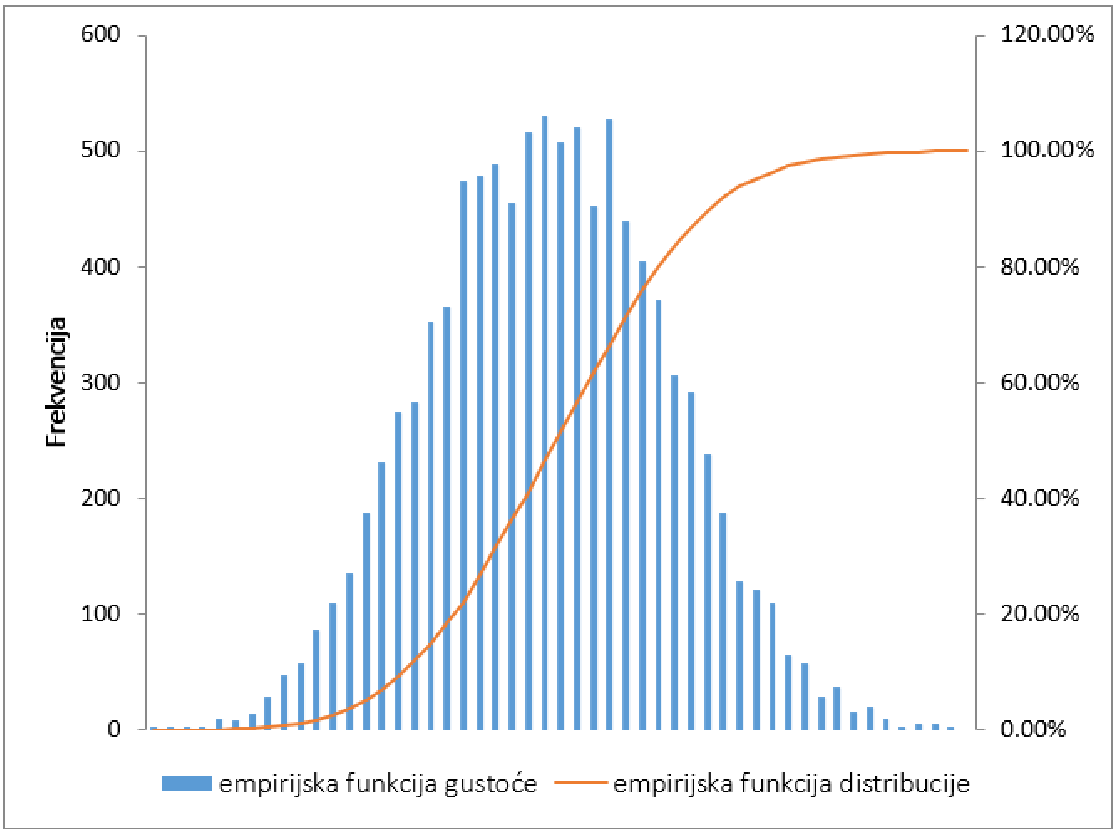
Broj simulacija: 10000.

Bitno za naglasiti prije samih rezultata: korišten je Pristup 2 - Uvjetna pogreška procjene parametara iz Poglavlja 5.2.

Godina nastanka	$\widehat{C_{i,J}^{CL}}$	Pričuva šteta	$\widehat{msep}_{C_{i,J}\|D_I}(\widehat{C_{i,J}^{CL}})^{1/2}$		$\widehat{Var}(C_{i,J}\|D_I)^{1/2}$		$\widehat{Var}\left(\widehat{C_{i,J}^{CL}}\middle\|D_I\right)^{1/2}$	
0	28,344,453	0	0	0.0%	0	*0%*	0	*0%*
1	28,888,584	1,143	4,885	427.2%	3,421	*49%*	3,487	*51%*
2	29,554,121	11,117	6,585	59.2%	4,955	*57%*	4,338	*43%*
3	29,678,840	85,192	16,020	18.8%	13,544	*71%*	8,555	*29%*
4	29,385,194	180,393	56,988	31.6%	50,661	*79%*	26,097	*21%*
5	26,348,497	325,196	83,412	25.6%	76,169	*83%*	33,997	*17%*
6	26,423,801	648,799	121,590	18.7%	112,096	*85%*	47,102	*15%*
7	23,751,940	1,699,885	258,889	15.2%	243,606	*89%*	87,635	*11%*
8	24,440,900	3,987,901	468,098	11.7%	443,366	*90%*	150,144	*10%*
9	26,155,841	13,734,416	964,203	7.0%	914,904	*90%*	304,364	*10%*
Ukupno	272,972,171	20,676,111	1,160,760	5.6%	1,054,011	*82%*	486,235	*18%*

Slika 9: Rezultati dobiveni bootstrap metodom za model ulančanih ljestvica

Pričuve dobivene bootstrap metodom neznačajno su različite od onih dobivenih analitičkom formom modela. Iznosi pričuva su skoro jednaki po svim godinama nastanka. Isto vrijedi i za pokazatelje varijabilnosti. Ono gdje se jedino razlikuju jest varijabilnost za stare godine. Međutim to je i za očekivati jer kod velikog broja simulacija gotovo sigurno će u nekima od njih dospjeti "veći" rezidual koji će im povećati varijabilnost. A kao i što smo već spomenuli za stare godine, obzirom da je ostao jedan ili nekoliko razvojnih perioda varijabilnost se nema gdje anulirati.



Slika 10: Empirijska distribucija bootstrap metode za model ulančanih ljestvica

Preostaje nam rezultat za koji smo nekoliko puta naglasili od kakve je vrijednosti - na slici 10 nalazi se distribucija pričuva šteta po svim godinama nastanka zajedno. Histogram je konstruiran na sljedeći način: od 10000 simulacija uzete su maksimalna i minimalna vrijednost, odredio se broj razreda (50) u koje su se potom rasporedile simulacije. Linija označava pripadnu funkciju distribucije.

Iz slike 10 bi se dalo zaključiti kako je ova distribucija normalna. Kako bi to potvrdili proveli smo Kolmongorov Smirnovljev test. Dobili smo sljedeće veličine:

$$Testna\ statistika\ T = 0.0132$$
$$95\%\ pouzdani\ interval = [-\infty, 0.0136]$$

pri čemu smo za kritičnu točku koristili procjenu $1.36/\sqrt{N}$, $N = 10000$. Dakle na temelju toga zaključujemo da uz značajnost od 5% ne odbacujemo nul hipotezu odnosno procjenjene pričuve su normalno distribuirane.

6.4 Bootstrap - Raspršeni Poissonov model

Preostali su nam rezultati bootstrap metode izgrađene na raspršenom Poissonovom modelu. Reziduale koji se koriste za uzorkovanje definirali smo u poglavlju 5.3, a njihove pretke komentirali u poglavlju 6.2.

Broj simulacija: 10000

Godina nastanka	$\widehat{C_{i,J}^{POI}}$	Pričuva šteta	$\widehat{msep_{C_{i,J}\|D_I}}(\widehat{C_{i,J}^{POI}})^{1/2}$		$\widehat{Var}(C_{i,J}\|D_I)^{1/2}$		$\widehat{Var}\left(\widehat{C_{i,J}^{POI}}\middle\|D_I\right)^{1/2}$	
0	28,344,453	0	0	0.0%	0	*0%*	0	*0%*
1	28,888,692	1,252	13,116	1047.5%	11,720	*80%*	5,887	*20%*
2	29,554,116	11,113	26,464	238.1%	22,667	*73%*	13,658	*27%*
3	29,679,092	85,444	60,137	70.4%	51,593	*74%*	30,897	*26%*
4	29,384,759	179,958	84,622	47.0%	73,960	*76%*	41,118	*24%*
5	26,350,560	327,259	111,717	34.1%	100,564	*81%*	48,656	*19%*
6	26,422,329	647,327	153,715	23.7%	139,117	*82%*	65,381	*18%*
7	23,751,524	1,699,468	247,833	14.6%	224,012	*82%*	106,019	*18%*
8	24,442,465	3,989,465	402,904	10.1%	348,593	*75%*	202,025	*25%*
9	26,163,330	13,741,906	976,922	7.1%	641,139	*43%*	737,101	*57%*
Ukupno	272,981,319	20,683,192	1,156,855	5.6%	781,901	*46%*	852,610	*54%*

Slika 11: Rezultati dobiveni bootstrap metodom za raspršeni Poissonov model

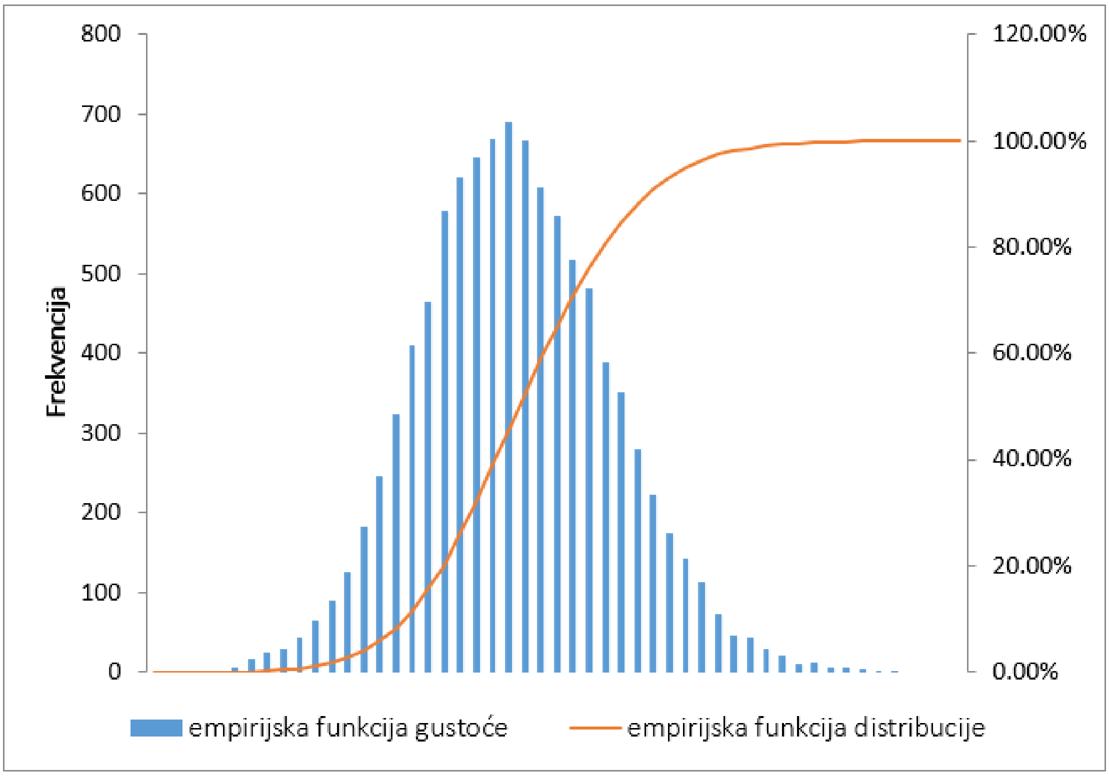
I ovdje se konačne pričuve neznačajno razlikuju od analitičkog modela. Varijabilnost je u principu slična, osim za stare godine. Tu imamo situaciju relativnog povećanja ali i drugačijeg omjera dvaju varijabilnosti u ukupnoj srednjekvadratnoj pogrešci. Komentar bi bio isti kao i u prethodnom bootstrap modelu - možemo očekivati da će u tolikom broju simulacija neki od većih reziduala dospjeti na stare godine i značajno povećati varijabilnot.

Na slici 12 nalazi se histogram frekvencija konstruiran na isti način kao i u prethodnom odjeljku. Iako bi se na prvi pogled reklo da su podaci normalno distribuirani, Kolmongorov Smirnovljev test upućuje na drugačiji zaključak:

$$Testna\ statistika\ T = 0.0207$$
$$95\%\ pouzdani\ interval = [-\infty, 0.0136]$$

pri čemu smo za kritičnu točku koristili procjenu $1.36/\sqrt{N}$, $N = 10000$. Dakle na temelju toga zaključujemo da uz značajnost od 5% odbacujemo nul hipotezu odnosno procijenjene pričuve nisu normalno distribuirane.

Mogli bismo dalje tražiti o kojoj distribuciji je riječ i zbog čega se ne radi o normalnoj, međutim ovdje nam je zanimljivo bilo vidjeti kako dvije metode na istom skupu podataka daju različite rezultate. Za procjenu raznih kvantila i pouzdanih intervala dovoljan nam je simulirani uzorak kojeg smo dobili, stoga ćemo se na ovome i zadržati.



Slika 12: Empirijska distribucija bootstrap metode za raspršeni Poissonov model

Literatura

[1] Basrak B. (2016) *Financijski praktikum: Generalizirani linearni modeli.* Predavanja, PMF-MO.

[2] Björkwall S. (2011) *Stochastic claims reserving in non-life insurance: Bootstrap and smoothing models.* Department od Mathematics, Stockholm University.

[3] Carrato A., McGuire G. and Scarth R. (2016) *A Practitioner's Introduction to Stochastic Reserving.*

[4] International Actuarial Association (2010) *Stochastic Modeling: Theory and reality from an actuarial perspective.* Treće izdanje. Ottava: International Actuarial Association.

[5] Mack T. (1993) *Measuring the Variability of Chain Ladder Reserve Estimates.* Munich Re.

[6] Taylor G and McGuire G. (2016) *Stochastic Loss Reserving Using Generalized linear Models.* Casualty Actuarial Society monograph series, number 3.

[7] Wüthrich M. V. and Merz M. (2008) *Stohastic claims reserving methods in insurance.* Chichester: John Wiley & Sons.

Sažetak

Cilj ovog rada bio je, kao što i sam naslov kaže, stohastički uokviriti dobro poznati algoritam metode ulančanih ljestvica i to kroz prizmu dva modela:

- Model ulančanih ljestvica bez pretpostavljene distribucije
- Raspršeni Poissonov model

Naime, sam algoritam je deterministički proces koji ne pruža nikakve informacije o pouzdanosti rezultata. Zbog toga ga je potrebno detaljnije opisati matematičkim rječnikom kako bi se kvantificirala pripadna neizvjesnost.

Prvi model prirodni je nastavak algoritma, koji stohastički prati njegovu osnovnu logiku. Njime se postepeno gradi put ka vremenskim nizovima pomoću kojih ćemo izmjeriti varijabilnosti uz sve nužne pretpostavke kao što je nezavisnost godina nastanka. Drugi model počiva na pretpostavci da su podaci distribuirani raspršenom Poissonovom distribucijom što je potpuno drugačiji pogled nego u Modelu ulančanih ljestvica. U pozadini pristupa je zapravo teorija generaliziranih linearnih modela uz ograničenje na Poissonovu distribuciju iz familije eksponencijalnih. Ovi modeli izabrani su za stohastički okvir zbog toga što daju identične očekivane pričuve šteta pri čemu se razlikuju u iznosu varijabilnosti odnosno razlikuju se u drugim i višim momentima. To je posljedica pretpostavki na kojima su izgrađeni. Također, oba imaju prirodnu podlogu za bootstrap metodu koja nam pruža dodatnu vrlo vrijednu informaciju - distribuciju pričuva šteta. Kako bi rezultati svih metoda bili usporedivi, za mjeru varijabilnosti korištena je uvjetna srednjekvadratna pogreška.

U drugom dijelu rada izabran je skup podataka na kojem su primijenjene sve 4 metode: Model ulančanih ljestvica bez pretpostavljene distribucije, Raspršeni Poissonov model, Bootstrap metoda modela ulančanih ljestvica bez pretpostavljene distribucije i Bootstrap metoda raspršenog Poissonovog modela. Modeli su obrađeni kroz nekoliko koraka počevši sa grafičkom analizom reziduala preko testiranja pretpostavki do zadnjeg dijela a to je prezentiranje i usporedba rezultata.

Summary

The aim of this paper was to make stochastic framework of well-known chain ladder algorithm through these two models:

- Chain ladder model without distributional assumption
- Overdispersed Poisson model

Chain ladder algorithm is deterministic process that does not provide any information about the reliability of the results. Because of that, it is necessary to describe it more detailed in terms of mathematical theory to achieve quantification of uncertainty.

First model is natural extension of algorithm because it follows its logic through stochastic modelling. It is gradually built up to time series model which will help us measure variability with all assumptions needed for that, such as independence of accident year. Second model is based on assumption that all incremental claims have overdispersed Poisson distribution which is completely different view than the Chain ladder model. In the background of this approach is the theory of generalized linear models with a restriction on the Poisson distribution from the exponential family. These models were chosen for stochastic framework because they produce identical expected claim reserve with different variability, more precisely with different second and higher moments. This is consequence of the assumptions they are built on. Also, both are suitable for extension with bootstrap method which provides us additional valuable information - the distribution of claim reserves. To make all the results comparable, we used mean square error of prediction as a measure of variability .

In the second part of paper specific data set was chosen on which all 4 models were applied: Chain ladder model without distributional assumption, Overdispersed Poisson model, Bootstrap method based on chain ladder model, Bootstrap method based on overdispersed Poisson model. The models were broke down through few stages beginning with graphical analysis of the residuals through testing specific assumptions to the last part which is presentation and comparison of the results.

Životopis

Tončica Topić rođena je u Splitu 25. kolovoza 1992. godine. Nakon završene Prirodoslovno - matematičke gimnazije u Splitu upisuje preddiplomski studij Prirodoslovno - matematičkog fakulteta Sveučilišta u Zagrebu, smjer Matematika. Godine 2016. diplomira na smjeru Financijska i poslovna matematika s temom *Bayesova statistika i procjena vrijednosti ulaganja.* Početkom 2017. godine zapošljava se u Croatia osiguranju u Sektoru za aktuaristiku gdje se do danas usavršava na području aktuaristike neživotnih osiguranja. Kao prirodni nastavak školovanja, 2018. godine upisuje Poslijediplomski specijalistički studij aktuarske matematike na Prirodoslovno-matematičkom fakultetu Sveučilišta u Zagrebu. Nakon položenih ispita na Poslijediplomskom specijalističkom studiju aktuarske matematike te dodatnih ispita u organizaciji Hrvatskog aktuarskog društva, u rujnu 2021. godine, dobiva ovlaštenje za obavljanje poslova ovlaštenog aktuara.

Od 2002. - 2018. godine Tončica se aktivno bavi sportom. Godine 2007. postaje član hrvatske mačevalačke reprezentacije, od kada je osam puta odnijela titulu državne prvakinje u kategoriji mač seniorke. Postigla je i zapažene međunarodne rezultate u svim dobnim uzrastima.