

Tartu Ülikool
Loodus-ja täppisteaduste valdkond
Matemaatika ja statistika instituut

Julius Pau
**ÜLDISTATUD LAMBDAJAOTUS: RAKENDUS
VARAKINDLUSTUSE KAHJUDELE**

Finants- ja kindlustusmatemaatika eriala

Magistritöö (30 EAP)

Juhendaja: Tõnu Kollo

Tartu 2017

Üldistatud lambdajaotus: rakendus varakindlustuse kahjudele

Magistritöö

Lühikokkuvõte Käesoleva magistritöö eesmärgiks on anda ülevaade paindlikust üldistatud lambdajaotuste perest ning uurida selle sobivust juriidiliste isikute varakindlustuse kahju suuruse modelleerimiseks. Seda eriti olukorras, kus eesmärgiks on leida sobiv tõenäosusjaotus kogu kahju suuruse varieeruvuspiirkonnale. Lisaks on toodud välja ka võrdlused mõnede kahjukindlustuses rohkem levinud jaotustega.

CERCS teaduseriala: P160 Statistika, operatsioonanalüüs, programmeerimine, finants- ja kindlustusmatemaatika

Märksõnad: üldistatud lambdajaotus, varakindlustus, raske sabaga jaotused

Generalized lambda distribution: Appliance on property insurance claims

Master thesis

Abstract The following Master thesis gives an overview of a flexible family of generalized lambda distributions and investigates its usability for modeling claim severity of commercial property insurance. Especially of interest are situations, where the aim is to find probability distribution able to model entire length of the data. In addition also comparison with several commonly used distributions in insurance practice is given.

CERCS research specialisation: P160 Statistics, operation research, programming, actuarial mathematics

Keywords: Generalized lambda distribution, property insurance, heavy tailed distributions

Sisukord

1	Sissejuhatus	5
2	Üldistatud lambda-jaotus	6
2.1	Tekkelugu	6
2.2	Definitsioon	7
2.3	Parameetrite ruum	8
2.4	Genereerimine	12
3	Jaotuse sobitamine	14
3.1	Momentide meetod	14
3.2	Kvantiilide meetod	17
3.3	Numbriline logaritmilise tõepära meetod	18
3.4	Veel üks kvantiilidel põhinev meetod	21
3.5	Võrestiku meetod (inglise keeles <i>Starship method</i>)	22
4	Andmete kirjeldus	24
4.1	Andmete päritolu	24
4.2	Andmete ülevaade	25
5	Sobitamine varakindlustuse kahjuandmetele	27
5.1	Sobitamine	28
5.2	Sobivuse kontroll	29
5.2.1	Kolmogorov-Smirnovi taasvaliku meetod	29
5.2.2	Mittesobivuse analüüs	30
5.2.3	Kvantiilide võrdlus	32

5.3	Võrdlus alternatiividega	36
5.3.1	Weibulli jaotus	36
5.3.2	Lognormaalne jaotus	37
5.3.3	Gammajaotus	38
5.3.4	Eksponentjaotus	39
5.3.5	Võrdluste kokkuvõte	40
6	Kokkuvõte	41
	Viited	42

1 Sissejuhatus

Kahju suurusele sobiva jaotuse leidmine on oluline osa kindlustuse riskide ja reserve hindamise protsessist. Tihtilugu on see ülesanne jagatud kaheks erinevaks osaks, kus ühes üritatakse leida sobivat jaotust kahjude empiirilisele jaotusele ning teises suurtele kahjudele jaotuse sabas. Selle lähenemise on tinginud asjaolu, et sageli esinevad kahjud, kus kahjusummad ei ole märkimisväärsed, on hästi modelleeritavad klassikalistes kindlustuses kasutatavate jaotustega (Weibulli jaotus, lognormaalne jaotus jt). Samuti on välja pakutud erinevaid meetodeid, mis vaatlevad ainult märkimisväärsed, ent harva esinevaid juhtumeid. Antud töös pakume üldistatud lambdajaotuse näol välja lähenemise, mis võiks potentsiaalselt katta kogu kahjusumma varieeruvuspiirkonna ning uurime antud jaotuse võimalikku sobivust juriidiliste isikute varakindlustus raske sabaga kahjude andmetele.

Töö on jagatud neljaks osaks, kus esimeses anname ülevaate üldistatud lambdajaotuse tekkelo kohta, esitame definitsiooni ja üldised omadused. Teises osas kirjeldame erinevaid meetodeid antud jaotuse parameetrite hindamiseks. Seejärel liigume edasi andmete kirjeldamise juurde, kus anname ülevaate andmete päritolust ning üldistest karakteristikutest. Töö lõppeb praktilise osaga, kus esmalt leitakse varakindlustuse andmete põhjal hinnanguid üldistatud lambdajaotuse parameetritele ning seejärel võrreldakse saadud tulemust mõnede kahjukindlustuses rohkem levinud jaotustega.

2 Üldistatud lambdajaotus

2.1 Tekkelugu

Üldistatud lambdajaotuse (inglise keeles *Generalized Lambda Distribution*, edaspidi GLD) pakkusid välja John S. Ramberg ja Bruce W. Schmeiser oma 1974. aasta artiklis *An Approximate Method for Generating Asymmetric random variables* [7]. Esialgu viidatigi antud jaotusele kui Ramberg-Schmeiseri jaotusele. Käsitletav jaotus pärineb otseselt üheparameetrilisest lambdajaotusest, mille pakkusid aastal 1947 välja Cecil Hastings, Frederick Mosteller, John W. Tukey, ja Charles P. Winsor artiklis [2] ja millele kirjanduses viidatakse kui Tukey lambdajaotusele. Kuna antud jaotuse jaotusfunktsiooni ei ole võimalik esitada analüütilisel kujul, siis defineeritakse see jaotus läbi jaotusfunktsiooni pöördfunktsiooni ehk kvantiilfunktsiooni:

$$Q(y) = Q(y, \lambda) = \frac{y^\lambda - (1 - y)^\lambda}{\lambda}, \quad (1)$$

kus $0 \leq y \leq 1$ ja $\lambda \neq 0$.

Üldistatud lambdajaotuse esimene kasutusala oli Monte Carlo simulatsioonide jaoks pseudojuhuslike arvude genereerimine asümmeetriliste juhuslike suuruste simuleerimiseks. Hiljem on üldistatud lambdajaotust tema suure paindlikkuse tõttu edukalt kasutatud väga erinevates valdkondades. Ühtedeks esimesteks elualadeks, kus GLD modelleerimisel kasutust leidis, olid ehitus ja meteoroloogia.

2.2 Defnitsioon

Üldistatud lambdajaotus on neljaparameetriline tõenäosusjaotus, mille jaotusfunktsiooni ei ole (sarnaselt Tukey lambdajaotusele) võimalik esitada analüütilisel kujul. Analüütiline esitus on aga tema kvantiilfunktsioonil, mida kasutatakse defineerimisel:

$$Q(y) = Q(y, \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \frac{y^{\lambda_3} - (1 - y)^{\lambda_4}}{\lambda_2} \quad (2)$$

kus $0 \leq y \leq 1$ ja $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \in \mathbb{R}$. Parameeter λ_1 on siin asukoha parameeter, λ_2 defineerib andmete ulatuse ning λ_3 ja λ_4 määravad jaotuse kuju. Viimase kahe parameetri täpsemat muutumispiirkonda ja mõju jaotuse kujule kirjeldame parameetruumi käsitlevas osas.

Üldistatud lambdajaotuse tihedusfunktsioon avaldub kvantiilfunktsiooni kaudu järgmisel kujul:

$$f(x) = \frac{\lambda_2}{\lambda_3 y^{\lambda_3 - 1} + \lambda_4 (1 - y)^{\lambda_4 - 1}} \quad (3)$$

kus $x = Q(y)$ ja $Q(y)$ on valemis (2) toodud kvantiilfunktsioon. Tulenevalt asjaolust, et tihedusfunktsioon on defineeritud läbi kvantiilfunktsiooni, on tihedusfunktsiooni graafiline esitamine tavapärasest erinev. Selleks et kanda joonisele konkreetse üldistatud lambdajaotuse tihedusfunktsioon, tuleb käituda järgnevalt: võtame hulga punkte y -teljel (näiteks 0,01, 0,02, ..., 0,99), leiame x väärtuse igas valitud punktis kasutades võrdust (2) ning seejärel saame võrdusest (3) leida tihedusfunktsiooni väärtuse $f(x)$. Punktid $(x, f(x))$ saame kanda graafikule ning need sujuva joonega ühendada.

Mainime siinkohal ära, et Marshall Fremier pakkus 1998. aastal artiklis [1] välja üldistatud lambdajaotuse alternatiivse parametrisatsiooni, mille puhul

on kvantiilfunktsioon järgneval kujul:

$$Q(y) = \lambda_1 + \frac{1}{\lambda_2} \left(\frac{y^{\lambda_1} - 1}{\lambda_3} - \frac{(1-y)^{\lambda_4} - 1}{\lambda_4} \right), \quad (4)$$

kus $0 \leq y \leq 1$. Kirjanduses on antud kujule viidatud kui FMKL parametri-
satsioonile, samas kui algset parametrisatsiooni tähistatakse lühendiga RS.
Peamine põhjus alternatiivse kuju väljapakumiseks oli asjaolu, et algselt
välja pakutud variandi puhul ei vasta igale parameetrite kombinatsioonile
tõenäosusjaotus, samas kui FMKL on määratud üle kogu parameeterruumi.
Keerukamast kujust tulenevalt on FMKL parametrisatsiooni puhul jällegi
jaotuse sobitamine märksa keerulisem. Antud töö raames keskendume alg-
selt välja pakutud kujule. RS parametrisatsiooniga kaasnevatest tingimustest
parameetrite väärtustele räägime lähemalt järgmises punktis.

2.3 Parameetrite ruum

Nagu eespool mainitud, pole üldistatud lambda-jaotuse tihedus RS pa-
rametrisatsiooni puhul leitav sugugi mitte iga $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ puhul. Järgneva
jaotuse käitumise kirjelduse aluseks on [4], lk. 11-37.

Tihedusfunktsiooni definitsioonist lähtuvalt saame väita, et $GLD(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$
tihedus määrab jaotuse, siis ja ainult siis, kui:

$$\frac{\lambda_2}{\lambda_3 y^{\lambda_3-1} + \lambda_4 (1-y)^{\lambda_4-1}} \geq 0 \quad (5)$$

iga $y \in [0, 1]$ korral. Sellest saame omakorda järeldada, et $GLD(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$
tihedus on määratud siis ja ainult siis, kui

$$g(y, \lambda_3, \lambda_4) = \lambda_3 y^{\lambda_3-1} + \lambda_4 (1-y)^{\lambda_4-1} \quad (6)$$

on samamärgiline (kas positiivne või negatiivne) iga $y \in [0, 1]$ korral suurusega λ_2 . See tähendab, et $GLD(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ tiheduse avaldis määrab jaotuse, kui $\lambda_2, \lambda_3, \lambda_4$ omavad kõik sama märki. Et paremini mõista, mida toodud tingimus täpsemalt tähendab, jagame (λ_3, λ_4) parameeterruumi järgnevatiks piirkondadeks:

$$R1 = \{(\lambda_3, \lambda_4) | \lambda_3 \leq -1, \lambda_4 \geq 1\}, \quad (7)$$

$$R2 = \{(\lambda_3, \lambda_4) | \lambda_3 \geq 1, \lambda_4 \leq -1\}, \quad (8)$$

$$R3 = \{(\lambda_3, \lambda_4) | \lambda_3 \geq 0, \lambda_4 \geq 0\}, \quad (9)$$

$$R4 = \{(\lambda_3, \lambda_4) | \lambda_3 \leq 0, \lambda_4 \leq 0\}, \quad (10)$$

$$R5 = \{(\lambda_3, \lambda_4) | \lambda_3 < 0, 0 < \lambda_4 < 1\}, \quad (11)$$

$$R6 = \{(\lambda_3, \lambda_4) | 0 < \lambda_3 < 1, \lambda_4 < 0\}, \quad (12)$$

$$R7 = \{(\lambda_3, \lambda_4) | -1 < \lambda_3 < 0, \lambda_4 > 1\}, \quad (13)$$

$$R8 = \{(\lambda_3, \lambda_4) | \lambda_3 > 1, -1 < \lambda_4 < 0\}. \quad (14)$$

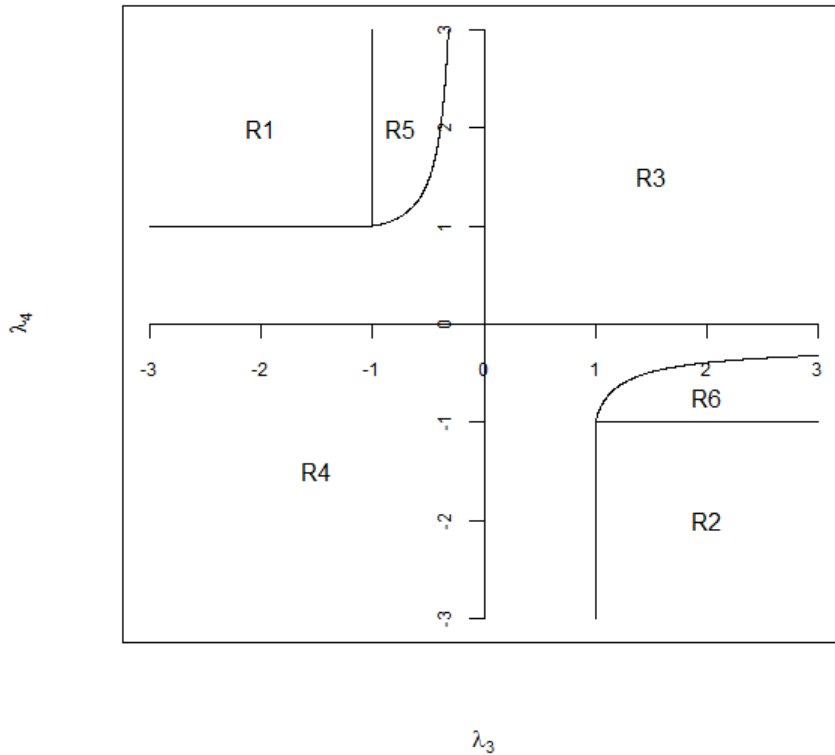
Kui neid piirkondi lähemalt uurida, siis selgub, et parameetrid piirkondadest $R1, R2, R3$ ja $R4$ määravad jaotuse eelduselt, et λ_2 on sobiva väärtusega. Teistpidi jällegi piirkondadest $R5$ ja $R6$ pärit parameetri väärtused ei võimalda mitte kunagi määrata tõenäosusetihedust. Piirkondadega $R7$ ja $R8$ on olukord keerulisem, sest selgub, et nendes leidub nii punkte, mis sobivad jaotuse määramiseks kui ka neid, mis ei ole lubatavad. Eraldavad väärtused nendes piirkondades on määratud järgmiste võrdustega:

$$\frac{(1 - \lambda_3)^{1-\lambda_3}}{(\lambda_4 - \lambda_3)^{\lambda_4 - \lambda_3}} (\lambda_4 - 1)^{\lambda_4 - 1} = \frac{-\lambda_3}{\lambda_4}, \quad (15)$$

$$\frac{(1 - \lambda_4)^{1-\lambda_4}}{(\lambda_3 - \lambda_4)^{\lambda_3 - \lambda_4}} (\lambda_3 - 1)^{\lambda_3 - 1} = \frac{-\lambda_4}{\lambda_3}. \quad (16)$$

Et olukorda paremini illustreerida, on kõik piirkonnad (7)-(14) toodud joonisel 1.

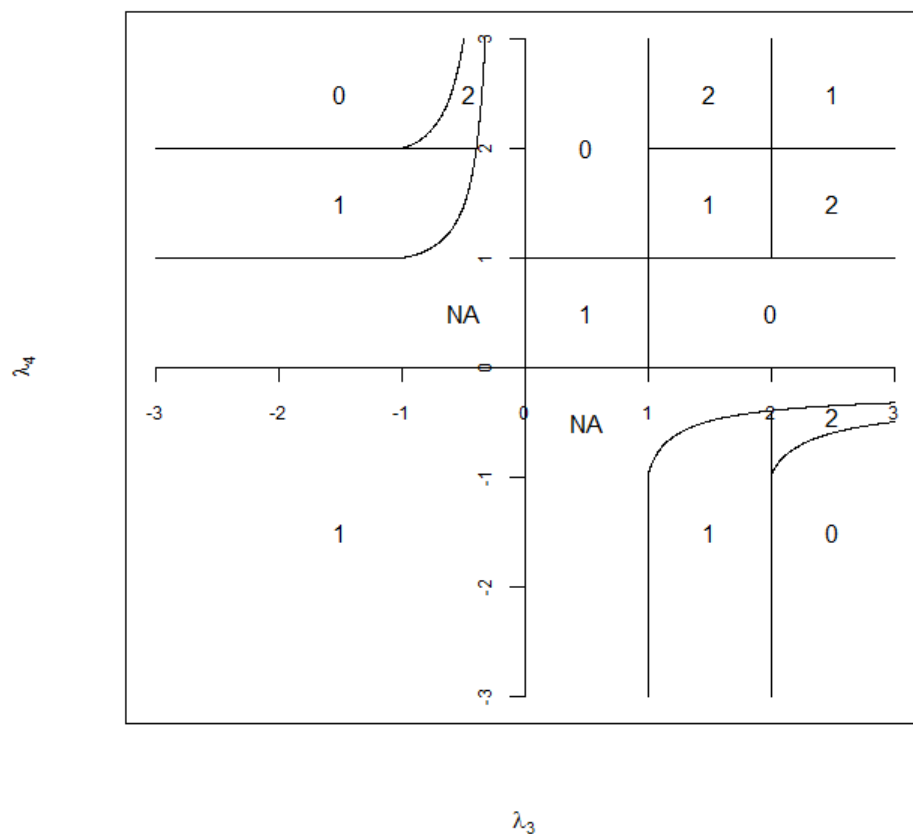
Joonis 1: Piirkonnad, kust valitud parameetrid määravad tõenäosustiheduse



Nüüd, kui oleme kirjeldanud lubatavad piirkonnad, liigume edasi tihedusfunktsiooni kuju juurde. Nagu juba mainitud, siis on üldistatud lambdajaotuse näol tegemist väga paindliku jaotusega ja seega on võimalike tihedusfunktsioonide

siooni kujude varieeruvus väga suur, kusjuures lokaalsete ekstreemumite arv võib olla nii null, üks, kui ka kaks. Arusaadavalt sõltub ekstreemumite arv λ_3 ja λ_4 valikust, kuid pole otseselt sõltuv kummagi parameetri kasvamisest või kahanemisest. Kuna iseseisvalt vaadeldavaid piirkondi on rohkem, siis on kõige otstarbekam illustreerida olukorda graafiliselt. Joonisel 2 selgub, kuidas lokaalsete ekstreemumite arv sõltub λ_3 ja λ_4 valikust

Joonis 2: Lokaalsete ekstreemumite arv piirkondades



Lisaks lokaalsete ekstreemumite arvu sõltuvusele λ_3 ja λ_4 väärtustest, on kasulik ka teada, kuidas funktsiooni positiivsete väärtustega määramispiirkond sõltub nende parameetrite väärtustest. Need seosed on toodud järgnevalt.

$$\lambda_3 > 0, \lambda_4 > 0 : [\lambda_1 - 1/\lambda_2, \lambda_1 + 1/\lambda_2], \quad (17)$$

$$\lambda_3 > 0, \lambda_4 = 0 : [\lambda_1, \lambda_1 + 1/\lambda_2], \quad (18)$$

$$\lambda_3 = 0, \lambda_4 > 0 : [\lambda_1 - 1/\lambda_2, \lambda_1], \quad (19)$$

$$\lambda_3 < 0, \lambda_4 < 0 : (-\infty, \infty) \quad (20)$$

$$\lambda_3 < 0, \lambda_4 = 0 : (-\infty, \lambda_1 + 1/\lambda_2], \quad (21)$$

$$\lambda_3 = 0, \lambda_4 < 0 : [\lambda_1 - 1/\lambda_2, \infty). \quad (22)$$

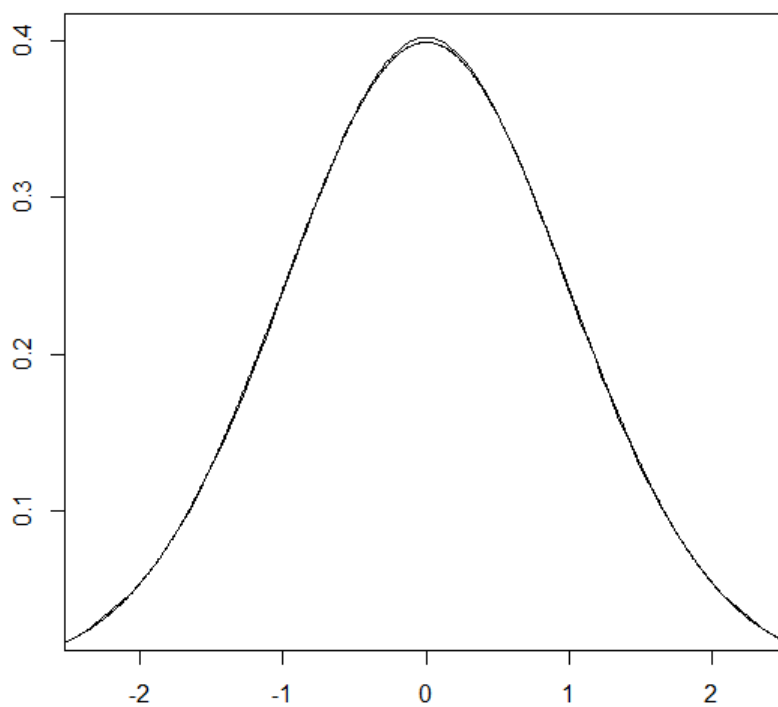
2.4 Genereerimine

Nagu sissejuhatuses mainitud, siis oli üldistatud lambda-jaotuse esimeseks kasutuselaks Monte Carlo simulatsioonide tarvis juhuslike suuruste genereerimine. Sellest lähtuvalt võib eeldada, et antud jaotusest juhuslike suuruste genereerimine peaks olema lihtne. Selgub, et see tõesti nii on ja see tuleneb asjaolust, et funktsioon on defineeritud läbi tema kvantiilfunktsiooni. Nagu me teame, siis kui $Q_X(y)$ on juhusliku suuruse X kvantiilfunktsioon ja U on ühtlase jaotusega, $U \sim U(0, 1)$, siis $Q_X(U)$ on sama jaotusega kui X . Seega piisab ühtlase jaotusega juhuslikust suurusest, et genereerida juhuslikke suuruseid ükskõik millisesest jaotusest, mida on võimalik lähendada üldistatud lambda-jaotusega.

Kirjeldatud juhuslike suuruste genereerimise lihtsus võib osutuda väga

käepäraseks, sest oma paindlikkuse tõttu suudab üldistatud lambdajaotus üsna hästi lähendada mitmeid levinud jaotuseid, mille puhul ei pruugi kvantiilfunktsioonil analüütilist kuju olla või siis on see keerukas. Näiteks annab standardsele normaaljaotusele hea lähendi $GLD(0, 0.1975, 0.1349, 0.1349)$. Võrdlus on toodud joonisel 3, kust on näha, et üldistatud lambdajaotus tõuseb tipus pisut kõrgemale, kuid on normaaljaotuse kõverale väga sarnane.

Joonis 3: $GLD(0, 0.1975, 0.1349, 0.1349)$ võrrelduna standardse normaaljaotuse tihedusega



3 Jaotuse sobitamine

Tulenevalt üldistatud lambdajaotuse paindlikkusest osutub jaotuse sobitamine andmetele üsnagi keeruliseks, sest nõuab nelja tundmatuga võrrandsüsteemi lahendamist. Sobivaima lahendi leidmiseks on välja pakutud mitmeid meetodeid, mis toovad kõik endaga kaasa mõne numbrilise meetodi rakendamise. See toob aga kaasa täiendavaid komplikatsioone juhul, kui lahendataval funktsioonil on mitu lokaalset ekstreemumit. Sellisel juhul sõltub saadud tulemus valitud algväärtustest ning lahendi sobivust tuleks kindlasti hoolikalt kontrollida. Mittesobivuse korral tuleks protsessi korrata uute algväärtustega. Järgnevalt kirjeldamegi enam levinud sobitamise meetodeid. Kindlasti pole tegemist ammendava nimekirjaga ning kirjandusest võib leida veel mitmeid meetodeid. Küll kasutavad enamuse neist meetoditest ühel või teisel viisil järgnevalt kirjeldatud momentide või kvantiilide meetodit.

3.1 Momentide meetod

Üldistatud lambdajaotuse sobitamine momentide meetodil seisneb esimese nelja momendi teoreetiliste avaldiste võrdsustamises vastavate valimi momentidega ning saadud võrrandsüsteemi lahendamises. Antud punkti aluseks on [4], lk. 41-66.

Üldistatud lambdajaotusega juhusliku suuruse X teoreetiliste karakteristikutena kasutame keskvaartust $\alpha_1 = E(X)$, dispersiooni $\alpha_2 = \sigma^2$, asümmeetriakordajat α_3 ja järsakuskordajat α_4 :

$$\alpha_1 = \mu = E(X) = \lambda_1 + \frac{A}{\lambda_2}, \quad (23)$$

$$\alpha_2 = \sigma^2 = E((X - \mu)^2) = \frac{B - A^2}{\lambda_2^2}, \quad (24)$$

$$\alpha_3 = \frac{E(X - E(X))^3}{\sigma^3} = \frac{C - 3AB + 2A^3}{\lambda_2^3 \sigma^3}, \quad (25)$$

$$\alpha_4 = \frac{E(X - E(X))^4}{\sigma^4} = \frac{D - 4AC + 6A^2B - 3A^4}{\lambda_2^4 \sigma^4}. \quad (26)$$

Siin

$$A = \frac{1}{1 + \lambda_3} - \frac{1}{1 + \lambda_4}; \quad (27)$$

$$B = \frac{1}{1 + 2\lambda_3} + \frac{1}{1 + 2\lambda_4} - 2\beta(1 + \lambda_3, 1 + \lambda_4); \quad (28)$$

$$C = \frac{1}{1 + 3\lambda_3} - \frac{1}{1 + 3\lambda_4} - 3\beta(1 + 2\lambda_3, 1 + \lambda_4) + 3\beta(1 + \lambda_3, 1 + 2\lambda_4); \quad (29)$$

$$D = \frac{1}{1 + 4\lambda_3} + \frac{1}{1 + 4\lambda_4} - 4\beta(1 + 3\lambda_3, 1 + \lambda_4) + 6\beta(1 + 2\lambda_3, 1 + 2\lambda_4) \quad (30)$$

$$- 4\beta(1 + \lambda_3, 1 + 3\lambda_4), \quad (31)$$

kus $\beta(a, b)$ on beeta-funktsioon:

$$\beta(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx \quad (32)$$

Oluline on mainida, et üldistatud lambda-jaotuste pere sisaldab ka jaotuseid, millel kõiki esimest nelja momenti ei eksisteeri. Et k -s moment eksisteeriks peavad kehtima võrratused:

$$\lambda_3 > -1/k, \quad (33)$$

$$\lambda_4 > -1/k. \quad (34)$$

Seega esimese 4 momendi olemasoluks piisab, kui nii λ_3 kui ka λ_4 on suuremad kui -0.25

Valimist saadud hinnangud, mis vastavad avaldistega (23), (24), (25), (26) kirjeldatud üldistatud lambdajaotuse teoreetilistele suurustele, saab leida järgnevate võrduste abil:

$$\hat{\alpha}_1 = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i; \quad (35)$$

$$\hat{\alpha}_2 = \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2; \quad (36)$$

$$\hat{\alpha}_3 = \frac{1}{n\hat{\sigma}^3} \sum_{i=1}^n (x_i - \bar{x})^3; \quad (37)$$

$$\hat{\alpha}_4 = \frac{1}{n\hat{\sigma}^4} \sum_{i=1}^n (x_i - \bar{x})^4. \quad (38)$$

Seega saame leida $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ lahendades võrrandsüsteemi $\{\alpha_i = \hat{\alpha}_i\}$ $i = 1, \dots, 4$. Lahendite leidmist lihtsustab asjaolu, et A, B, C, D ei sisalda λ_1 ja λ_2 . Samas taandub λ_2 avaldistest (25), (26) välja, ja saame esmalt lahendada võrrandid $\alpha_3 = \hat{\alpha}_3$ ja $\alpha_4 = \hat{\alpha}_4$ ning leida λ_3 ja λ_4 hinnangud. Saadud väärtuseid kasutame järgnevalt, et avaldada λ_2 ja seejärel λ_1 . Siiski on saadav kahest võrrandist koosnev süsteem küllaltki keeruline ja täpset analüütilist lahendit sellele ei leidu ning leppida tuleb numbriliselt lahendades saadud ligikaudsete tulemustega.

Lõpetuseks rõhutame, et antud meetodi kasutamise eelduseks on nelja esimese lõpliku momendi olemasolu. Seega jäävad jaotustest hulgast välja need, kus kõiki lõplikke momente ei eksisteeri. Lisaks on oluline meeles pidada, et erinevad parameetrite väärtused võivad anda tulemuseks samad momendid, seega tuleks iga saadud lahendi puhul testida selle sobivust empiiriliste andmetega.

3.2 Kvantiilide meetod

Üldistatud lambda-jaotuse parameetrite hindamine kvantiilide meetodil seisneb nelja kvantiilidel põhineva statistiku kasutamisel, mis võrdsustatakse seejärel vastavate teoreetiliste suurusetega. Tekkiva võrrandsüsteemi lahendamisel leitakse hinnangud parameetritele $\lambda_1, \lambda_2, \lambda_3, \lambda_4$. Antud punkti aluseks on [4], lk. 153-168.

Tähistame p -kvantiili hinnangut valimist $\hat{\pi}_p$. Võtame kasutusele järgmised neli empiiriliste kvantiilide kaudu määratud suurust:

$$\hat{\rho}_1 = \hat{\pi}_{0.5}; \quad (39)$$

$$\hat{\rho}_2 = \hat{\pi}_{1-u} - \hat{\pi}_u; \quad (40)$$

$$\hat{\rho}_3 = \frac{\hat{\pi}_{0.5} - \hat{\pi}_u}{\hat{\pi}_{1-u} - \hat{\pi}_{0.5}}; \quad (41)$$

$$\hat{\rho}_4 = \frac{\hat{\pi}_{0.75} - \hat{\pi}_{0.25}}{\hat{\rho}_2}, \quad (42)$$

kus u on suvaline arv vahemikus 0 kuni 0,25

p -kvantiili hinnangu $\hat{\pi}_p$ leidmiseks tuleb $(n+1)p$ kirjutada kujul $r + a/b$, kus r on positiivne täisarv ning a/b on murd, mille väärtus võib olla null. Kui y_1, y_2, \dots, y_n on valimi x_1, x_2, \dots, x_n variatsioonirea elemendid, siis

$$\hat{\pi}_p = y_r + \frac{a}{b}(y_{r+1} - y_r) \quad (43)$$

Valemis (2) defineeritud kvantiilfunktsioonist saame leida vastavad teoreetilised suurused järgmiselt [4] lk. 155:

$$\rho_1 = Q\left(\frac{1}{2}\right) = \lambda_1 + \frac{\left(\frac{1}{2}\right)^{\lambda_3} - \left(\frac{1}{2}\right)^{\lambda_4}}{\lambda_2}; \quad (44)$$

$$\rho_2 = Q(1-u) - Q(u) = \frac{(1-u)^{\lambda_3} - u^{\lambda_4} + (1-u)^{\lambda_4} - u^{\lambda_3}}{\lambda_2}; \quad (45)$$

$$\rho_3 = \frac{Q\left(\frac{1}{2}\right) - Q(u)}{Q(1-u) - Q\left(\frac{1}{2}\right)} = \frac{(1-u)^{\lambda_4} - u^{\lambda_3} + \left(\frac{1}{2}\right)^{\lambda_3} - \left(\frac{1}{2}\right)^{\lambda_4}}{(1-u)^{\lambda_3} - u^{\lambda_4} + \left(\frac{1}{2}\right)^{\lambda_4} - \left(\frac{1}{2}\right)^{\lambda_3}}; \quad (46)$$

$$\rho_4 = \frac{Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right)}{\rho_2} = \frac{\left(\frac{3}{4}\right)^{\lambda_3} - \left(\frac{1}{4}\right)^{\lambda_4} + \left(\frac{3}{4}\right)^{\lambda_4} - \left(\frac{1}{4}\right)^{\lambda_3}}{(1-u)^{\lambda_3} - u^{\lambda_4} + (1-u)^{\lambda_4} - u^{\lambda_3}}. \quad (47)$$

3.3 Numbriline logaritmilise tõepära meetod

Täiendusena seni kirjeldatud meetoditele pakkus Steve Su oma 2006 aasta artiklis [8] välja meetodi, kus esimese sammuna leitakse kas kvantiilide või momentide meetodit kasutates algväärtused, mis seejärel on sisendiks logaritmilise tõepärafunktsiooni numbrilisel maksimeerimisel. See meetod on käesoleva töö raames oluline, sest see on realiseeritud tarkvarapakettis R, kus on teostatud töö praktiline osa. Pakettis R on realiseeritud nii RS kui ka FMKL parametrisatsiooni sobitamine, kusjuures esimene kasutab lähte- väärtuste saamiseks kvantiilide meetodit ja teine momentide meetodit. Kuna antud töös keskendume peamiselt RS parametrisatsioonile, siis toome siinkohal ära meetodi kirjeldusest ainult selle osa. FMKL parametrisatsiooni jaoks välja töötatud protseduuriga on võimalik tutvuda artiklis [8]. Üldistatud lambdajaotuse RS parametrisatsiooni sobitamine koosneb järgmistest

sammudest.

1. Vali λ_3 ja λ_4 jaoks muutumisvahemik ning valitavate lähteväärtuste arv. Vaikimisi kasutatav muutumisvahemik on $(-1, 5; 1, 5)$ ning valitavaid lähteväärtuseid on 10 000. Valik tehakse kasutades pseudojuhuslike arvude generaatorit, et kahemõõtmeline parameeterruum saaks ühtlaselt kaetud.

2. Kasutades valitud λ_3 ja λ_4 väärtuseid arvuta võrrandist (45) λ_2 ning võrrandist (44) λ_1 . Saadud tulemuste hulgast tuleb eemaldada variandid, mis ei määra jaotust või ei kata valimi ulatust. Järelejäänud $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ kombinatsioonide hulgast tuleb valida see, mille korral järgnev avaldises (48) on minimaalne:

$$\sqrt{(\hat{\rho}_3 - \rho_3)^2 + (\hat{\rho}_4 - \rho_4)^2} \quad (48)$$

3. Kasutades valitud parameetrite kombinatsiooni ja kvantiilfunktsiooni valemist (2), arvuta valimi elementidele $1, 2, \dots, n$ vastavad jaotusfunktsiooni väärtused u_i . Kuna üldistatud lambdajaotuse jaotusfunktsiooni analüütilisel kujul esitada ei ole võimalik, siis tuleb väärtused u_i leida numbriliselt.

4. Asenda eelmises punktis saadud väärtused alltoodud võrdusesse

$$ML_{RS} = \sum_{i=1}^n \ln \left[\frac{\lambda_2}{\lambda_3 u_i^{\lambda_3 - 1} + \lambda_4 (1 - u_i)^{\lambda_4 - 1}} \right]. \quad (49)$$

5. Logaritmilise tõepärameetodi hinnangute leidmiseks tuleb kasutada mõnda numbrilist meetodit. R paketi on realiseeritud Nelder-Mead simpleksmeetod [6], mille abil saadud tulemuse sobivust tuleks kindlasti kontrollida ja vajadusel korrata protsessi uute algväärtustega. Nelder-Mead simpleksmeetodil n muutujaga funktsiooni minimeerimiseks läbiviidav algoritm on järgmine:

I - Esimese sammuna tuleb valida n -mõõtmelises ruumis $n + 1$ punkti, mis moodustavad vaadeldava simpleksi ja leida funktsiooni väärtused $y_i, i = 1, \dots, n+1$ igas valitud punktis. Siinjuures tähistame $y_h = \max(y_i)$ ja $y_l = \min(y_i)$ ning neile vastavaid punkte P_h ja P_l .

II - Järgmise sammu eesmärgiks on vaadeldavas simpleksis vahetada välja punkt, mis annab suurima funktsiooni väärtuse, s.t. funktsiooni väärtus antud punktis on y_h . Uue punkti valimiseks on ette nähtud 3 meetodit: peegeldamine, kokkutõmbumine ja laiendamine. Neile vastavad uued punktid leitakse järgnevate võrduste kaudu:

$$P_r = (1 + \alpha)\bar{P} - \alpha P_h; \quad (50)$$

$$P_c = \beta P_h + (1 - \beta)\bar{P}; \quad (51)$$

$$P_e = \gamma P_r + (1 - \gamma)\bar{P}, \quad (52)$$

kus \bar{P} on punktide $i \neq h$ tsentroid ning α, β, γ on vastavalt peegelduskordaja, kokkutõmbumiskordaja ja laiendamiskordaja, mille kasutatavad vaikeväärtused on tarkvarapeketis R vastavalt 1, 0,5 ja 2. Kui $y_l \leq y_r \leq y_h$, siis P_h vahetatakse välja punktiga P_r . Kui $y_r < y_l$, s.t. peegeldus on andnud tulemuseks punkti, mille puhul funktsiooni väärtus

on väiksem, kui üheski teises simpleksi punktis ja $y_e < y_l$, siis vahetatakse P_h välja punktiga P_e . Samas kui $y_e > y_l$, siis jäädakse vahetuseks punkti P_r juurde. Kui nende sammude järel selgub, et uueks simpleksi punktiks valiti P_r , kuid $y_r > y_i$ iga $i \neq h$, s.t. y_r oleks uus maksimum, siis vaadatakse, kas funktsiooni väärtus on väiksem punktis P_h või P_r ja kasutatakse väiksema tulemuse andnud punkti avaldises (51) P_h rollis, et leida punkt P_c ja kasutada seda P_h asendusena. Kui nüüd peaks juhtuma, et $y_c > \min(y_h, y_r)$, siis vahetatakse kõik punktid P_i välja punktidega $(P_i + P_l)/2$

III - Kui eelmisel sammul välja vahetatav punkt või punktid on leitud, liigutakse tagasi protsessi algusesse ja hakatakse leidma vahetust uuele punktile P_h . Protsess lõpetatakse, kui $\sqrt{\sum(y_i - \bar{y})^2/n}$ on väiksem, kui etteantud konstant. Tarkvarapakettis R realiseeritud vaikeväärtus on ligikaudu $1,49 \cdot 10^{-8}$

3.4 Veel üks kvantiilidel põhinev meetod

Lisaks peatükis 3.2 kirjeldatud meetodile pakkus Steve Su 2008. aastal välja uue meetodi [3], lk. 557-583, mis sarnaneb paljuski eelmises peatükis kirjeldatud lahendusega, kuid kasutab parameetritele hinnangute leidmiseks teoreetiliste ja empiiriliste kvantiilide vahede ruutude summa minimeerimist. Erinevus numbrilise logaritmilise tõepära meetodist algab algoritmi 3. sammust. Antud meetodi puhul näevad edasised sammud välja järgmised:

3. Jaga intervall $(0,1)$ tuhandeks võrdse pikkusega vahemikuks ning kasutades

kvantiilfunktsiooni valemis (2), avalda 1000 teoreetilist kvantiili q_i parameetrite $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ kaudu.

4. Arvuta eelmises punktis leitud teoreetilistele kvantiilidele vastavad valimi hinnangud \hat{q}_i ning asenda need järgmisesse avaldisse:

$$\sum_{i=1}^{1000} (q_i - \hat{q}_i)^2 \quad (53)$$

5. Eelmises punktis toodud avaldist minimeeriva parameetrite kombinatsiooni $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ leidmiseks tuleb kasutada mõnda numbrilist meetodit ja sarnaselt eelmises punktis kirjeldatuga, kasutatakse tarkvarapaketi R vaiki-misi Nelder-Mead simpleksmeetodit. Punktis 2. leitud parameetrite kombi-natsiooni kasutatakse numbrilise minimeerimise algpunktina.

3.5 Võrestiku meetod (inglise keeles *Starship method*)

Võrestiku meetod, mille pakkusid 1999. aastal välja Robert King ja He-len MacGillivray [5], püüab pakkuda lahendust teiste meetoditega kaasnevale probleemile, kus lahendi leidmiseks tuleb kasutada mõnda numbrilist mee-todit. Selle asemel keskendutakse ainult valitud punktides jaotuse sobivuse testide läbiviimisele. Võrestiku meetod koosneb järgmistest sammudest:

1. Valitakse 4-mõõtmelise ruumi osa, kust sobivat lahendit $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ otsida soovitakse.
2. Eelmises sammus valitud ruumi osale asetatakse 4-mõõtmeline võre, mille tulemusel moodustub $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ punktide hulk, kust lahendit otsitakse

3. Defineeritud võre punktides hakatakse kontrollima antud punktis määratud jaotuse sobivust valimile. Selleks kasutatakse Anderson-Darlingi kooskõla testi teoreetilise ja empiirilise jaotuse võrdlemiseks [11]. Kuna selle statistiku kriitilised väärtused on teada ainult teatud jaotuste puhul, mille hulgas ei ole üldistatud lambdajaotust, siis tuleb leida valimi elementidele x_1, \dots, x_n vastavad, $\lambda_1, \lambda_2, \lambda_3, \lambda_4$, poolt määratud jaotusfunktsiooni, väärtused $F(x_1), \dots, F(x_n)$. Kuna jaotusfunktsiooni väärtused on sellisel juhul ühtlase jaotusega $U(0, 1)$ ning ühtlase jaotuse jaoks on teada Anderson-Darlingi statistiku kriitilised väärtused, siis saamegi kontrollida jaotusfunktsiooni väärtuste sobivust ühtlase jaotusega. Kuna üldistatud lambdajaotuse jaotusfunktsioon ei ole analüütilisel kujul esitatav, siis tuleb vastavad väärtused arvutada numbriliselt. Anderson-Darlingi statistiku väärtus leitakse järgneva valemi abil:

$$A^2 = -\frac{1}{n} \sum_{i=1}^n (2i - 1)(\ln(z_i) + \ln(1 - z_{n+1-i})) - n, \quad (54)$$

kus $z_i = F(y_i)$ ja y_i $i = 1, \dots, n$ on valimi variatsioonirida. $F(x)$ tähistab siin teoreetilist jaotusfunktsiooni.

4. Lõplikuks lahendiks valitakse parameetrite kombinatsioon $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, mis annab kõige väiksema Anderson-Darlingi statistiku väärtuse. Teise variandina võib protsessi lõpetada juba esimese parameetrite kombinatsiooni juures, mis annab kriitilisest väärtusest väiksema statistiku väärtuse. Sellisel juhul on võimalik vähendada arvutusteks kuluvat aega.

5. Lisaks eelnevalt kirjeldatud sammudele, mis pärinevad otseselt artiklist [5], on tarkvarapaketi R realiseeritud lisasamm, kus punktis 4. saadud väärtused

on sisendiks Anderson-Darlingi statistiku väärtuse minimeerimiseks Nelder-Mead simpleksmeetodil.

Võrestiku meetodi näol on tegemist lahendusega, mis toimib hästi paljudes situatsioonides, kuid probleemiks on asjaolu, et sobiva lahendi leidmiseks vajalik aeg kasvab kiiresti koos valimi mahu suurenemisega. Lisaks võib arvutusteks kuluda palju aega, kui otsinguks kasutatav võre on halvasti valitud. Sobiva ruumi defineerimist võib lihtsustada peatükis 2.3 toodud parameeter-ruumi kirjelduse abil.

4 Andmete kirjeldus

Antud peatükis anname ülevaate töö praktilises osas kasutatud andmetest. Esmalt selgitame andmete päritolu ja seejärel liigume edasi statistilise kirjeldamise juurde.

4.1 Andmete päritolu

Kasutatavate andmete näol on tegemist ühe Eestis tegutseva kindlustusseltsi juriidiliste isikute varakindlustuse toote kahjudega, mis on teatatud kindlustusseltsile ajavahemikus 01.01.2002-31.12.2012. Kindlustusseltsi jaoks realiseerunud kahjusummasid on suurendatud poliisis ette nähtud omavastutuse võrra, et saada toimunud kahju tegelik suurus. Lisaks on andmetele rakendatud ehitushinna indeksi muutusest tulenev korrektsioon, et saada hinnangut toimunud kahju nüüdisväärtusele 2012. aastal. Rakendatud korrektsiooni kordajad on toodud tabelis 2. Märkime veel, et tulenevalt juriidiliste isikute varakindlustuse toote eripärast, kus kindlustatakse kohati äärmiselt

suuri riske, kuid kahjusagedused on madalad, ei anna realiseerunud kahjud adekvaatset hinnangut tegelikult võimaliku maksimaalse kahju suurusele. Positiivse tõenäosusega võib realiseeruda oluliselt suuremaid kahjusid, kui on valimi maksimaalne element.

4.2 Andmete ülevaade

Kasutatavas andmestikus on kokku 2802 vaatlust, kusjuures minimaalne väärtus on 2 eurot ning maksimaalne 1 288 968 eurot. Keskmise teatatud juhtumi kahjusumma on 7 352 eurot. Täpne ülevaade jaotuse kvantiilidest on leitav alltoodud tabelis 1. Nagu näha, on tegemist suure varieeruvusega andmetega. Kui kahjusumma kokku on 20 600 178 eurot, siis 10 kõige suuremat juhtumit moodustavad sellest koguni 6,7 miljonit eurot, ehk ligi 30%. Seega on selge, et jaotuse sobitamise ja hilisema sobivuse testimise juures saab määravaks saba käitumine, s.t. et sobitatud jaotus peab andma adekvaatse hinnangu kõige suuremate kahjude toimumise sagedusele ja suurusele.

Tabel 1: Valimi kvantiilid

Tõenäosus	Kvantiil	Tõenäosus	Kvantiil
0	2	0.55	1311
0.001	6	0.6	1594
0.005	37	0.65	1953
0.01	55	0.7	2394
0.05	109	0.75	2990
0.1	170	0.8	3969
0.15	232	0.85	5339
0.2	315	0.9	8812
0.25	394	0.95	20555
0.3	495	0.99	109657
0.35	598	0.995	234476
0.4	724	0.999	762716
0.45	882	1	1288968
0.5	1048		

Tabel 2: Ehitushinna korrektsioonid

Aasta	Ehitushinna korrektsioon
2002	1.32
2003	1.29
2004	1.24
2005	1.19
2006	1.11
2007	0.99
2008	0.96
2009	1.05
2010	1.07
2011	1.04
2012	1.00

5 Sobitamine varakindlustuse kahjuandmetele

Käesolevas peatükis liigume töö praktilise osa juurde, kus kasutades eelnevates peatükkides kirjeldatud teooriat, üritame sobitada üldistatud lambda-jaotust juriidiliste isikute varakindlustuse kahjudele. Selleks kasutame tarkvarapaketi R realiseeritud meetodeid, mis asuvad paketi *GLDEX*. Saadud tulemuste jaoks toome välja parameetrite hinnangud ning võrdleme parameetrite poolt määratud jaotuse sobivust valimiga. Lisaks paketi *GLDEX* tehnilisele kirjeldusele [9] on töö praktilises osas kasutatud ka Steve Su artiklit [10] antud paketi kasutamisest.

5.1 Sobitamine

Sobitamise käigus leidsime hinnangud parameetritele neljal erineval viisil. Saadud tulemused on toodud tabelis (3). Parameetritele hinnangu leidmiseks kasutatud meetodid on järgnevad:

ML - punktis 3.3 kirjeldatud sobitamise meetod.

MM - punktis 3.1 kirjeldatud sobitamise meetod.

QS - punktis 3.4 kirjeldatud sobitamise meetod.

STAR - punktis 3.5 kirjeldatud sobitamise meetod.

Erinevalt ülejäänud kolmest meetodist on antud juhul kasutatud FMKL parametrisatsiooni, sest seda meetodit pole RS parametrisatsioonile paketi R realiseeritud. Seda saab selgitada asjaoluga, et nagu varasemalt mainitud, on FMKL parametrisatsiooni puhul oluliselt vähem kitsendusi, millele parameetrite $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ väärtused vastama peavad.

Tabel 3: Hinnangud parameetrite väärtustele ümardatuna nelja komakohani

Tüüp	λ_1	λ_2	λ_3	λ_4
ML	126.2212	-0.0012	-0.0200	-1.0725
MM	-1128.9119	0.0000	-0.0003	-0.2116
QS	-1355.7432	-0.0004	-0.0219	-0.8577
STAR	699.7949	0.0016	0.8839	-1.2738

Nagu tabelis 3 näha, on eri meetoditega saadud hinnangud väga erinevad.

5.2 Sobivuse kontroll

Käesolevas peatükis võrdleme hinnatud parameetritega määratud jaotuste sobivust kahjude andmestikuga. Selleks viime läbi mõned sobivuse kontrollid, millest esimesena vaatleme Kolmogorov-Smirnovi taasvaliku meetodit. Hinnangutega, mille puhul Kolmogorov-Smirnovi taasvaliku meetod hindab sobivuse piisavalt heaks, liigume edasi kvantiilide võrdlemise juurde. Vahepeal toome ära mõningad selgitused, miks osad parameetrite hindamise meetodid ei anna andmetega head sobivust.

5.2.1 Kolmogorov-Smirnovi taasvaliku meetod

Antud meetod seisneb 1000 iseseisva Kolmogorov-Smirnovi testi läbi viimises, kus igaüheks valitakse juhuslikult 90% valimi elementidest, ning võrreldakse neid sama arvu teoreetilisest jaotusest genereeritud juhuslike väärtustega. See meetod on realiseeritud *GLDEX* pakettis. Saadud tulemused on toodud tabelis 4.

Tabel 4: Kolmogorov-Smirnovi taasvaliku meetodi tulem 1 000 korduse puhul

ML	MM	QS	STAR
963	0	0	959

Tabelis toodud arvud näitavad testide arvu, mille puhul jäädid nullhüpoteesi juurde, et empiirilise ja teoreetilise jaotuse vahel on kooskõla. Nagu näha, siis nii ML kui ka STAR meetodi puhul jäädakse rohkem kui 950 sõltumatu KS testi puhul nullhüpoteesi juurde. See tähendab, et vähem kui viiekümnel juhul osutus empiirilise ja teoreetilise jaotusfunktsiooni maksimaalne vahe

suuremaks, kui Kolmogorov-Smirnovi testi kriitiline piir. Seega on üsna tugev alus väite, et valimi sobivus teoreetilise jaotusega on hea. Samas ei jääda nii MM ega ka QS meetodi puhul mitte ühtegi korda nullhüpoteesi juurde, millest saame järeldada, et antud meetoditega saadud parameetrite hinnangud ei anna jaotust, mis sobiks kahjude valimiga.

5.2.2 Mittesobivuse analüüs

Enne kui liigume edasi ML ja STAR meetoditega saadud parameetrite poolt määratud jaotuste kvantiilide analüüsi juurde, uurime põhjuseid, miks MM ja QS meetoditega saadud hinnangud ei paista valimile sobivat. Selleks tuletame esmalt meelde, mis on kummagi sobitamismeetodi aluseks, s.t. mis kriteeriumi alusel parameetreid hinnatakse.

MM ehk momentide meetodi puhul üritatakse leida parameetrite kombinatsiooni, mis annab kõige parema momentide sobivuse. See lähenemine tekitab aga töös kasutatavate andmete kontekstis kohe küsimusi, sest võib eeldada, et antud andmed on raske sabaga ning seega kõiki momente ei eksisteeri. Antud seisukohta toetab ka asjaolu, et kõigi ülejäänud meetoditega saadud hinnangute puhul on mõne parameetri hinnang väiksem kui $-0,25$, mis tähendab, et kõiki momente ei eksisteeri. ML ja STAR meetodite puhul on minimaalne parameetri hinnang isegi väiksem kui -1 , mis tähendab, et ühtegi momenti ei eksisteeri.

QS ehk kvantiilide meetodi puhul on eesmärgiks minimeerida valimi ja teoreetiliste kvantiilide erinevus. See tähendab, et kui leida kõikide meetoditega saadud parameetrite hinnangutele vastavate jaotuste kvantiilid ja arvutada nende ning valimi kvantiilide vahede ruutude summa, siis peaks antud

meetod andma väikseima arvu. Kui selline arvutus 1000 kvantiili jaoks teha, siis selgub, et nii see tõesti ka on. Siinkohal täpseid tulemusi välja tooma ei hakka, sest isegi QS meetodi puhul on see üheteistkohaline arv. Teiste meetodite puhul on tulemus suurusjärgus kümme kuni sada korda suurem. Antud lähenemine peaks seega andma ka parima sobivuse kvantiilide tasemel. Kvantiilide võrdlus on toodud tabelis 5 ja nagu näha, siis vastupidiselt ootusele on kvantiilide sobivus üsna kehv ning lisaks sellele katab antud jaotus ka negatiivseid väärtuseid, kuigi kasutatavas andmestikus on ainult positiivsed suurused. Küll paneme tähele, et sobivus on küllaltki hea 99% kvantiili puhul ja selle põhjal võib ka aimata, miks jaotuse üldine sobivus mitterahuldav on. Nimelt, kui üritatakse minimeerida kvantiilide erinevust ja tegemist on raske sabaja jaotusega, siis mõjutavad suuremad kvantiilid sobitamise tulemust ebaproportsionaalselt palju. Seda põhjusel, et proportsionaalselt üsnagi väike erinevus kõrge kvantiili puhul võib absoluutarvuna anda oluliselt suurema vahe kui suur erinevus madala kvantiili puhul.

Tabel 5: Kvantiilide meetodil sobitatud jaotuse kvantiilide võrdlus valimiga

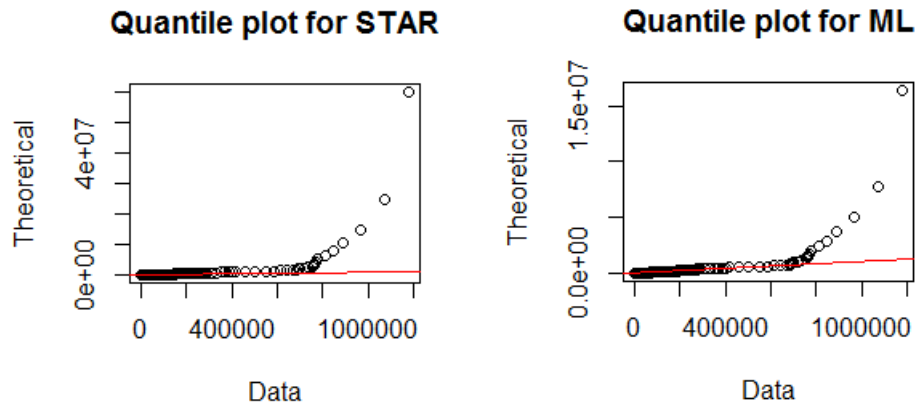
Tõenäosus	QS kvantiilid	Valimi kvantiilid
0.01	-1575	55
0.05	-1407	109
0.1	-1259	170
0.25	-797	394
0.5	433	1053
0.75	3757	3015
0.9	12572	8952
0.95	25714	20924
0.99	112983	109657

5.2.3 Kvantiilide võrdlus

Edasi liigume kvantiilide võrdlemise juurde. Siinkohal jätame välja jaotused, mille puhul jõudsime eelmises punktis järelduseni, et sobivust ei ole, s.t. kus Kolmogorov-Smirnovi taasvaliku testi tulemused olid nullid. Seega jäävad järgi ainult ML ja STAR hinnangud. Nende mõlema graafikud on toodud joonisel 4.

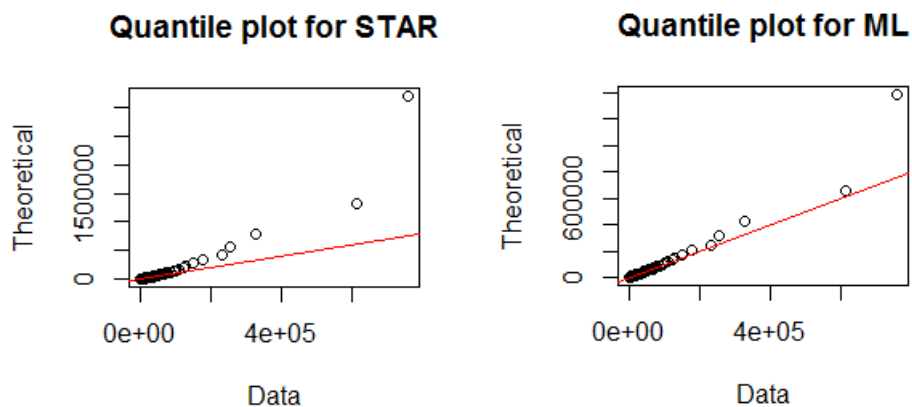
Nagu jooniselt näha, siis kvantiilide kattuvus jaotuse sabal ei ole hea. Seeda mõlema variandi puhul. Siinkohal tuleb meele pidada, et vaikeväärtusena kasutab tarkvarapakett R joonise tegemiseks 10 000 kvantiili. Samas on meie andmestikus alla kolme tuhande vaatluse ning samuti teame, et positiivse tõenäosusega võib tekkida veel oluliselt suuremaid kahjusid, seega on mõist-

Joonis 4: kvantiilide võrdlus (10 000 kvantiili)



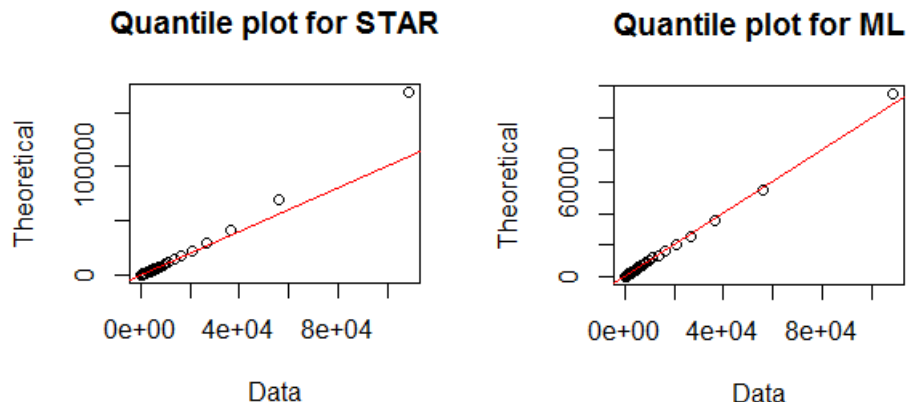
lik kontrollida sobivust väiksema arvu kvantiilide puhul. Joonisel 5 toodud graafik näitabki 1 000 kvantiili võrdlust. Nagu näha, siis on tulemus ML meetodi jaoks juba üsna hea, ainult viimane vaatlus on joonest märkimisväärselt eemal. STAR puhul on sobivus mõnevõrra halvem ja suurem hulk viimaseid kvantiile erineb.

Joonis 5: kvantiilide võrdlus (1 000 kvantiili)



Kuigi eelmine joonis andis juba põhjuse pidada ML meetodil saadud parameetreid üsna sobivaks, vaatame veel ka 100 kvantiili graafikut. See on toodud joonisel 6. Nagu näha, siis võib sellisel tasemel sobivust pidada ML meetodi puhul juba väga heaks, aga STAR-i puhul jääb viimased punktid joonest jällegi üsna kaugele, millest saame järeldada, et sobivus viimaste kvantiilide puhul ei ole hea

Joonis 6: kvantiilide võrdlus (100 kvantiili)



Et saada veelgi paremat pilti ML meetodil saadud parameetrite sobivusest, toome ära ka kvantiilide väärtuste võrdluse juba varem välja toodud valimi kvantiilidega, mida on võimalik näha tabelis 6. Nagu sellest tabelist näha, on kvantiilide sobivus kuni 95%-ni isegi väga hea. 99,9% kvantiil on ainukene toodud näidetest, kus erinevus on juba märkimisväärne, kuid siinkohal peab meeles pidama, et valimis on alla kolme tuhande vaatluse ja seega tähendab nõnda kõrge kvantiili arvutus kahe suure väärtuse vahel interpoleerimist. Võttes arvesse ka asjaolu, et tegemist on raske sabaga andmetega, mille arengu kogu võimalik ulatus ei ole teada, leiab autor, et ML meetodil

sobitatud jaotus on igati kohane kirjeldama juriidiliste isikute varakindlustuse kahju suuruse jaotust.

Tabel 6: Valimi kvantiilide võrdlus ML meetodil sobitatud jaotusega

Tõenäosus	Valim	Jaotus	Tõenäosus	Valim	Jaotus
0.001	6	2	0.55	1314	1256
0.005	37	36	0.6	1605	1524
0.01	55	54	0.65	1962	1872
0.05	109	122	0.7	2401	2339
0.1	170	187	0.75	3016	3002
0.15	232	254	0.8	4011	4009
0.2	315	326	0.85	5396	5719
0.25	394	407	0.9	8952	9227
0.3	495	498	0.95	20924	20196
0.35	600	602	0.99	109657	116794
0.4	725	725	0.995	234476	246421
0.45	883	869	0.999	762716	1387970
0.5	1053	1043			

5.3 Võrdlus alternatiividega

Järgnevalt võrdleme, kuidas üldistatud lambda-jaotusega saadud tulemus käitub võrrelduna teiste kindlustuses enam levinud kahjujaotustega. Vaadeldavad jaotused on Weibulli jaotus, lognormaalne jaotus, gammajaotus ja eksponentjaotus. Võrdluse aluseks on jaotuse sobitamine kogu andmestikule, sest see on käesoleva töö peamine eesmärk. Seega ei uuri me eraldi juhte, kus mõni allolevatest jaotustest võiks anda paremaid tulemusi, kui sobitamine oleks tehtud eraldi kas väiksematele või suurematele kahjudele. Kõik allolevad võrdlused on läbi viidud tarkvarapakettis R. Tulemusi võrreldes keskendume peamiselt graafikute analüüsile, mis autori hinnangul annab kõige parema ülevaate sobivusest kogu jaotuse ulatuses.

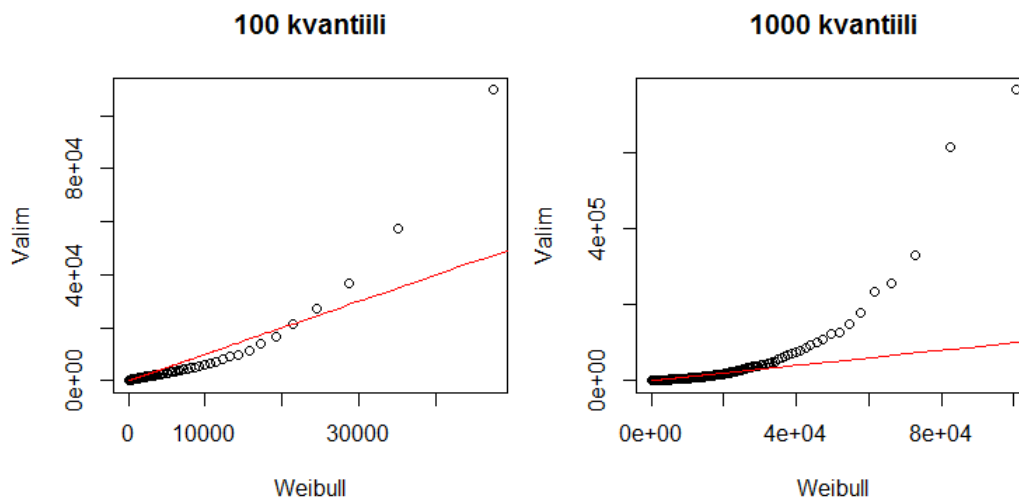
5.3.1 Weibulli jaotus

Weibulli jaotus on kaheparameetriline tõenäosusjaotus, mille jaotusfunktsioon on defineeritud järgmiselt:

$$F(x) = 1 - e^{-(x/\lambda)^k}. \quad (55)$$

Jaotuse parameetrite hindamiseks on kasutatud tarkvarapakettis R reliseeritud funktsiooni *fitdist* paketist *fitdistrplus*. Tulemuseks on järgnevad hinnangud: $k = 0,5392$ ja $\lambda = 2792,8340$. Antud parameetrite poolt määratud jaotuse kvantiilide võrdlus valimiga on toodud joonisel 7. Nagu näha, siis on madalamate kvantiilide puhul märgata teatavat sarnasust, kuid juba alates 95% kvantiilist hakkab empiiriline jaotus kasvama kiiremini kui sobitatud Weibulli jaotuse kvantiilid.

Joonis 7: Weibull võrrelduna valimiga



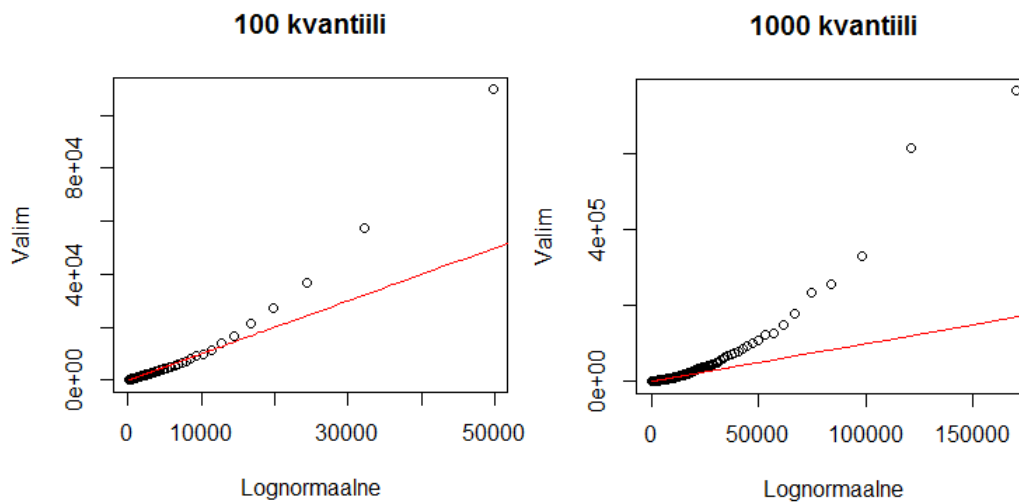
5.3.2 Lognormaalne jaotus

Lognormaalne jaotus on kaheparameetriline tõenäosusjaotus, mida on kõige lihtsam defineerida läbi tema tihedusfunktsiooni:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right). \quad (56)$$

Jaotuse parameetrite hindamiseks on kasutatud tarkvarapaketi R reliseeritud funktsiooni *fitdist* paketist *fitdistrplus*, mille rakendamise tulemuseks on parameetrite hinnangud $\mu = 7,0792$ ja $\sigma = 1,6062$. Antud parameetrite poolt määratud jaotuse kvantiilide võrdlus valimiga on toodud joonisel 8. Nagu näha, siis on madalamate kvantiilide sobivus parem kui see oli Weibulli jaotuse puhul, kuid samuti 95% kvantiilist alates hakkavad valimi kvantiilid kasvama märksa kiiremini.

Joonis 8: Lognormaalne võrrelduna valimiga



5.3.3 Gammajaotus

Gammajaotus on kaheparameetriline tõenäosusjaotus, mille jaotusfunktsioon avaldub järgmiselt:

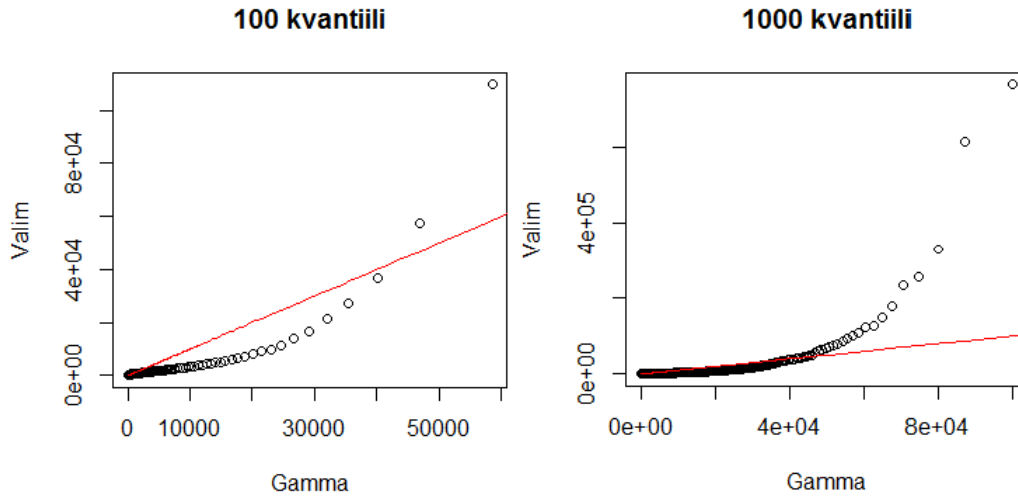
$$F(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}, \quad (57)$$

kus $\Gamma(k)$ on gammafunktsioon:

$$\Gamma(k) = \int_0^{\infty} t^{k-1} e^{-t} dt. \quad (58)$$

Jaotuse parameetrite hindamiseks on kasutatud tarkvarapaketi R reliseeritud funktsiooni *fitdistr* paketist *MASS*. Tulemuseks on parameetrite hinnangud $k = 0,3646$ ja $\theta = 20352$. Antud parameetrite poolt määratud jaotuse kvantiilide võrdlus valimiga on toodud joonisel 9.

Joonis 9: Gamma võrrelduna valimiga



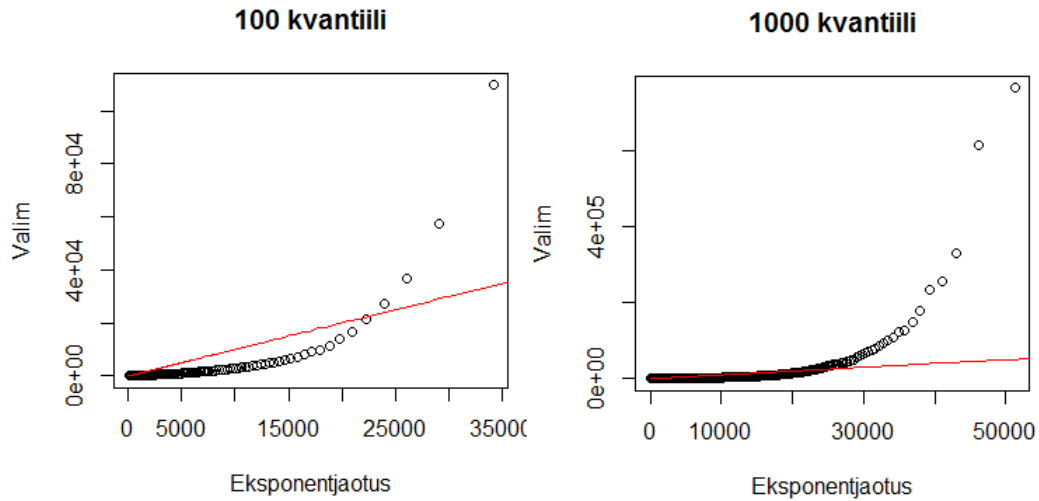
5.3.4 EkspONENTJAOTUS

EkspONENTJAOTUS on üheparameetriline tõenäosusjaotus, mille jaotusfunktsioon avaldub järgmiselt:

$$F(x) = 1 - e^{-\lambda x} \quad (59)$$

Jaotuse parameetri hindamiseks on kasutatud tarkvarapaketi R reliseeritud funktsiooni *fitdistr* paketi *MASS*. Tulemuseks on parameetri hinnang $\lambda = 0.00013479$. Antud parameetri poolt määratud jaotuse kvantiilide võrdlus üldistatud lambdajaotusega on toodud joonisel 10. Nagu näha, siis on sobivus halb ja isegi 100 kvantiili puhul on kattuvaid punkte vähe.

Joonis 10: Eksponentjaotuse võrrelduna valimiga



5.3.5 Võrdluste kokkuvõte

Võrdluse tulemuste põhjal saab kokkuvõtteks öelda, et ükski vaadeldud jaotustest ei sobitunud juriidiliste isikute varakindlustuse andmetele paremini, kui üldistatud lambdajaotus. Weibulli ja lognormaalse jaotuse puhul võis küll täheldada teatavat sobivust madalamate kvantiilide puhul, kuid paremal sabas jäid erinevused siiski suureks. Seega vaadeldud jaotuste hulgas ei leidu alternatiivi, mis annaks üldistatud lambdajaotusest parema tulemuse või oleks isegi tõsiseltvõetavaks alternatiiviks

6 Kokkuvõte

Üldistatud lambdajaotus on paindlik tõenäosusjaotus, kuid parameetrite arv ning üsnagi keerukas parameeterruum muudavad parameetrite hindamise keerukaks. Samas on tegemist jaotuste perega, mis suudab hästi lähendada mitmeid levinud jaotusi ning sobib ka raske sabaga andmete puhul. Koos asjaoluga, et jaotuse kvantiilfunktsioon on esitatav võrdlemisi lihtsal analüütilisel kujul, on tegemist väga käepärase töövahendiga. Seda kinnitavad ka empiirilised katsed juriidiliste isikute varakindlustuse kahju andmetel, mille puhul üldistatud lambdajaotus andis igati hea lähendi, mis autori hinnangul on praktikas kasutatav. Samas ükski võrdlusena vaadeldud jaotustest ei suutnud vaadeldavate andmetega rahuldavat kooskõla saavutada.

Sellest lähtuvalt soovitab autor jaotust sobitades kaalude variandina ka üldistatud lambdajaotust kogu jaotuse kirjeldamiseks.

Viited

- [1] Freimer, Marshal & Kollia, Georgia & Mudholkar, Govind S. & Lind, C Thomas (1998) A study of the generalized Tukey lambda family. *Communications in Statistics - Theory and Methods* **17**, No. 10, 3547-3567
- [2] Hastings, Cecil Jr. & Mosteller, Frederick & Tukey, John Q. & Winsor, Charles P. (1947) Low moments for small samples: A comparative study of order statistics *The Annals of Mathematical Statistics* **18**, No 3, 413-426
- [3] Karian, Zaven A. & Dudewicz, Edward J. (2010) *Handbook of Fitting Statistical Distributions with R*. CRC Press, Boca Raton
- [4] Karian, Zaven A. & Dudewicz, Edward J. (2000) *Fitting Statistical Distributions The Generalized Lambda Distribution and Generalized Bootstrap Methods*. CRC Press, Boca Raton
- [5] King, Robert A.R. & MacGillivray, H.L. (1999) A Starship estimation method for the generalized lambda distribution *Australian and New Zealand Journal of Statistics* **41**, No. 3, 353-374
- [6] Nelder, J. A. & Mead, R. (1965) A simplex method for function minimization *Computer Journal* **7**, No 4, 308-313
- [7] Ramberg, John S. & Schmeiser Bruce W. (1974) An approximate method for generating asymmetric random variables *Communications of the ACM* **17**, No. 2, 78-82

- [8] Su, Steve (2007) Numerical maximum log likelihood estimation for generalized lambda distribution *Computational Statistics & Data Analysis* **51**, 3983-3998
- [9] Su, Steve & Wuertz, Diethelm & King, Robert & Dean, Benjamin (2012) Fitting single and mixture of generalised lambda distributions (RS and FMKL) using various methods: in R GLDEX <https://cran.r-project.org/web/packages/GLDEX/GLDEX.pdf>
- [10] Su, Steve (2007) Fitting single and mixture of generalized lambda distribution to data via discretized and maximum likelihood methods: GLDEX in R *Journal of Statistics Software* **21**, No. 9, 1-17
- [11] Stephens, Michael A. (1979) *The Anderson-Darling statistic* Technical report no. 39, Stanford University Department of Statistics

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Julius Pau,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose "Üldistatud lambdaajaotus: rakendus varakindlustuse kahjudele", mille juhendaja on Tõnu Kollo,

1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 15.05.2017