UNIVERZITET U BEOGRADU

MATEMATIČKI FAKULTET

Ivana D. Ilić

OCENJIVANJE INDEKSA REPA RASPODELE KORIŠĆENJEM NEKOMPLETNIH UZORAKA

doktorska disertacija

Beograd, 2012

UNIVERSITY OF BELGRADE

MATHEMATICAL FACULTY

Ivana D. Ilić

ON TAIL INDEX ESTIMATION USING SAMPLES WITH MISSING OBSERVATIONS

Doctoral Dissertation

Belgrade, 2012

Mentor: Prof. dr Pavle Mladenović, Univerzitet Beograd, Matematički fakultet.

Članovi komisije:

1. Prof. dr Slobodanka Janković, Univerzitet Beograd, Matematički fakultet.

2. Prof. dr Ljiljana Petrović, Univerzitet Beograd, Ekonomski fakultet.

Datum odbrane:

____

Zahvalnost

Posebnu zahvalnost dugujem svom mentoru Prof. dr Pavlu Mladenoviću od koga potiče osnovna ideja na kojoj se baziraju dobijeni rezultati i koji je svojim korisnim primedbama i komentarima doprineo boljem kvalitetu teze.

Zahvaljujem se Prof. dr Slobodanki Janković i Prof. dr Ljiljani Petrović na ukazanim propustima u tekstu doktorata i na savetima koji su značajno odredili konačnu verziju teze.

Takodje, zahvaljujem se koleginici Jeleni Višnjić, koja mi je pružala stalnu podršku i pomogla u završnoj fazi rada. Za tehničku pomoć i konačni izgled korica doktorata veliko hvala mojim dragim prijateljma Snežani i Rodoljubu Avramović. Zahvaljujem se roditeljima i bratu na moralnoj i intelektualnoj podršci. Posebno hvala mom suprugu i kćerkama na nesebičnoj ljubavi i razumevanju.

Naslov teze:

OCENJIVANJE INDEKSA REPA RASPODELE KORIŠĆENJEM NEKOMPLETNIH UZORAKA

Abstrakt:

Teza obrađuje ocenjivanje indeksa pravilne promenljivosti na nekompletnom uzorku zavisnih slučajnih veličina sa raspodelom teškog repa. Pod pretpostavkom ekstremalne zavisnosti dokazuje se konzistentnost ocene geometrijskog tipa, kao i konzistentnost i asimptotska normalnost Hilove ocene. Navode se primeri procesa koji zadovoljavaju zahteve dokazanih teorema i na kojima se mogu primeniti dobijeni rezultati.

Ključne reči:

Hilova ocena, nekompletni uzorci, indeks repa raspodele, ocena geometrijskog tipa, ekstremalna zavisnost.

Naučna oblast:

Matematika

Uža naučna oblast:

Verovatnoća i statistika

UDK broj:
519.21
519.2

AMS klasifikacija:
62 G 32
60 G 70

Thesis title:

ON TAIL INDEX ESTIMATION USING SAMPLES WITH MISSING OBSERVATIONS

Abstract:

For the sequence of heavy-tailed and possibly dependent random variables with the missing observations the estimation of the tail-index is considered. Under minimal but verifiable assumption of "extremal dependence" we proved the consistency of geometric-type estimator (Brito and Freitas, 2003). We extended results from Mladenovic and Piterbarg (2008) and proved the consistency and the asymptotic normality of the Hill estimator. Illustrative examples are provided.

Keywords:

Hill estimator, incomplete samples, tail index, geometric-type estimator, extremal dependence.

Scientific field:

Mathematics

Specialized scientific field:

Probability and Statistics

UDK number:
519.21
519.2

AMS classification:
62 G 32
60 G 70

"Creativity is the ability to introduce order into the randomness of nature."

Sadržaj

Posvećeno mojoj majci...

Poglavlje 1

UVOD

Centralna tema teze je ocenjivanje indeksa pravilne promenljivosti α koji figuriše u izrazu funkcije raspodele teškog repa. Ocenjivanje se vrši pod pretpostavkom da nam nisu dostupni svi podaci i formira se model nekompletnog uzorka. Takodje, zahteva se da važi *ekstremalna zavisnost* medju elementima uzorka koja predstavlja tip zavisnosti u domenu visokih kvantila.

Uvodno poglavlje započinjemo objašnjavanjem osnovnih elemenata na kojima se zasniva teorija ekstremnih vrednosti, kao što je pojam teškog repa i pojam ekstremnog dogadjaja.

Drugo poglavlje je posvećeno klasičnoj teoriji ekstremnih vrednosti, osnovnim postulatima i teoremama na kojima zasnivamo kasnije dobijene rezultate.

U trećem poglavlju bavimo se tipovima ocena Paretovog indeksa α, vršimo komparaciju njihovih svojstava i uporedjujemo asimptotsko ponašanje.

Naredno poglavlje analizira tipove i oblike nekompletnih podataka i opisuje načine kojima se tretiraju ovakvi tipovi uzoraka.

Peto poglavlje predstavlja centralni deo teze i daje nove rezultate kao i primere procesa koji zadovoljavaju pretpostavljene uslove i na kojima su teoreme praktično primenljive. Posmatra se nekompletan niz medjusobno zavisnih slučajnih veličina koje imaju zajedničku raspodelu teškog repa. Uz pretpostavku ekstremalne zavisnosti, dokazujemo konzistentnost ocene geometrijskog tipa, kao i konzistentnost i asimptotsku normalnost Hilove ocene. Koncepte dokaza u osnovi baziramo na Hilovom radu (Hill (2010)), radu Brito i Freitas (2003), kao i na radu Mladenović i Piterbarg (2008).

Poslednje poglavlje posvećeno je opštem zaključku teze i diskusiji rezultata.

1.1 Teški repovi

Koji repovi raspodela se mogu smatrati teškim? Do nedavno, nije postojala potpuno precizna definicija raspodela koje imaju teške repove. Različiti autori su koristili izraze poput: ,,*teški rep*", ,,*debeo rep*", ,,*Paretova raspodela*", ,,*sub-eksponencijalna raspodela*" itd., često pod tim podrazumevajući stabilne zakone ili Studentovu t-raspodelu.

Prema definiciji kod Reznika (Resnick (1997)), reći ćemo da je raspodela koja ima teški rep data sa:

$$1 - F(x) = x^{-\alpha} L(x), \quad x > 0, \tag{1.1}$$

gde je nepoznata funkcija L sporo veličina u beskonačnosti:

$$\lim_{t \to \infty} \frac{L(tx)}{L(t)} \to 1 \quad \text{za } x > 0. \tag{1.2}$$

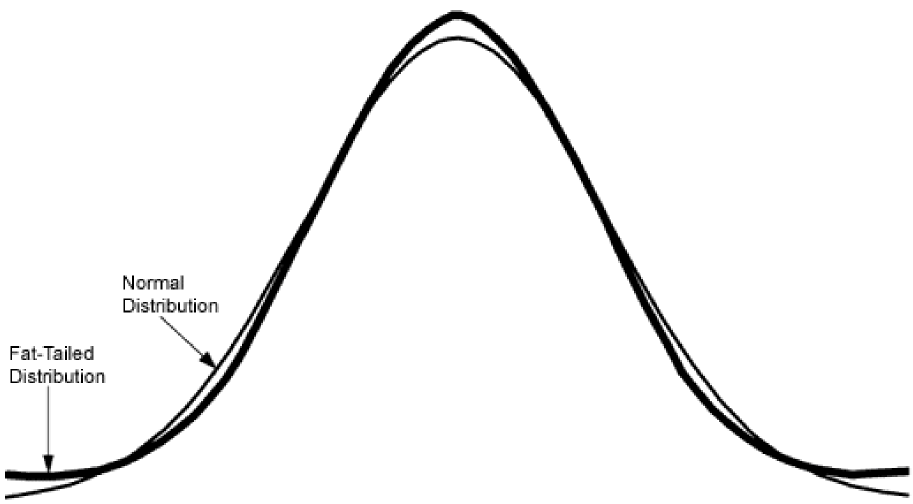
Skup svih sporo promenljivih funkcija u beskonačnosti označićemo sa $\mathcal{PP}_0$ (pogledati **Definiciju 2.6**).

Konstanta $\alpha > 0$ koja se pojavljuje u (1.1) naziva se ,,*indeks repa raspodele*", ,,*indeks pravilne promenljivosti*", ,,*Paretov indeks*" ili ,,*parametar oblika repa*". Ona predstavlja glavnu karakteristiku koja se mora uzeti u obzir prilikom opisivanja debljine repa raspodele. Podaci sa raspodelom koja zadovoljava (1.1) se mogu naći u širokom spektru najrazličitijih oblasti, kao što su: osiguranje, biznis, finansije, industrija, telekomunikacije, saobraćaj, ekonomija, sociologija i geologija. Raspodele sa takvim repovima imaju značajnu ulogu u modeliranju ekstremnih dogadjaja, što je predmet proučavanja teorije ekstremnih vrednosti o čijim rezultatima će biti reči u narednom poglavlju. Rep raspodele verovatnoća neke slučajne veličine odredjen je verovatnoćom da slučajna veličina uzima vrednosti veće od nekog datog praga. Ukoliko je raspodela te veličine normalna ili eksponencijalna ta verovatnoća veoma brzo opada ka nuli, s obzirom da odgovarajuća gustina raspodele teži ka nuli eksponencijalnom brzinom. Medjutim, ukoliko se radi o raspodelama koje zadovoljavaju (1.1), odnosno ako su u pitanju raspodele *Pareto*-tipa, kako se iz same formule za rep raspodele može videti, verovatnoća uzimanja velikih vrednosti opada stepenom brzinom, dakle sporije nego u slučaju normalne raspodele. Zato se kaže da je rep raspodele koja zadovoljava (1.1) "*teži*" ili "*deblji*" od repa normalne raspodele. Na *slici 1.1* se može videti razlika u repovima izmedju normalne i neke "*teške*" raspodele koja zadovoljava (1.1). *Slika 1.2* još preciznije pokazuje tu razliku. Na berzi su verovatnoće ekstremnih dogadjaja mnogo veće nego što bi se moglo opisati normalnom distribucijom, što opravdava široku upotrebu raspodela debelih repova prilikom modeliranja retkih dogadjaja u pomenutim domenima.

Primećeno je da postoje neke tipične zajedničke osobine podataka koji se pojavljuju u navedenim oblastima: oni su medjusobno zavisni, teških repova i ekstremni dogadjaji se često pojavljuju u klasterima. Moderna statistika se intenzivno bavi ovakvim raspodelama naročito u oblasti finansija pokušavajući da reši problem predvidjanja velikih krahova na berzi.

1.2 Black swans

Nakon teske finansijske krize 2008. godine investitori Wall Street-a pokušavali su da dokuče da li je postojala mogućnost da se zaštite od iznenadnog kolapsa i na



SLIKA 1.1: Poredjenje normalne raspodele sa raspodelom teškog repa

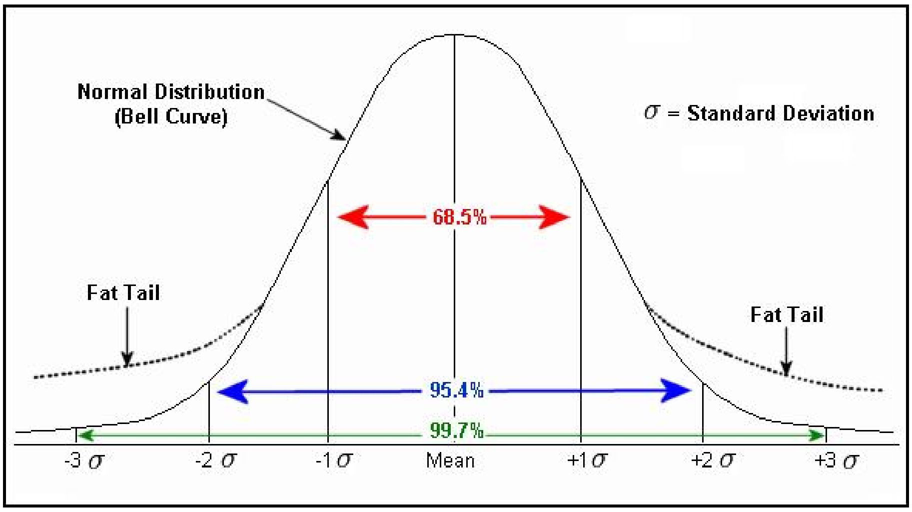
koji način ce moći da se predvidi buduće pojavljivanje ekstremuma sa negativnim efektom ili *crnih labudova*-nepoželjnih dogadjaja koji uzrokuju masovne gubitke.

Pojam *"black swans"* prvi put susrećemo kod esejiste Nasima Taleba (Nassim Taleb (2007)) čiji se rad zasniva na pojmovima slučajnosti i verovatnoće. Crni labudovi predstavljaju ekstremne dogadjaje koji mogu imati Kao posledicuce širokih razmera u različitim domenima. Tipično se pojavljuju bez neke odredjene učestalosti, tako da kao takvi, navodi Taleb (2007), mogu lako izmaknuti primeni standardnih analitičkih metoda i verovatnoća njihovog pojavljivanja može greškom biti procenjena kao jako mala. U takvim slučajevima prihvata se procena da su takvi dogadjaji retki i da se mogu pojaviti tek jednom u million godina. Otud i potiče naziv crni labud kao matafora za redak, iznenadjujuć dogadjaj. Taleb u svojoj knjizi navodi neke pojave koje se mogu smatrati crnim labudovima: internet, personalni kompjuter, prvi svetski rat i napad 11. septembra. Pojam *"black swan"* potiče iz Latinske izreke ,,*rara avis in terris nigroque simillima cygno*", što u prevodu znači ,,*retka ptica na zemlji, veoma slična crnom labudu.*"

U vreme kada je nastala ova izreka pretpostavljalo se da crni labudovi ne postoje.

Analogija koja je preuzeta govori o krhkosti bilo kog sistema mišljenja. Ukoliko je skup odredjenih zaključaka donesen na osnovu nekih fundamentalnih postulata od kojih se pokazalo da je bar jedan netačan, u ovom slučaju, pojavljivanje crnog labuda, onda se svi drugi zaključci koji su proistekli iz te teorije dovode u ozbiljno pitanje. Ova fraza se koristila u Londonu jos u 16. veku i označavala je nemoguć dogadjaj. U to vreme se znalo samo za labudove sa belim perjem. Nemačka ekspedicija je 1697. godine u Zapadnoj Australiji otkrila crnog labuda. Opet se može napraviti analogija, jer se termin i kasnije zadržao i naznačavajući da se pretpostavljena nemogućnost pojave nekog dogadjaja može kasnije opovrgnuti.

Ono što Taleb naziva crnim labudom ima tri atributa: najpre, to je podatak autsajder (eng. *outlier*) s obzirom da leži van regularnih očekivanja i ništa u prošlim dogadjanjima ne može ukazivati da postoji verovatnoća njegovog pojavljivanja; drugo, nosi sa sobom ekstremno veliki uticaj; i treće, uprkos svom specijalnom statusu čovek ga ipak može na osnovu svog iskustva i logike objasniti pa i predvideti. Dakle: retkost, veliki uticaj i retrospektivna predvidljivost. Crnim labudovima se, po Talebu, mogu objasniti kroz istoriju progresi odredjenih ideja i religija, globalna dešavanja u svetskoj ekonomiji, kao i pravci u naseg ličnog postojanja.



Slika 1.2: Normalna raspodele i raspodela debelog repa

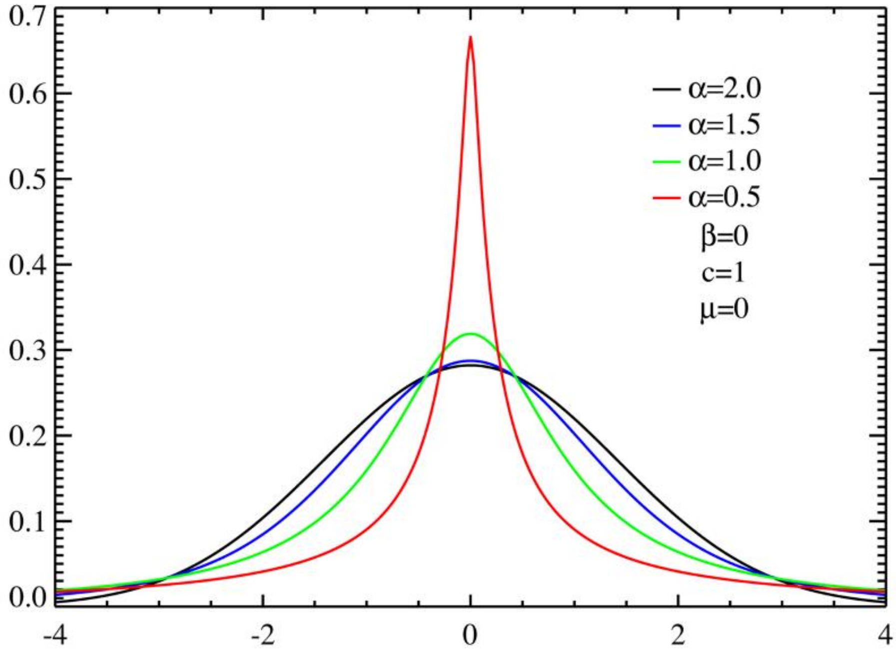
1.3 Metodi detektovanja teških repova

Postoji veliki broj procedura pomoću kojih se može proveriti da li je u pitanju raspodela teškog repa. Recimo, formiranje takozvane QQ ravni (eng. *quantile-quantile plot*). Podsetimo se da kvantil funkcije F definiše zapravo inverz F^{-1}. Dakle, ukoliko je p takvo da je $0 < p < 1$, kvantil reda p predstavlja ono x za koje važi da je $F(x) = p$. Ukoliko F ima tačke diskontinuiteta definišemo inverz funkcije raspodele kao $F^{-1} = \inf\{t : F(t) \geq y\}$, o čemu će biti više reči u narednim poglavljima. U grafikon se ubacuju kvantili empirijske funkcije raspodele nasuprot kvantilima pretpostavljene funkcije raspodele. Ukoliko je grafik približno prava linija može se sa velikom verovatnoćom reći da se radi o slučaju teškog repa (pogledati za više detalja Embrechts i dr. (1997, str. 292-293)). Na *slici 1.5* prikazana je QQ ravan procene oblika raspodele, gde je pretpostavljena raspodela *Pareto*-tipa sa ocenjenim indeksom $\alpha = 0,708$ pomoću neke od postojećih metoda ocene parametara repa raspodele (pogledati treće poglavlje za više detalja o ocenjivanju). Druga procedura koju ćemo pomenuti je bazirana na srednjem ekscesu funkcije:

$$M(x) = E\{X - x|X > x\}$$

(pogledati Embrechts i dr. (1997, str. 296-355) i Novak (2002)).

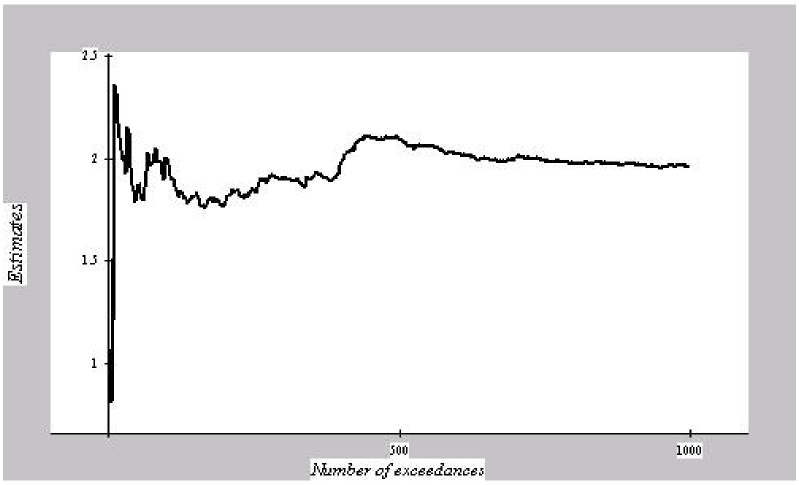
Ako je grafik funkcije $M(x)$ linearan, može se zaključiti da se radi o raspodeli sa debelim desnim repom. Još jedna procedura koja se koristi je i uporedjivanje histograma raspodele koju ispitujemo sa histogramom odgovarajuće normalne raspodele sa istom srednjom vrednošću i disperzijom (pogledati za više detalja Luenberger (1998)). Naravno, u slučaju podataka se raspodelom teškog repa, standardna devijacija ne može više biti indikator ekstremnih dešavanja i teško da se treba uzeti u obzir kao mera rizika. Podsetimo se da standardna devijacija opisuje prosečno odstupanje od srednje vrednosti. U slučaju teških repova se može desiti da samo jedan element bude iste veličine kao suma svih ostalih elemenata uzorka.



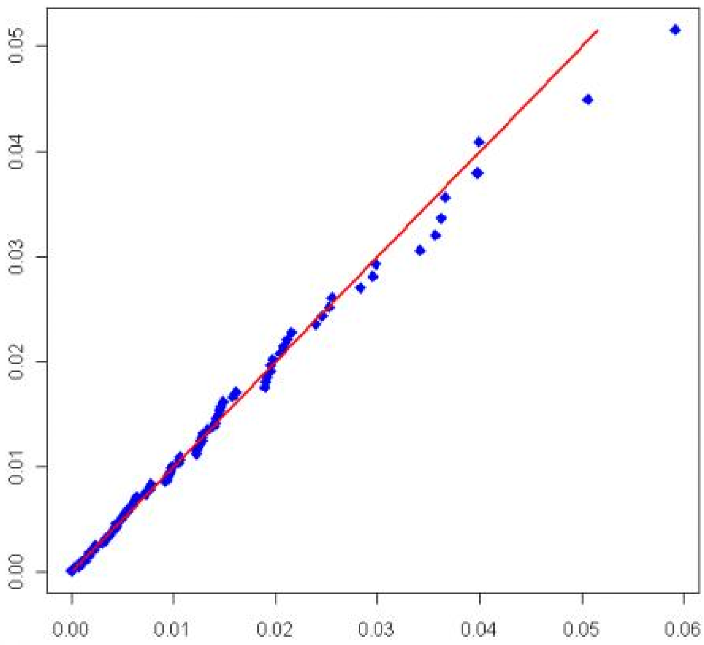
SLIKA 1.3: Poredjenje raspodela u odnosu na različite parametre oblika

Ono što nas posebno interesuje je metod odredjivanja debljine repa putem odgovarajućih ocena Paretovog indeksa α. Ukoliko je pouzdanom ocenom ocenjeno da je $0 < \hat{\alpha} < 2$, vrlo je verovatno da se radi o debelom desnom repu u smislu definicije (1) i tada postoje samo momenti reda manjeg od α, tj. raspodela ima beskonačnu varijansu. S druge strane, ako je procenjeno odgovarajućom ocenom da je $\hat{\alpha} \geq 2$ radi se o lakom repu. Pogledati *sliku 1.3* za grafički prikaz pomenutih razlika u debljini repova. Normalnoj raspodeli odgovara parametar $\alpha = 2$.

Problem pouzdanog zaključivanja o indeksu repa i ekstremnim kvantilima je bio predmet istraživanja jos od 60-tih godina (Fama, Roll (1968)). Do nedavno, zaključci su bili veoma pesimistični; videti ,,*Hill's horor ravan*" (Resnick (1997)) ili ,,*MLE horror plot*" (Embrechts i dr. (1997)), ukazujući na poteškoće prilikom ocenjivanja i donošenja konačnih odluka o vrednosti parametra oblika repa raspodele. Na *slici 1.4* predstavljena je Hilova *horor* ravan na osnovu koje se može oceniti



Slika 1.4: Hilova horor ravan



Slika 1.5: QQ ravan uopštene *Pareto*-raspodele sa parametrom $\alpha = 0,708$

vrednost traženog indeksa α, u delu grafika gde se pokazuje odredjena stabilnost, odnosno linearnost.

U narednom poglavlju objasnićemo matematičke osnove teških repova i bavićemo se fundamentalnim postulatima teorije ekstremnih vrednosti.

Poglavlje 2

TEORIJA EKSTREMNIH VREDNOSTI

2.1 Granične raspodele za maksimume

Teorija ekstremnih vrednosti (eng. *Extreme Value Theory*) pruža matematički okvir unutar koga možemo da formalizujemo opservacije i zaključke o ponašanju funkcije raspodele u levom ili desnom repu. Kritična pitanja o kojima smo govorili u uvodnom poglavlju a koja se odnose na verovatnoću dogadjanja kraha na berzi, ekonomske krize ili bilo kakvog kolapsa u različitim domenima, podrazumevaju poznavanje očekivanog ponašanja na krajevima funkcije raspodele u statističkom smislu. Teorija ekstremnih vrednosti nam omogućava da pomoću ekstremnih opservacija izmerimo debljinu repa. Dobijene rezultate možemo proširiti na delove još neregistrovanih empirijskih podataka. Na taj način se može simulirati funkcija raspodele teškog repa i u velikoj meri poboljšati prognoziranje teško predvidivih i nestabilnih procesa u ekonomiji, osiguranju, finansijamna, hidrologiji i drugim oblastima od interesa. Postoje neke zajedničke osobine ekstremnih dogadjaja: obično imaju značajan uticaj u konačnom bilansu (velike isplate nakon prirodnih katastrofa u oblasti osiguranja), teško su predvidivi (često su uzrokovani prirodnim procesima) i retki su.

Prema Centralnoj Graničnoj Teoremi (eng. *Central Limit Theorem*) normalna raspodela predstavlja graničnu funkciju raspodele uzoračkih proseka. Slična ideja se pojavljuje i kada govorimo o uzoračkim ekstremumima. Klasa raspodela ekstremnih vrednosti poklapa se sa mogućim graničnim raspodelama uzoračkog maksimuma. Kao što ćemo videti, postoji neka vrsta analogije izmedju Centralne Granične Teoreme i Fisher-Tippet teoreme koja predstavlja osnovnu teoremu Teorije ekstremnih vrednosti i prvi značajni rezultat u ovoj oblasti.

Neka je $\{X_t\} = \{X_t : 1 \leq t \leq n\}$ niz medjusobno nezavisnih slučajnih veličina sa istom funkcijom raspodele $F(x) = P\{X \leq x\}$.

Definišimo slučajne veličine

$$M_n = \max\{X_1, X_2, ..., X_n\}$$

i

$$m_n = \min\{X_1, X_2, ..., X_n\}.$$

Teorija ekstremnih vrednosti nizova slučajnih veličina bavi se ponašanjem raspodela ovih slučajnih veličina kad $n \to \infty$. U slučaju razmatranja maksimuma, procenjujemo da li postoje nizovi realnih konstanti $a_n > 0$ i b_n, $n \in N$ takvi da važi

$$\lim_{n\to\infty} P\{\frac{M_n - b_n}{a_n} \leq x\} = G(x) \tag{2.1}$$

za svaku tačku neprekidnosti x neke nedegenerisane funkcije raspodele $G(x)$. To znači da će, ukoliko takvi nizovi postoje, $\frac{M_n-b_n}{a_n}$ konvergirati u raspodeli ka $G(x)$ kad $n \to \infty$. U tom slučaju nizove a_n i b_n zovemo normirajućim konstantama, a funkciju raspodele $G(x)$ graničnom raspodelom linearno normiranog maksimuma M_n.

Veličine M_n i m_n predstavljaju ekstremne vrednosti datog niza, i u mnogim primenama nalazimo neophodnost odredjivanja njihove granične raspodele. Recimo, može da nas zanima verovatnoća da će neka buduća opservacija prevazići odredjenu vrednost u toku posmatranog perioda. Na primer, nivo reke iznad odredjene granice može izazvati probleme, buka veća od uobičajene pri radu mašine može značiti kvar, jačina vetra iznad dozvoljene može takodje predstavljati odredjeni rizik, količina šeera u krvi iznad dozvoljene može imati negativan uticaj na stanje organizma, pa se u ovim slučajevima pokazalo važno odredjivanje granične raspodele maksimuma. Slične primere možemo navesti i za minimume (gde se odredjuju verovatnoće da će buduća opservacija biti ispod odredjenog nivoa), pri čemu se slični rezultati koje ćemo navesti za maksimume mogu dobiti i za minimume na osnovu jednakosti:

$$\min\{X_1, X_2, ..., X_n\} = -\max\{-X_1, -X_2, ..., -X_n\}. \tag{2.2}$$

U ovom poglavlju ćemo se baviti nizovima nezavisnih slučajnih veličina sa istom funkcijom raspodele $F(x)$ i problemima vezanim za asimptotsko ponašanje maksimuma. Neka je $u_n = a_n x + b_n$. Tada važi

$$\begin{aligned} P\{\frac{M_n - b_n}{a_n} \leq x\} &= P\{M_n \leq a_n x + b_n\} \\ &= P\{M_n \leq u_n\} \\ &= P\{\max\{X_1, X_2, ..., X_n\} \leq u_n\} \\ &= P\{X_1 \leq u_n\}P\{X_2 \leq u_n\} \cdots P\{X_n \leq u_n\} \\ &= F^n(u_n). \end{aligned}$$

Ako za nizove konstanti $a_n > 0$ i b_n, $n \in N$ i neku nedegenerisanu funkciju raspodele $G(x)$ važi relacija (2.1), onda kažemo da funkcija raspodele F pripada oblasti privlačenja za maksimume funkcije raspodele G. Skup svih takvih funkcija označavaćemo sa $D(G)$.

Dalje navodimo, bez dokaza, nekoliko poznatih teorema koje čine osnovu teorije ekstremnih vrednosti nizova nezavisnih slučajnih veličina, koje mogu dati odgovor na pitanja: koje funkcije $G(x)$ se mogu pojaviti kao granične funkcije raspodele linearno normiranog maksimuma od n nezavisnih slučajnih veličina sa istom funkcijom raspodele $F(x)$, kako za datu zajedničku funkciju raspodele $F(x)$ niza nezavisnih slučajnih veličina odrediti da li postoji i koja je granična funkcija raspodele linearno normiranog maksimuma M_n i kako odrediti normirajuće konstante a_n i b_n.

Teorema 2.1. *[**Fisher-Tippett teorema, granični zakoni za maksimume**] Neka je $\{X_t\}$ niz medjusobno nezavisnih identično raspodeljenih slučajnih veličina (engl. independent identically distributed random variables). Ako postoje normirajuće konstante $a_n > 0$ i $b_n \in \mathbb{R}$ za koje važi da*

$$a_n^{-1}(M_n - b_n) \xrightarrow{d} G, \tag{2.3}$$

tada G pripada jednom od sledećih tipova funkcija raspodele:

Freseova (Frechét)*:*

$$\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0, \\ exp\{-x^{-\alpha}\}, & x > 0, \alpha > 0 \end{cases}$$

Vejbulova (Weibull):

$$\Psi_\alpha(x) = \begin{cases} exp\{-(-x)^{\alpha}\}, & x \leq 0, \alpha > 0 \\ 1, & x > 0, \end{cases}$$

Gumbelova (Gumbel)*:*

$$\Lambda(x) = exp\{-e^{-x}\}, \qquad x \in \mathbb{R}. \quad \blacksquare$$

Iako se, u smislu matematičkog modelovanja tipovi Λ, Φ_α i Ψ_α veoma razlikuju, sa matematičke tačke gledišta oni su jako bliski. Zaista, primetimo da za $X > 0$ važe sledeće ekvivalencije:

$$X \text{ ima df } \Phi_\alpha \Leftrightarrow \ln X^\alpha \text{ ima df } \Lambda \Leftrightarrow -X^{-1} \text{ ima df } \Psi_\alpha,$$

gde smo sa *df* označili funkciju raspodele (eng. *distribution function*).

Definicija 2.2 (**Raspodele ekstremnih vrednosti**)**.** Funkcije raspodele Λ, Φ_α i Ψ_α koje su prezentovane u teoremi zovu se standardne raspodele ekstremnih vrednosti, a odgovarajuće slučajne promenljive-standardne ekstremalne slučajne promenljive. Takodje, funkcije raspodele istog tipa kao neka od Λ, Φ_α i Ψ_α raspodela nazivaju se raspodele ekstremnih vrednosti, a odgovarajuće slučajne veličine se zovu ekstremalne slučajne veličine.

Definicija 2.3 (Funkcije raspodele istog tipa). Za slučajne promenljive X i Y kažemo da imaju isti tip raspodele ako postoje realni brojevi $a > 0$ i b takvi da je

$$Y =^d aX + b.$$

To znači da se funkcija raspodele za Y može dobiti pomoću funkcije raspodele za X linearnom transformacijom argumenta. Dakle, može se zaključiti da je granični zakon u (2.3) jedinstven do na afine transformacije. Ukoliko se kao granična vrednost pojavi $G(ax+b)$, onda je granična vrednost takodje i $G(x)$ samo za druge normirajuće konstante. Zaista, ako je

$$\lim_{n\to\infty} P\{a_n^{-1}(M_n - b_n) \leq x) = G(ax+b),$$

onda je $G(x)$ granična vrednost pri jednostavnoj promeni normirajućih konstanti:

$$\lim_{n\to\infty} P\{\widetilde{a}_n^{-1}(M_n - \widetilde{b}_n) \leq x) = G(x),$$

gde je $\widetilde{a}_n = a_n/a$ i $\widetilde{b}_n = b_n - ba_n/a$.

Primer 2.1. ***(Maksimum Košijevih slučajnih veličina)*** *Neka je $\{X_t\}$ niz medjusobno nezavisnih slučajnih veličina sa identičnom standardnom Košijevom funkcijom raspodele. Gustina ove raspodele data je sa*

$$f(x) = (\pi(1+x^2))^{-1}, \quad x \in \mathbb{R}.$$

Prema Lopitalovom (L'Hospital) pravilu dobijamo

$$\lim_{n\to\infty} \frac{\overline{F}(x)}{(\pi x)^{-1}} = \lim_{n\to\infty} \frac{f(x)}{\pi^{-1}x^{-2}} \lim_{n\to\infty} \frac{\pi x^2}{\pi(1+x^2)} = 1,$$

što daje $\overline{F}(x) \sim (\pi x)^{-1}$. Odavde sledi da je

$$\begin{aligned} P\{M_n \leq \frac{nx}{\pi}\} &= (1 - \overline{F}(\frac{nx}{\pi}))^n \\ &= (1 - \frac{1}{n}(\frac{1}{x} + \mathrm{o}(1)))^n \\ &\to exp\{-x^{-1}\} = \Phi_1(x), \quad x > 0. \quad \blacksquare \end{aligned}$$

Teorema 2.1 predstavlja početak razvoja teorije ekstremnih vrednosti kao jedan od centralnih rezultata u teoriji verovatnoće i statistike. Granične vrednosti za maksimume definisali su Fišer i Tipet (1928). Prvi precizni dokaz formulisao je Gnedenko (1943). De Haan (1971) takodje dolazi do istih zaključaka na temu graničnih vrednosti maksimuma slučajnih veličina ali uz primenu pravilne promenljivosti, što će biti od izuzetne važnosti za dalji razvoj teorije ekstremnih vrednosti.

2.2 Maksimalni domeni privlačenja raspodela EV (ekstremnih vrednosti)

U prethodnoj sekciji definisali smo raspodele ekstremnih vrednosti kao granične raspodele maksimuma identično raspodeljenih i nezavisnih slučajnih veličina.

Definicija 2.4 (**Maksimalni domen atrakcije**). Za slučajnu veličinu X (ili za funkciju raspodele te slučajne promenljive) kažemo da pripada maksimalnom domenu privlačenja funkcije raspodele ekstremnih vrednosti G, ukoliko postoje konstante $a_n > 0$ i $b_n \in \mathbb{R}$ takve da važi

$$a_n^{-1}(M_n - b_n) \xrightarrow{d} G. \tag{2.4}$$

To zapisujemo $X \in MDA(G)$ ili $F \in MDA(G)$.

Primetimo da je (2.4) ekvivalentno sa

$$\lim_{n\to\infty} P(M_n \leq a_n x + b_n) = \lim_{n\to\infty} F^n(a_n x + b_n) = G(x). \tag{2.5}$$

Definicija 2.5 (**Parametri položaja i razmere**). Gumbelova funkcija raspodele sa parametrima položaja μ i razmere σ odredjena sa $\Lambda(\frac{x-\mu}{\sigma})$, Frešeova raspodela sa paremetrima položaja i razmere sa $\Phi_\alpha(\frac{x-\mu}{\sigma})$ i Vejbulova sa $\Psi_\alpha(\frac{x-\mu}{\sigma})$.

Neka je

$$x_F = sup\{t : F(t) < 1\}. \tag{2.6}$$

U daljem tekstu ovu tačku zvaćemo krajnja desna tačka domena funkcije raspodele F. U narednim paragrafima opisaćemo domene privlačenja funkcija raspodela ekstremnih vrednosti koje su opisane Teoremom 2.1.

2.2.1 Maksimalni domen privlačenja funkcije $\Phi_\alpha(x)$

U ovoj sekciji karakterišemo maksimalni domen privlačenja funkcije Φ_α, za $\alpha > 0$. Prema Tejlorovoj formuli, važi da je

$$1 - \Phi_\alpha(x) = 1 - exp\{-x^{-\alpha}\} \sim x^{-\alpha}.$$

Kao što se može uočiti, verovatnoća uzimanja velikih vrednosti slučajne veličine koja ima Frešeovu raspodelu opada stepenom brzinom.

Definicija 2.6 (**Pravilno promenljive funkcije u beskonačnosti**). Neka je $\mathbb{R}$ skup realnih brojeva i $\mathbb{R}_+ = (0, \infty)$ skup pozitivnih realnih brojeva. Funkcija $F : \mathbb{R}_+ \to \mathbb{R}_+$ je *pravilno veličina u beskonačnosti* ukoliko postoji broj $\rho \in \mathbb{R}$ tako da za svaki pozitivan broj x važi

$$\lim_{t\to\infty} \frac{F(tx)}{F(t)} = x^\rho. \tag{2.7}$$

Broj ρ se naziva *indeks* ili *eksponent* pravilne promenljivosti. Za funkciju F koja je pravilno veličina u beskonačnosti sa indeksom ρ, kažemo da je $\rho-$*veličina* i pišemo $F \in \mathcal{PP}_\rho$, gde oznaka $\mathcal{PP}_\rho$ predstavlja skup svih pravilno promenljivih funkcija sa eksponentom promenljivosti ρ. Skup svih sporo promenljivih funkcija u beskonačnosti koje predstavljaju podklasu klase $\mathcal{PP}_\rho$ definisali smo u prvom poglavlju i njihov skup označili sa $\mathcal{PP}_0$ (slučaj kada je $\rho = 0$).

Pokazaćemo da se maksimalni domen atrakcije funkcije Φ_α sastoji od funkcija čiji je desni rep pravilno promenljiv sa indeksom $-\alpha$.

Sledeća teorema, koju navodimo bez dokaza, daje dovoljan uslov za pripadnost neke funkcije raspodele F maksimalnom domenu atrakcije funkcije ekstremnih vrednosti Φ_α.

Teorema 2.7. *[**Maksimalni domen privlačenja** $\Phi_\alpha(x)$] F pripada maksimalnom domenu atrakcije funkcije Φ_α, $\alpha > 0$, ako i samo ako je $\overline{F}(x) = x^{-\alpha}L(x)$ za neku sporo promenljivu funkciju L. Ako $F \in MDA(\Phi_\alpha)$, onda*

$$a_n^{-1}M_n \xrightarrow{d} \Phi_\alpha, \tag{2.8}$$

gde se normirajuće konstante a_n mogu odrediti na sledeći način:

$$\begin{aligned} a_n = F^{-1}(1-n^{-1}) &= inf\{x \in \mathbb{R} : F(x) \geq 1 - n^{-1}\} \\ &= inf\{x \in \mathbb{R} : (1/\overline{F})(x) \geq n\} \\ &= (1/\overline{F})^{-1}(n), \end{aligned}$$

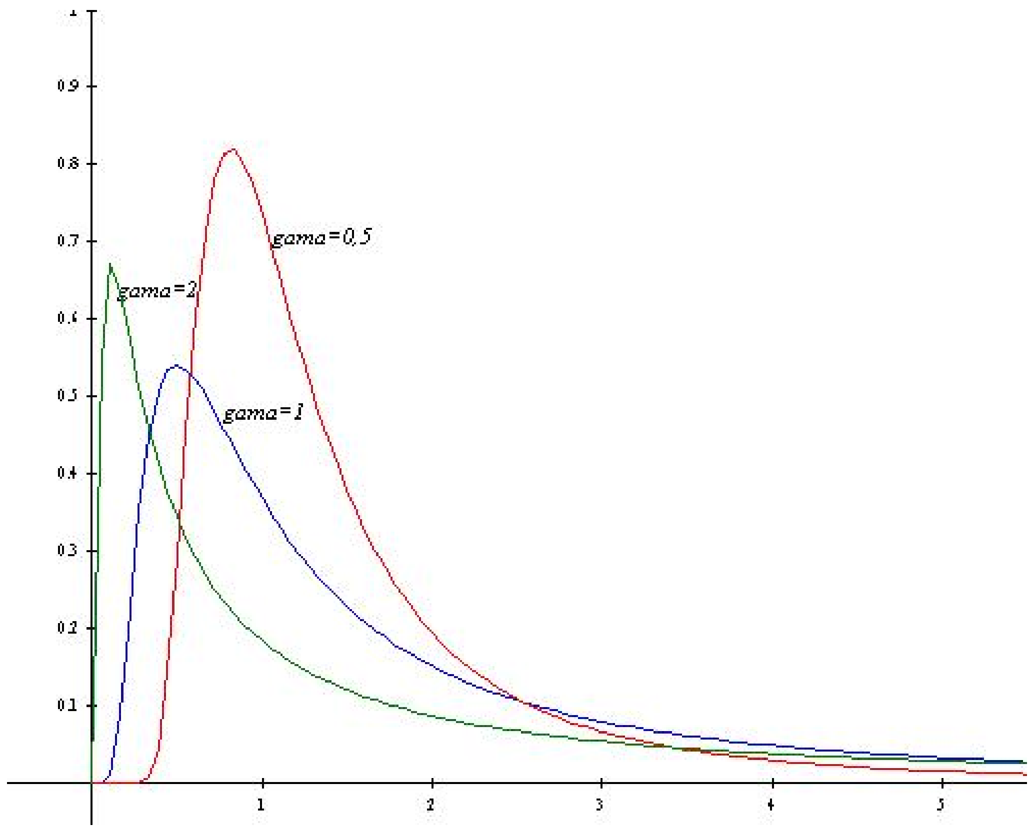
dok se konstante b_n mogu smatrati jednakim nuli. ■

Ovaj rezultat implicira da svaka funkcija raspodele $F \in MDA(\Phi_\alpha)$ ima beskonačnu krajnju desnu tačku $x_F = \infty$. Na osnovu Teoreme 2.6, zaključujemo da je

$$F \in MDA(\Phi_\alpha) \Leftrightarrow \overline{F} \in \mathcal{PP}_{-\alpha}.$$

Dakle, na ovaj način bi bila opisana karakterizacija $MDA(\Phi_\alpha)$.

Ova klasa obuhvata raspodele *veoma teških repova*, u smislu da je $E(X^+)^\delta = \infty$, za $\delta > \alpha$. Ova osobina ih čini veoma podesnim za modelovanje velikih isplata u osiguranju, velikih fluktuacija i skokova cena, itd.



SLIKA 2.1: Poredjenje Frešeovih raspodela za parametre $\gamma = 1/\alpha$, $\mu = 0$ i $\sigma = 1$

Slika 2.1 pokazuje Frešeove funkcije raspodele za različite parametre indeksa oblika $\gamma = 1/\alpha$. Na *slici 2.1* se može primetiti da za veće γ, odnosno za α bliže nuli, rep raspodele postaje sve deblji ako skoncentrišemo pažnju na oblast vrlo visokih kvantila, što znači da su veće verovatnoće retkih dogadjaja.

Primer 2.2 [***Raspodele Pareto-tipa***]

- *Paretova raspodela*
- *Košijeva raspodela*
- *Burova raspodela*

- *Stabilna raspodela sa eksponentom* $\alpha < 2$.

2.2.2 Maksimalni domen privlačenja funkcije $\Psi_\alpha(x)$

Ovaj paragraf opisuje maksimalnu oblast privlačenja Vejbulove funkcije raspodele $\Psi_\alpha(x) = exp\{-(-x)^{-\alpha}\}$, za $\alpha > 0$. Važna, ali ne tako očigledna činjenica je da funkcije raspodele koje su u domenu privlačenja Vejbulove raspodele ekstremnih vrednosti imaju konačnu krajnju desnu tačku x_F. Kao što je već napomenuto u paragrafu 2.1 ovog poglavlja, funkcije Φ_α i Ψ_α su blisko povezane, tačnije važi

$$\Psi_\alpha(-x^{-1}) = \Phi_\alpha(x).$$

Dakle, očekujemo da će i njihove oblasti privlačenja $MDA(\Phi_\alpha)$ i $MDA(\Psi_\alpha)$ pokazati odredjenu analogiju, što i pokazuje naredna teorema.

Teorema 2.8. *[**Maksimalni domen privlačenja** $\Psi_\alpha(x)$] Funkcija raspodele F pripada maksimalnom domenu privlačenja Vejbulove funkcije raspodele ekstremnih vrednosti $\Psi_\alpha(x)$, za $\alpha > 0$, ako i samo ako $x_F < \infty$ i $\overline{F}(x_F - x^{-1}) = x^{-\alpha}L(x)$ za neku sporo promenljivu funkciju L. Ako $F \in MDA(\Psi_\alpha)$ onda je*

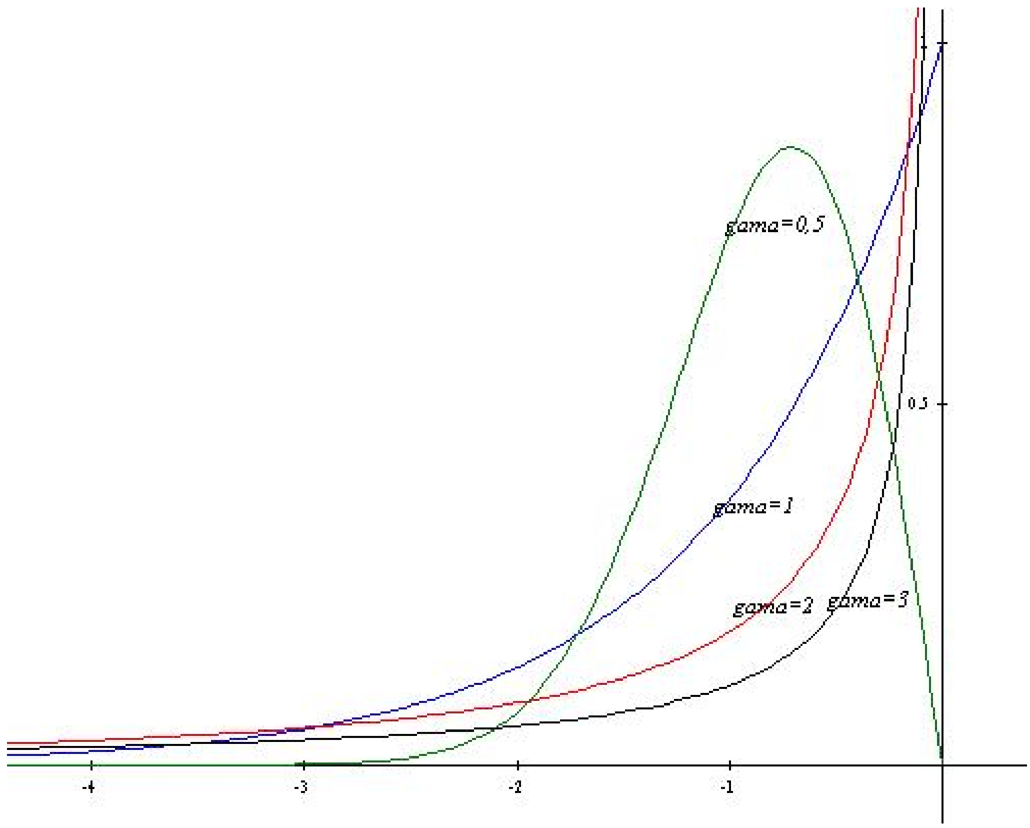
$$a_n^{-1}(M_n - x_F) \xrightarrow{d} \Psi_\alpha, \tag{2.9}$$

gde normirajuće konstante a_n mogu biti izabrane kao $a_n = x_F - F^{-1}(1 - n^{-1})$ i $b_n = x_F$. ■

Na osnovu Teoreme 2.8 zaključujemo da važi

$$F \in MDA(\Psi_\alpha) \Leftrightarrow x_F < \infty, \quad \overline{F}(x_F - x^{-1}) \in \mathcal{PP}_{-\alpha}.$$

Dakle, $MDA(\Psi_\alpha)$ se sastoji od funkcija raspodele F čiji je domen ograničen sa desne strane. Upravo zato one možda nisu najbolji izbor za modelovanje ekstremnih dogadjaja u osiguranju i finansijama, naročito zbog činjenice da je $x_F < \infty$. U praktičnim situacijama, kada postoji desna ograničenost vrednosti uzorka koja



SLIKA 2.2: Poredjenje Vejbulovihvih funkcija raspodela za različite parametre γ, $\mu = 0$ i $\sigma = 1$

je veoma visoka, nećemo želeti da koristimo model u kome je $x_F < \infty$, već će mnogo podesnije biti raspodele koje su u domenu Frešeove ili čak Gumbelove funkcije raspodele ekstremnih vrednosti (sto ćemo videti u paragrafu 2.2.3). Pogledati *sliku 2.2* koja uporedjuje Vejbulove funkcije raspodele za razne vrednosti parametra γ.

Primer 2.3 [***Funkcije raspodele koje pripadaju*** $MDA(\Psi_\alpha)$]

- *Uniformna raspodela na intervalu* $(0, 1)$
- *Beta raspodela*

- *Funkcija raspodele koja se u desnoj krajnjoj tački ponaša prema stepenom zakonu, tj. važi da je*

$$\overline{F} = K(x_F - x)^{\alpha}, \quad x_F - K^{-1/\alpha}x \leq x_F, \quad K, \alpha > 0.$$

2.2.3 Maksimalni domen privlačenja funkcije $\Lambda(x)$

Ovaj paragraf opisuje maksimalnu oblast privlačenja Gumbelove funkcije raspodele $\Lambda(x) = exp\{-exp\{-x\}\}$.

Sledeća teorema nam daje potrebne i dovoljne uslove pod kojima za neku funkciju raspodele važi da $F \in MDA(\Lambda)$.

Teorema 2.9. *[**Maksimalni domen privlačenja** $\Lambda(x)$] Označimo sa F funkciju raspodele i definišimo sledeću funkciju:*

$$H(x) = \frac{1}{1 - F(x)}. \tag{2.10}$$

Tada su sledeći uslovi ekvivalentni:

1. *$F \in MDA(\Lambda)$, tj. postoje nizovi konstanti $a_n > 0$ i $b_n \in \mathbb{R}$, $n \in \mathrm{N}$, tako da za svaki realan broj x važi da je*

$$\lim_{n\to\infty} F^n(a_n x + b_n) = exp\{-exp\{-x\}\} \tag{2.11}$$

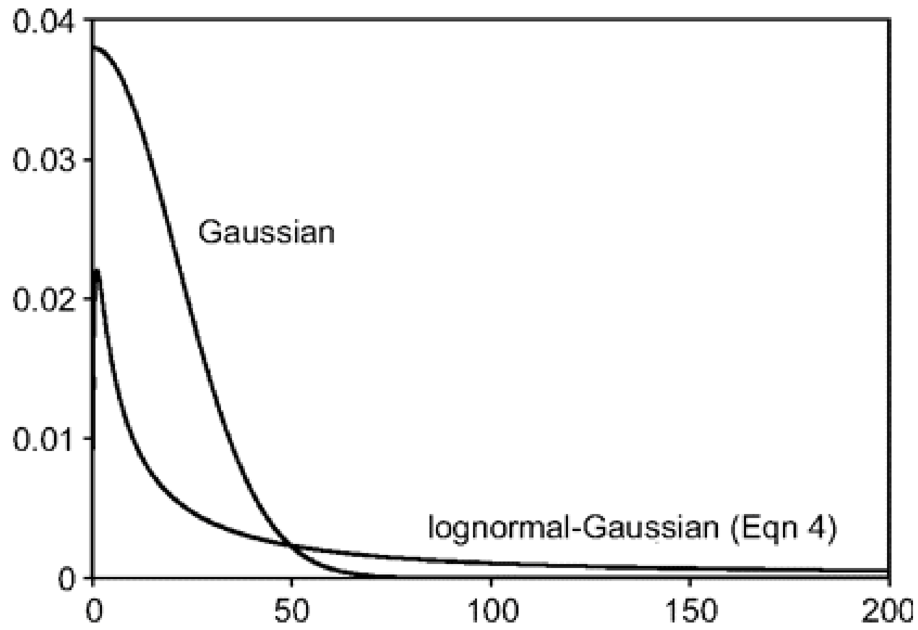
2. *Postoji funkcija $g : (c, x_F) \to \mathbb{R}_+$ takva da za svaki realan broj x važi da je*

$$\lim_{t\to\infty} \frac{1 - F(t + xg(t))}{1 - F(t)} = exp(-x) \tag{2.12}$$

3. *Postoji funkcija $a : (c, \infty) \to \mathbb{R}_+$ takva da za svaki realan broj x važi da je*

$$\lim_{t\to\infty} \frac{H^{-1}(tx) - H^{-1}(t)}{a(t)} = \ln(x). \quad \blacksquare \tag{2.13}$$

Na *slici 2.3* prikazana je razlika u debljini repova normalne i lognormalne raspodele, iako su obe raspodele predstavnici Gumbelovog maksimalnog domena atrakcije.



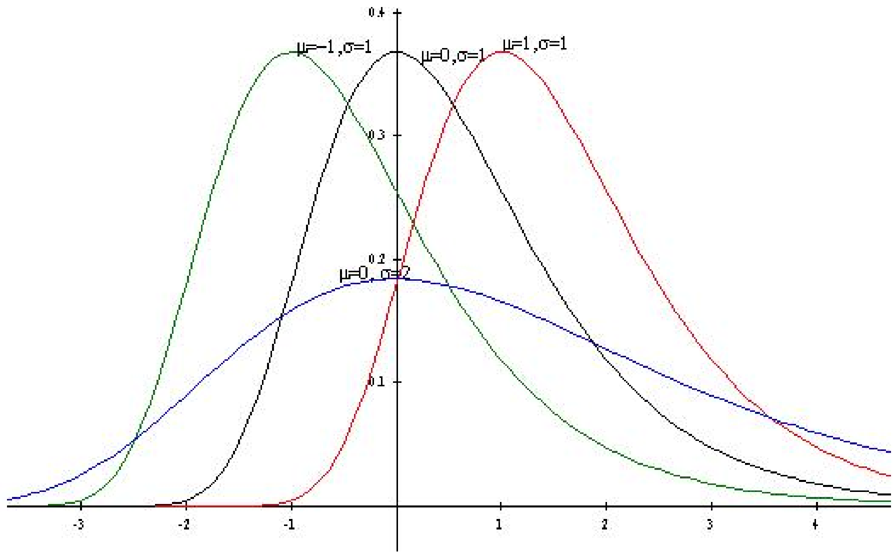
SLIKA 2.3: Poredjenje desnog repa normalne i lognormalne raspodele

Iz prethodnog teksta se može videti da $MDA(\Lambda)$ sadrži funkcije raspodele sa veoma različitim repovima koji se kreću od *srednje teških*, kao što je slučaj kod lognormalne raspodele, pa do *lakih repova*, kao što je slučaj kod normalne raspodele. Takodje su moguća oba slučaja što se tiče krajnje tačke $x_F < \infty$ i $x_F = \infty$. *Slika 2.4* prikazuje Gumbelove raspodele za različite parametre položaja i razmere.

Pogledati *sliku 2.4* za grafički prikaz Vejbulovih funkcija raspodele za različite parametre $\gamma = 1/\alpha$. Na slici se može videti da su levi repovi deblji što je veći parametar γ, odnosno manji parametar α.

Primer 2.4 [***Funkcije raspodele koje pripadaju*** $MDA(\Lambda)$]

- *Eksponencijalna raspodela*



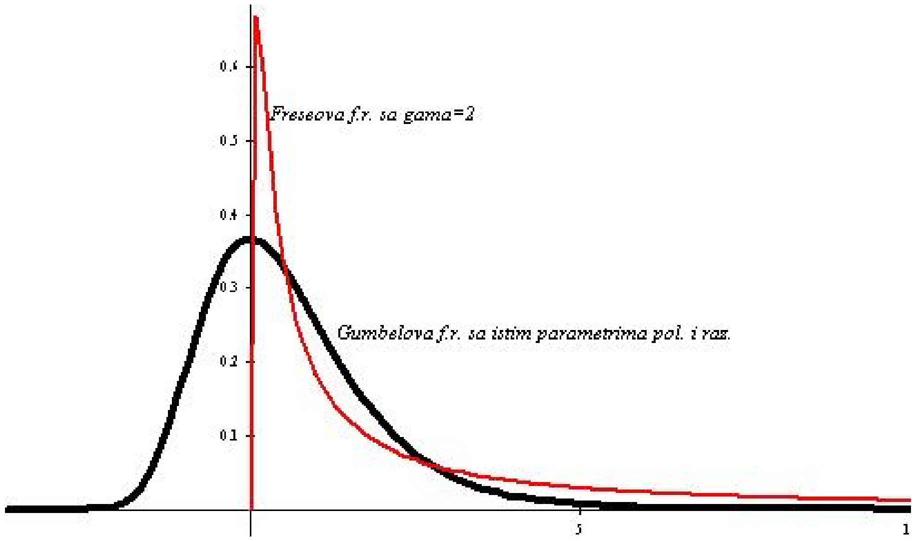
SLIKA 2.4: Poredjenje Gumbelovih funkcija raspodele za različite parametre položaja i razmere

- *Lognormalna raspodela*
- *Normalna raspodela*
- *Erlangova raspodela*
- *Funkcije raspodele koje pokazuju eksponencijalno ponašanje u krajnjoj desnoj tački, tj. za koje važi da je*

$$\overline{F}(x) = Kexp\{-\frac{\alpha}{x_F - x}\}, \quad x < x_F, \quad K, \alpha > 0.$$

Videti *sliku 2.5* za prikaz razlike izmedju Gumbelove i Frešeove funkcije raspodele.

Na kraju ovog po glavlja predstavljamo dva primera preuzeta iz realnog života kao ilustraciju opisanih raspodela ekstremnih vrednosti.

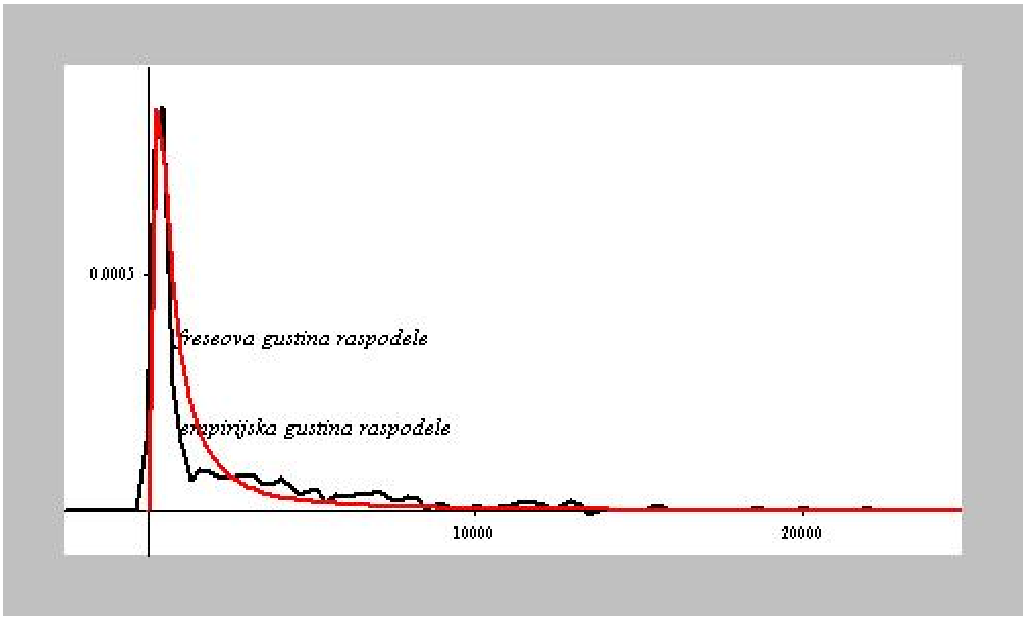


SLIKA 2.5: Poredjenje Frešeove i Gumbelove raspodele

Primer 2.5 *[Funkcija raspodele najvećih plata igrača bejzbola Američke lige]*

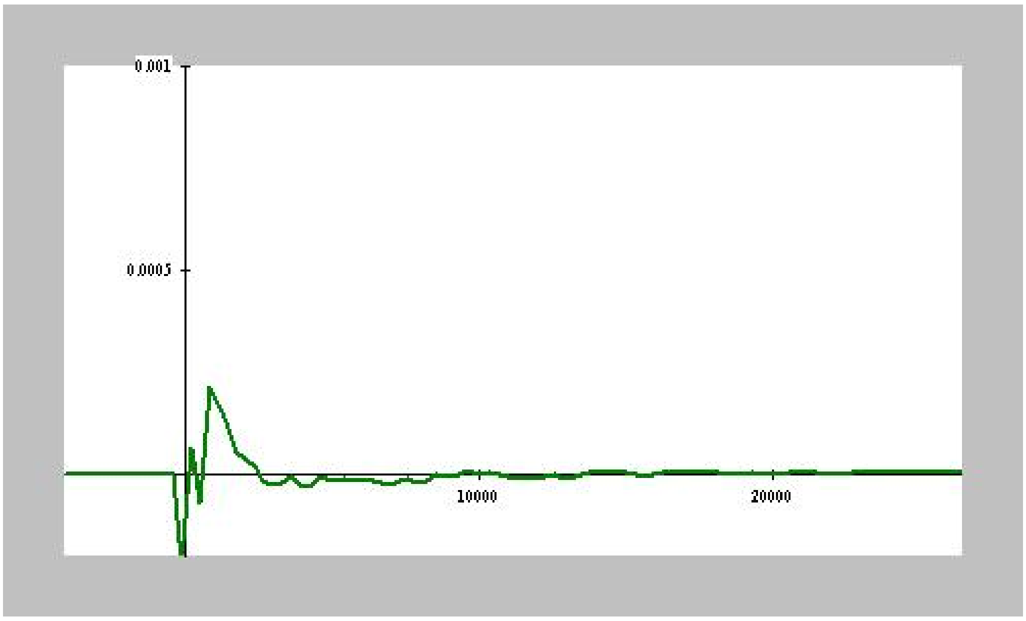
U ovom primeru posmatramo realne podatke čiju empirijsku funkciju raspodele želimo da procenimo. Analiziramo raspodelu ekstremnih vrednosti plata najbolje plaćenih bejzbol igrača u SAD. Podaci su preuzeti sa interneta 2003. godine i odnose se na trinaest najprestižnijih timova koji pripadaju Američkoj ligi (New York Yankees, Boston Red Sox, Cleveland Indians, Toronto Blue Jays, Baltimore Orioles, Tampa Bay Devil Rays, Kansas City Royals, Minnesota Twins, Chicago White Sox, Detroit Tigers, Seattle Mariners, Oakland Athletics, Texas Rangers). Obim uzorka je 380. Pogledati sliku 2.6 za empirijsku funkciju raspodele ovih podataka.

Raspodela je vrlo slična Frešeovoj, tako da možemo da uporedimo odgovarajuću Frešeovu raspodelu sa istim parametrima α, μ *(parametar položaja) i* σ *(parametar razmere) pri čemu su kao ocene maksimalne verodostojnosti za parametre dobijeni sledeći rezultati:* $\alpha = 0,5$, $\mu = 0$ *i* $\sigma = 592$. *Na slici 2.6 uporedjivana je empirijska funkcija raspodele sa Frešeovom funkcijom raspodele sa parametrima* $\alpha = 0,5$, $\mu = 0$ *i* $\sigma = 592$. *Takodje, pogledati i sliku 2.7 koja pokazuje kvalitet modelovanja. Ovde je više nego očigledno da je raspodela upravo Frešeova, kao to se može i testirati* *primenom testa količnika verodostojnosti, gde se pripadnost Gumbelovoj raspodeli odbacuje za male vrednosti verovatnoće p.*



SLIKA 2.6: Uporedjivanje empirijske i Frešeove funkcije raspodele

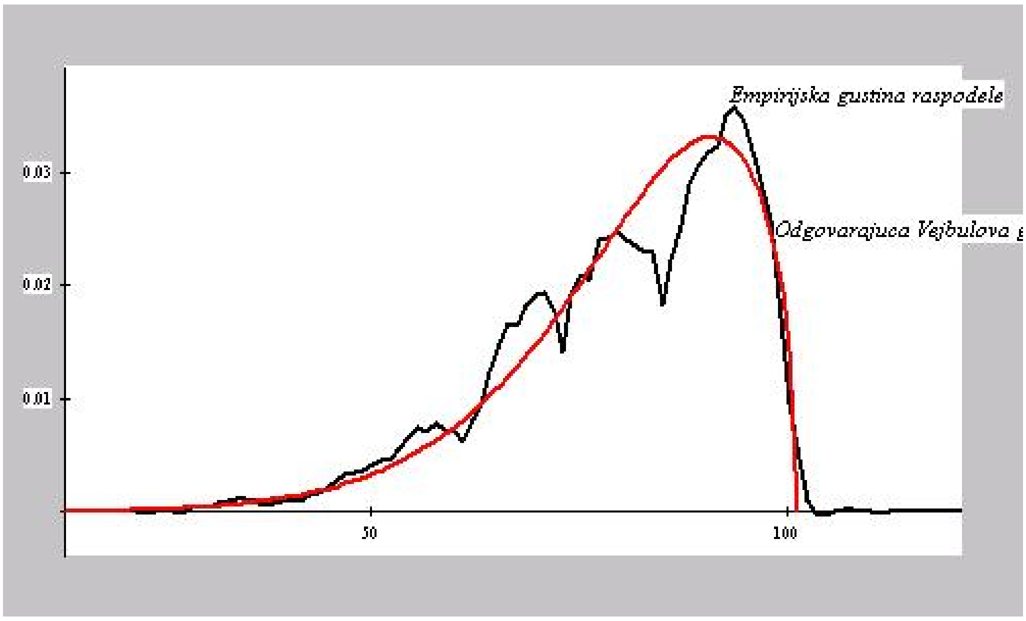
Primer 2.6 *[**Maksimalne mesečne temperature**] Podaci koji se ovde analiziraju predstavljaju maksimalne mesečne temperature vazduha na Havajima, merene u letnjim mesecima: junu, julu i avgustu u toku perioda od deset godina. Vrednosti su izražene u Farenhajtima. Podaci su preuzeti sa interneta. (NationalClimatic Data Center, Asheville, NC). Cilj analize je pre svega odredjivanje najpribližnije raspodele slučajne promenljive T, koja u ovom slučaju predstavlja temperaturu. Gumbel je isticao veliku važnost posmatranja samih podataka, detaljnog crtanja grafikona i njihove analize. S obzirom da se radi o maksimalnim temperaturama, očekujemo da će raspodela naših podataka biti istog tipa kao jedna od tri parametarske familije raspodela: Gumbelova, Frešeova ili Vejbulova. Na slici 2.8 su prikazane empirijska i pretpostavljena Vejbulova raspodela. Linija crne boje predstavlja uzoračku raspodelu, a drugu liniju crvene boje dobili smo modelirajući Vejbulovu raspodelu sa najpribližnijim parametrima. To smo postigli zahvaljujući opciji MDE (Minimal distance estimator), koji izvrava najpribližnije ocenjivanje i aproksimaciju date raspodele.*



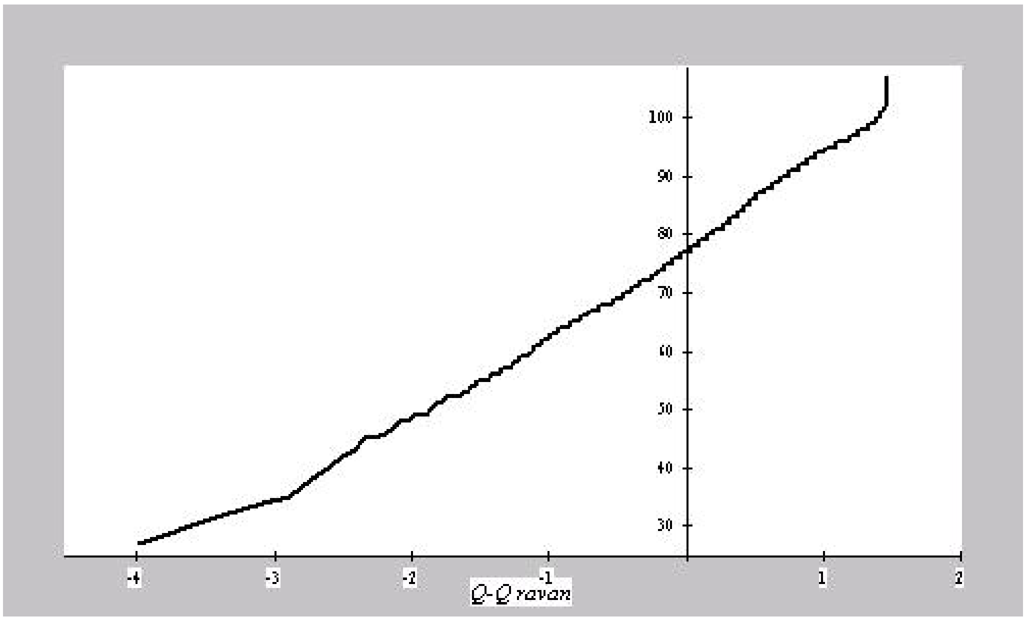
SLIKA 2.7: Razlika izmedju empirijske i ocenjene Frešeove raspodele

Ocenjeni parametri su redom: $\alpha = -1,47$, $\sigma = 22,41$ *i* $\mu = 101,241$. *Procenjena tačka koja je desni nosač je 107. Sve nam to ukazuje na Vejbulovu raspodelu. Koristili smo i test količnika verodostojnosti, gde smo Gumbelovu raspodelu odbacili.*

Na slici 2.9 data je QQ ravan koja nam, kako smo u ranijem tekstu naveli, može pokazati kvalitet modelovanja. Uporedili smo kvantile empirijske funkcije raspodele sa odgovarajućim kvantilima pretpostavljene Vejbulove funkcije raspodele. Kako je ranije rečeno, aproksimacija je dobra ukoliko je grafik približno prava linija.



Slika 2.8: Uporedjivanje empirijske i Vejbulove funkcije raspodele



Slika 2.9: Poredjenje odgovarajućih raspodela (QQ ravan)

Poglavlje 3

OCENJIVANJE INDEKSA PRAVILNE PROMENLJIVOSTI

3.1 Uvod

Jedan od problema sa kojim se susreću istraživači prilikom modelovanja raspodela sa pravilno promenljivim repovima jeste ocenjivanje parametra pravilne promenljivosti. Prve statističke studije o ponašanju skokova cena na berzi pokazuju važnost odredjivanja debljine repa empirijske raspodele (pogledati Blattberg, Gonedes (1974)). Efekat ne-normalnosti u raspodelama koje se dobijaju u finansijskim serijama je privukao značajnu paznju javnosti naročito u vreme neželjenih i iznenadnih dogadjaja kao što su: krah berze 1987, Azijska kriza 1997 i finansijska kriza u Rusiji 1998 (pogledati Rubinstein (1994) koji sugeriše neophodnost fokusiranja pažnje na modeliranje raspodela vrednosti akcija). Takodje, različiti slučajevi finansijskih turbulencija u novije vreme dokazuju da se dogadjaji za koje se smatralo da se mogu pojaviti "*jednom u hiljadu godina*" ipak pojavljuju mnogo frekventnije (pogledati prvo poglavlje i značenje *"crni labud"*). Ova činjenica pokazuje važnost pažljivog modeliranja ekstremnih finansijskih dogadjaja prilikom procene rizika. Teorija ekstremnih vrednosti, o kojoj je bilo reči u prethodnom poglavlju, obezbedjuje matematički okvir unutar koga se mogu istraživati asimptotske ekstremalne karakteristike stacionarnih raspodela. Ona omogućava da donosimo zaključke i van ranga observiranih uzoračkih veličina, dakle omogućava predvidjanje ekstremnih dogadjaja. U tom cilju, esencijalno je ocenjivanje indeksa repa posmatrane raspodele za koje ova teorija nudi više različitih pristupa (pogledati sledeće radove za više informacija o ocenjivanju indeksa repa raspodele sa akcentom na praktičnim aplikacijama u oblasti finansija: Koedijk i dr. (1990), Dacorogna i dr. (1993), Danielsson, de Vries (1997a), Huisman i dr. (2001), Jondeau, Rockinger (1998), Lux (2000) i McNeil, Frey (2000)).

Najpoznatija i najviše korišćena ocena u finansijskim serijama je Hilova ocena (Hill, 1975). Pokazano je da je ona najpogodnija ocena u slučaju finansijskih podataka. Semiparametarski pristup je zasnovan na pretpostavci da se posmatrana funkcija raspodele nalazi u domenu privlačenja Frešeove funkcije raspodele ekstremnih vrednosti. Ova pretpostavka generalno važi za teške repove finansijskih raspodela. Medjutim, ova ocene je najoptimalnija kada se radi o uzorku nezavisnih slučajnih veličina sa istom Paretovom funkcijom raspodele. Većina postojećih ocena zasniva se na odredjenom broju k maksimalnih statistika poretka uzorka veličine n (videti Csörgo i dr. (1985) za prikaz ocena ovog tipa). Tačnost ocena, naravno, zavisi od preciznosti u biranju broja maksimalnih statistika poretka. Izbor najoptimalnijeg broja ekstremnih statistika poretka k predstavlja dugi niz godina veliki izazov za istraživače (videti Embrechts i dr. (1997) i njihove reference za više podataka o ovoj temi). Može se naći asimptotski najoptimalnija vrednost za k, u smislu najmanje srednjekvadratne greške. Poslednjih godina, predlagane su različite procedure za biranje optimalnog broja statistika poretka sa asimptotski najmanjom srednjekvadratnom greškom (kao primere pogledati sledeće radove: Dekkers, de Haan (1993), Beirlant i dr. (1996), Drees, Kaufmann (1998) i Danielson i dr. (2001)). De Haan i dr. (1998) predlažu tzv. *bootstrap* metod kojim se u dva koraka odredjuje asimptotski najmanja srednjekvadratna greška, pri čemu za razliku od drugih poznatih metoda nije neophodno poznavanje parametra drugog reda.

Indeks repa raspodele je indikator verovatnoće da se desi velika devijacija kod slučajne promenljive, pa kao takav ima široku primenu kod finansijskih vremenskih serija. Ocenjivanjem indeksa pravilne promenljivosti na osnovu uzorka medjusobno nezavisnih i identično raspodeljenih slučajnih veličina bavio se veliki broj naučnika i istraživača poslednjih godina. Takodje, indeks repa je ocenjivan i u slučajevima medjusobno zavisnih slučajnih veličina pod različitim početnim uslovima.

Sem Hilove ocene, u literaturi su predlagane: Pikandsova ocena (Pikands (1975)), Moment ocena (Dekkers, Einhmal, De Haan (1989)), ocena geometrijskog tipa (Brito, Freitas 2003), Shultze, Steinbach (1996) i dr. De Haan (1981) predlaže ocenu čija su svojstva analizirana u radu de Haan, Resnick (1980), a Hall (1982) dokazuje njenu asimptotsku normalnost pod još nekim dodatnim uslovima nego što predpostavljaju prethodni autori. Na temu ocenjivanja indeksa repa postoji brojna literatura, kao na primer: Csörgó, Mason (1985), Haeusler, Teugels (1984), Teugels (1981a, 1982b), De Meyer, Teugels (1983), Gawronski, Stadmüler (1984), Hall, Welsh (1984), Du Mouchel (1983), Welsh (1984), Dekkers (1989), Drees (1995), Csörgó, Viharos (1998), de Haan, Peng (1998) and Bacro, Brito (1995).

3.2 Ocenjivanje indeksa pravilne promenljivosti

U funkcijama raspodele ekstremnih vrednosti Φ_α i Ψ_α figuriše parametar α o kome smo govorili na početku prvog poglavlja. Neophodnost što preciznijeg ocenjivanja ove konstante naročito se pokazala u domenima od interesa o kojima smo ranije govorili, a u kojima se pojavljuju raspodele debelih repova. Korišćenjem odgovarajućeg uzorka ekstremnih veličina, možemo oceniti indeks α i na taj način proceniti kako se ponaša funkcija raspodele datog uzorka u svom repu, odnosno možemo predvideti šta će se dogadjati sa kvantilima višeg reda. Naravno, u modelovanju se koriste još mnoge druge karakteristike funkcije raspodele kao što su zakrivljenost i modalnost što prevazilazi okvire naše teme. Postoji veliki broj ocena ovog indeksa i veoma intenzivno se radilo na poboljšavanju njihovih osobina i preciznosti. Naredni paragraf započinjemo Hilovom ocenom parametra α (Hill (1975)), a u nastavku ćemo analizirati i uporedjivati još dve poznatije ocene parametra α: Pikandsovu (Pikands (1975)) i Moment ocenu (Dekkers, Einhmal, de Haan (1989)).

3.2.1 γ-parametarizacija

Ako uvedemo smenu $\gamma = 1/\alpha$ za Frešeovu raspodelu i $\gamma = -1/\alpha$ za Vejbulovu raspodelu, onda se funkcije raspodele ekstremnih vrednosti mogu zapisati u obliku parametarske familije koja zavisi od jednog parametra na sledeći način:

$$G_\gamma(x) = \begin{cases} exp\{-(1+\gamma x)^{-1/\gamma}\}, & \text{ako je } \gamma \neq 0 \\ exp\{-e^{-x}\}, & \text{ako je } \gamma = 0. \end{cases}$$

U vezi sa γ-parametarizacijom familije raspodela ekstremnih vrednosti primetimo da je granična vrednost izraza $(1+\gamma x)^{-1/\gamma}$ jednaka e^{-x} kad $\gamma \to 0$.

Definicija 3.1 (**Generalisane Paretove raspodele**). Ovaj tip raspodela dat je sledećom formulom:

$$G_{\gamma,\beta}(x) = \begin{cases} 1-(1+\frac{\gamma x}{\beta})^{-1/\gamma}, & \text{ako je } \gamma \neq 0 \\ 1-exp\{-\frac{x}{\beta}\}, & \text{ako je } \gamma = 0, \end{cases} \tag{3.1}$$

gde je β parametar razmere. Takodje, nosač raspodele $G_{\gamma,\beta}(x)$ je $x \geq 0$, ako je $\gamma \geq 0$, odnosno skup $0 \leq x \leq -\beta/\gamma$ ukoliko je $\gamma < 0$.

3.2.2 Hilova ocena

Verovatno najpopularnija ocena parametra γ je Hilova ocena i definisana je na sledeći način:

$$\gamma_{k,n}^H = \frac{1}{k}\sum_{t=1}^{k} \ln X_{(t)} - \ln X_{(k+1)}, \tag{3.2}$$

gde $X_{(n)}, ..., X_{(1)}$ predstavlja niz statistika poretka u rastućem redosledu, a $k = k_n$ je niz pozitivnih celih brojeva koji zadovoljava

$$1 \leq k_n < n, \quad \lim_{n\to\infty} k_n = \infty \quad \text{i} \quad \lim_{n\to\infty} k_n/n = 0. \tag{3.3}$$

Obično se indeks $\gamma = 1/\alpha$ naziva indeks ekstremnih vrednosti (eng. *extreme value index-EVI*).

Asimptotsko ponašanje ove ocene analizirano je od strane velikog broja autora pod različitim uslovima. Osobine ocene su intenzivno razmatrane u slučaju kada su X_n identično raspodeljene i nezavisne slučajne promenljive (videti kao primer Davis and Resnick (1984), Haeusler and Teugels (1985)). Dekkers, Einmahl and de Haan (1989) su proširili značenje Hilove ocene kao ocene indeksa regularne varijacije na ocenu indeksa raspodele ekstremnih vrednosti. Kao prilično značajan problem u korišćenju ove ocene predstavlja izbor optimalnog broja maksimalnih statistika potretka, odnosno izbor broja k koji figuriše u formuli. Ukoliko je taj broj mali, javiće se mali bijas ali velika varijansa, a ukoliko je k suviše veliki broj, biće veliki bijas, a varijansa će biti mala.

Teorema 3.2. *[Osobine Hilove ocene] Neka je X_n strogo stacionaran niz sa marginalnom funkcijom raspodele F koja za neko $\alpha > 0$ i $L \in \mathcal{R}_0$ zadovoljava*

$$\overline{F} = P(X > x) = x^{\alpha} L(x), \quad x > 0.$$

Označimo sa $\gamma^H = \gamma^H_{k,n}$ Hilovu ocenu.

1. *(Slaba konzistencija) Pretpostavimo da je zadovoljen jedan od sledećih uslova:*
   - *X_n je niz nezavisnih slučajnih promenljivih (Mason (1982))*
   - *X_n je niz slabo zavisnih slučajnih promenljivih (Rootzen, Leadbetter (1982, 1988), Hsing (1991))*
   - *X_n je linearan proces (Resnick, Starica (1995),(1998))*

   Ukoliko $\lim_{n\to\infty} k_n = \infty$ *i* $\lim_{n\to\infty} k_n/n = 0$, *kad* $n \to \infty$, *sledi*

$$\gamma^H \xrightarrow{P} \gamma.$$

2. *(Jaka konzistencija) (Deheuvels, Hausler, Mason (1988)) Ako* $k/n \to 0$, $k/\ln\ln n \to \infty$ *kad* $n \to \infty$ *i* X_n *je niz nezavisnih slučajnih promenljivih, onda*
$$\gamma^H \xrightarrow{a.s} \gamma.$$

3. *(Asimptotska normalnost) Ako su zadovoljeni još neki dodatni uslovi za* k i F *i ako je* X_n *niz nezavisnih slučajnih veličina, onda*
$$\sqrt{k}(\gamma^H - \gamma) \xrightarrow{d} N(0, \frac{1}{\gamma^2}). \quad \blacksquare$$

Rezultat o asimptotskoj normalnosti Hilove ocene omogućava odredjivanje intervala poverenja nepoznatog parametra $\gamma > 0$.

Na osnovu Teoreme 3.2 ne može se smatrati da su svojstva Hilove ocene uvek dobra. Tačnije, teorema pokazuje da u opštem slučaju za Hilovu ocenu važe neka standardna statistička svojstva. Medjutim, za bolje performanse ocene porebni su dodatni uslovi vezani za oblik repa raspodele $\overline{F}$ kao i odredjeni zahtevi vezani za broj statistika poretka $k = k(n)$. Zapravo, da bi se dokazala asimptotska normalnost Hilove ocene γ^H potrebni su nam dodatni uslovi pravilne promenljivosti drugog reda koji se odnose na $\overline{F}$ koji nisu proverljivi u praksi, što je otežavajuća okolnost.

Može se pokazati da je u slučaju niza nezavisnih identično raspodeljenih slučajnih veličina Hilova ocena konzistentna, kao i da je asimptotski normalna. Medjutim postoje primeri (videti Embrechts i dr. (1997)) na kojima se može videti da je brzina konvergencije Hilove ocene u nekim slučajevima uzoraka prilično spora. Ukoliko sporo promenljiva funkcija L koja figuriše u formuli za rep raspodele bitno odstupa od konstante, može se javiti priličan bijas Hilove ocene. U cilju smanjenja bijasa pojavile su se i neke nove ocene: Beirlant i dr. (1999), Feuerverger, Hall (1999), Guillou, Hall (2001) i Ling, Peng (2004).

Što se tiče problema vezanog za utvrdjivanje optimalnog broja k, u praksi se obično primenjuje Hilova ravan, tj. najoptimalniji broj statistika poretka se odredjuje grafičkim putem, tako što se posmatra grafik

$$\{(k, \gamma_{k,n}^{H}) : k = 2, ...n\}.$$

Za detaljniju analizu korišćenja Hilove ravni u praksi videti Resnick, Starica (1997). Pogledati *sliku 1.4* Hilove ravni kojom se može grafičkim putem proceniti vrednost parametra, tako što se bira oblast u kojoj je grafikon linearan i stabilan. Korišćeno je simulirano ocenjivanje parametra γ u zavisnosti od broja k na Frešeovoj raspodeli sa parametrom $\gamma = 2$.

3.2.3 Pikandsova ocena

Pikandsova ocena parametra $\gamma = \alpha^{-1}$ definiše se na sledeći način:

$$\widehat{\gamma}_{k,n}^{P} = \frac{1}{\ln 2} \ln \frac{X_{(k)} - X_{(2k)}}{X_{(2k)} - X_{(4k)}}, \tag{3.4}$$

gde $X_{(n)}, ..., X_{(1)}$ predstavlja niz statistika poretka niza $\{X_t\}$ poredjanih u rastućem redosledu.

Teorema 3.3. *[Osobine Pikandsove ocene] Neka je $\{X_t\}$ niz medjusobno nezavisnih slučajnih veličina sa zajedničkom funkcijom raspodele F za koju važi da pripada maksimalnom domenu privlačenja neke od funkcija raspodele ekstremnih vrednosti $MDA(G_\gamma)$, gde je $\gamma \in \mathbb{R}$. Označimo sa $\widehat{\gamma}^P = \widehat{\gamma}_{k,n}^P$ Pikandsovu ocenu parametra γ. Tada važe sledeća tvrdjenja:*

1. *(Slaba konzistencija) Ako $k \to \infty$, $k/n \to 0$, kad $n \to \infty$, onda važi da*

$$\widehat{\gamma}^P \xrightarrow{P} \gamma, \quad n \to \infty.$$

2. *(Jaka konzistencija) Ako $k/n \to 0$, $k/\ln\ln n \to \infty$ kad $n \to \infty$, tada*

$$\widehat{\gamma}^P \xrightarrow{a.s} \gamma, \quad n \to \infty.$$

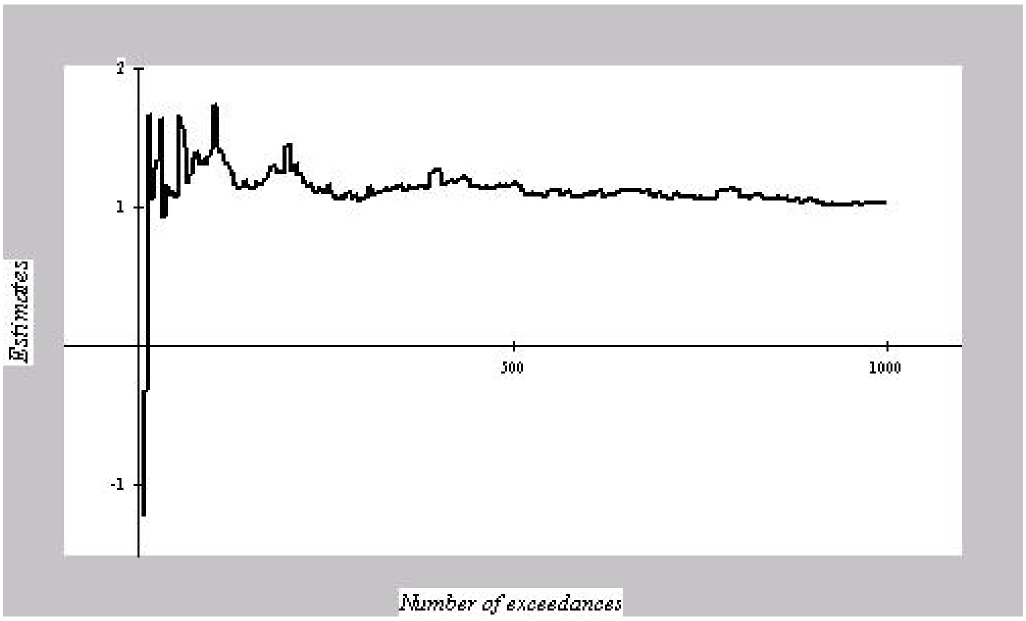
3. *(Asimptotska normalnost) Pod još nekim dodatnim uslovima za k i F (videti Dekkers, de Haan (1989), p.1799),*

$$\sqrt{k}(\widehat{\gamma}^P - \gamma) \xrightarrow{d} \mathrm{N}(0, \nu(\gamma)), \quad n \to \infty,$$

gde je

$$\nu(\gamma) = \frac{\gamma^2(2^{2\gamma+1} + 1)}{(2(2^\gamma - 1)\ln 2)^2}.$$

Kao i u slučaju Hilove ocene u formuli za Pikandsovu ocenu pojavljuje se odredjeni broj statistika poretka $k = k(n)$. Kao značajan instrument u odredjivanju optimalnog broja k koristi se



Slika 3.1: Pikandsova ravan

Pikandsova ravan koja predstavlja grafik

$$\{(k, \widehat{\gamma}^P_{k,n}) : k = 1, ...n\}.$$

Broj k treba izabrati iz onog dela x-ose u kome je grafik pretežno horizontalna linija (pogledati *sliku 3.1.* na kojoj je simulirano ocenjivanje parametra γ u zavisnosti od broja k na Frešeovoj raspodeli sa parametrom $\gamma = 1$).

3.2.4 Moment ocena

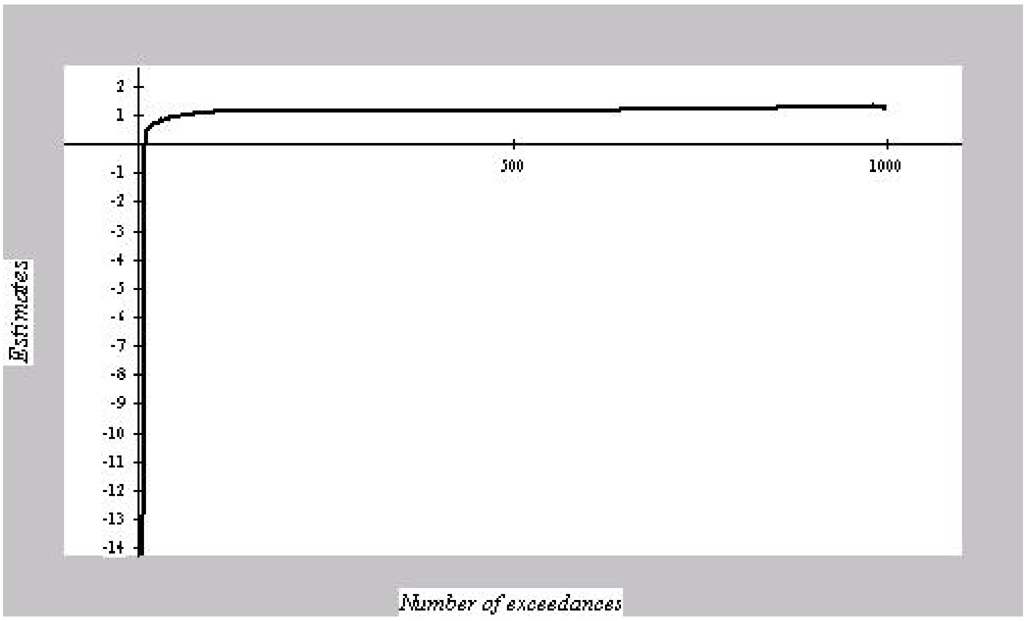
Ocena parametra γ metodom momenata nastala je u cilju proširenja Hilove ocene za širi spektar vrednosti, tj. za parametre $\gamma \in \mathbb{R}$. Ranije smo napomenuli (pogledati paragraf 3.2.2) da je Hilovu ocenu moguće koristiti samo u slučajevima kada je $\gamma > 0$. Dekkers, Einhmal, De Haan (1989) predlažu sledeću ocenu parametra γ koja bi pokrivala sve realne vrednosti:

$$\gamma^M = 1 + \gamma_1^H + \frac{1}{2}(\frac{\gamma_1^H}{\gamma_2^H} - 1)^{-1}, \tag{3.5}$$

gde je $\gamma_1^H = \frac{1}{k}\sum_{j=1}^{k}(\ln X_{j,n} - \ln X_{k+1,n})$ Hilova ocena i $\gamma_2^H = \frac{1}{k}\sum_{j=1}^{k}(\ln X_{j,n} - \ln X_{k+1,n})^2$. S obzirom da se ovi izrazi mogu interpretirati kao empirijski momenti, ocena je nazvana *Moment ocena.* I u slučaju ove ocene, može se posmatrati odgovarajući grafik vrednosti ocene u zavisnosti od parametra k. Utvrdjuje se ocena parametra repa tako što se posmatra deo grafika koji je približno jednak pravoj liniji (videti *sliku 3.2*, na kojoj je simulirano ocenjivanje parametra γ u zavisnosti od broja k na Frešeovoj raspodeli sa parametrom $\gamma = 1$).

3.2.5 Poredjenje ocena parametra pravilne promenljivosti

Logično pitanje koje se postavlja u modelovenju vremenskih serija je kako izabrati optimalnu ocenu indeksa repa raspodele. Odgovor na to pitanje predstavlja veliki izazov i u priličnoj meri zavisi od parametra drugog reda $\rho \leq 0$, koji upravlja konvergencijom količnika $\overline{F}(tx)/\overline{F}(t)$. Pokazalo se da Hilova ocena ima optimalne



SLIKA 3.2: Moment ravan

osobine kada sporo promenljiva funkcija L, koja figuriše u formuli za rep raspodele, zadovoljava da

$$\lim_{t\to\infty} \frac{\overline{F}(tx)/\overline{F}(t) - x^{-\alpha}}{a(t)} = x^{-\alpha}\frac{x^{\rho}-1}{\rho},$$

za svako $x > 0$ postoji, gde je $a(t)$ merljiva funkcija konstantnog znaka i gde se podrazumeva da je desna strana jednakosti jednaka nuli za parametar $\rho = 0$.

Asimptotska disperzija Hillove ocene ima oblik α^2/k . S obzirom da asimptotska disperzija recipročno zavisi od broja statistika poretka, onda bi bilo logično uzeti veći broj ekstremnih statistika kako bi disperzija bila manja. Na žalost, pokazalo se da se može desiti da u tom slučaju postoji povećanje pristrasnosti (Goldie and Smith (1987)). Tu dolazimo do pitanja, od čega zavisi i u kojoj meri postoji raskorak izmedju ove dve veličine.

Pokazalo se (Embrechts i dr.(1997), *Teorema 6.4.9*) da u velikoj meri veličina pristrasnosti i varijanse kod Hilove ocene zavisi upravo od medjusobnog odnosa parametara prvog i drugog reda, tj. parametara ρ i γ.

Hilova ocena je, kako je već objašnjeno, ograničena na raspodele sa pozitivnim parametrom γ, dok su Pikandsova i Moment ocena primenljive na čitavu klasu parametara, tj. za $\gamma \in \mathbb{R}$. Ipak, nema preciznog odgovora koju ocenu treba upotrebiti. U slučaju da se radi o pozitivnom parametru, preporuka je da se koristi Hilova ocena, jer se može pokazati (videti Smith, Weissman (1985), (1987)) da, u slučaju da je parametar drugog reda $\rho = 0$, najmanju srednjekvadratnu grešku ima Hilova ocena.

Ako se radi o širem spektru vrednosti za γ, onda vršimo poredjenje Pikandsove i Moment ocene. Medjutim, njihova asimptotska efikasnost u velikoj meri zavisi od medjusobnog odnosa parametara α i ρ. Pokazano je da za $\gamma > -2$ Moment ocena ima manju varijansu nego Pikandsova ocena. Takodje, pokazalo se da Pikandsova ocena u nekim slučajevima može pokazati odredjeni stepen nestabilnosti. Na ovu temu mogu se pogledati radovi: Danielson, de Vries (1997), Hall (1982), Anderson (1984) i Davis, Resnick (1984).

U primerima 3.1 i 3.2 koji slede prikazaćemo ponašanje Hilove ocene, Pikandsove ocene i Moment ocene prilikom procene parametra repa simulirane Frešeove raspodele sa indeksom debljine repa $\gamma = 1$ i simulirane Paretove raspodele sa parametrom debljine repa $\gamma = 2$, parametrom položaja $\mu = 0$ i parametrom razmere $\sigma = 1$.

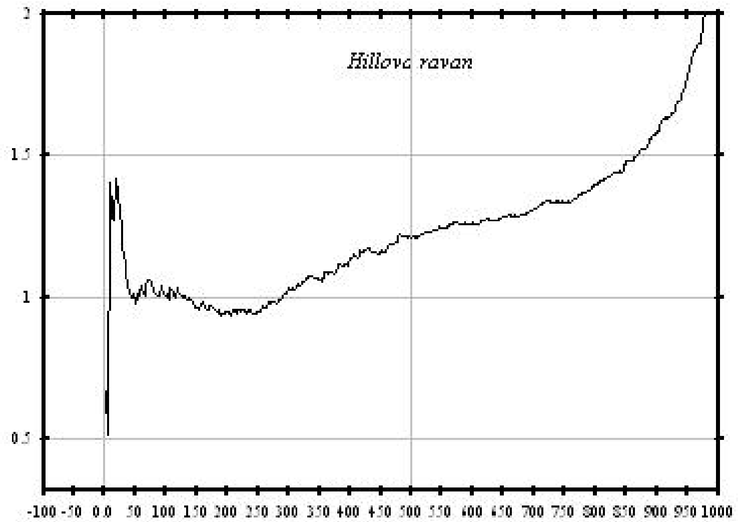
Primer 3.1. *[**Simulacija različitih ocena Paretovog indeksa**]*

U ovoj simulaciji koristili smo 1000 podataka sa Frešeovom funkcijom raspodele sa parametrom $\gamma = 1$. Posmatrali smo tri osnovne ocene parametra teskog repa čija smo svojstva prethodno opisali.

Svaka ocena opisana je tabelom sa brojem korišćenih ekstremnih statistika, kao i odgovarajućim grafikonom (slika 3.3, slika 3.4, slika 3.5 slika 3.6). Za statističku obradu podataka korišćen je softverski paket Xtremes.

Tabela 1. ***HILOVA OCENA***

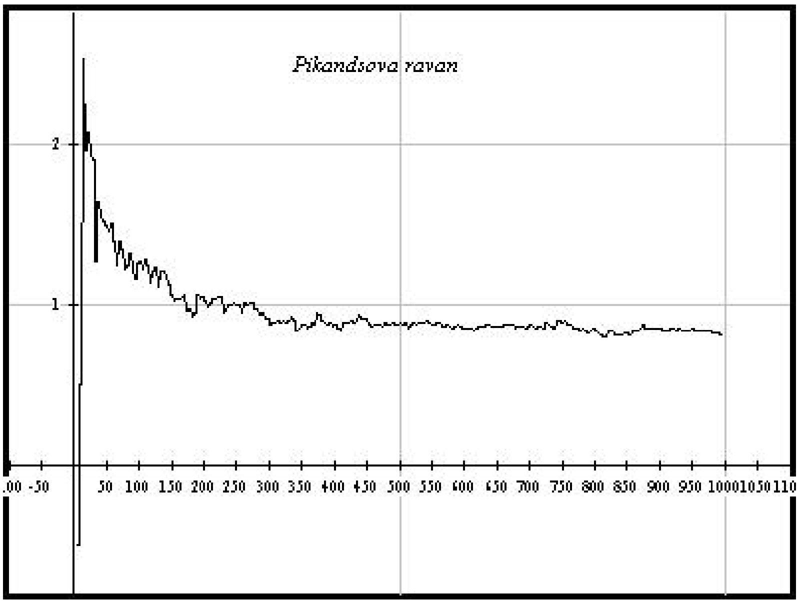
Broj ekstremnih statistika	*Hilova ocena*
90	*1.12*
100	*0.999*
105	*1.011*
110	*0.973*
120	*0.983*
130	*0.922*
140	*0.947*
150	*0.924*
155	*0.931*
160	*0.995*



Slika 3.3: Hilova ocena

Tabela 2. ***PIKANDSOVA OCENA***

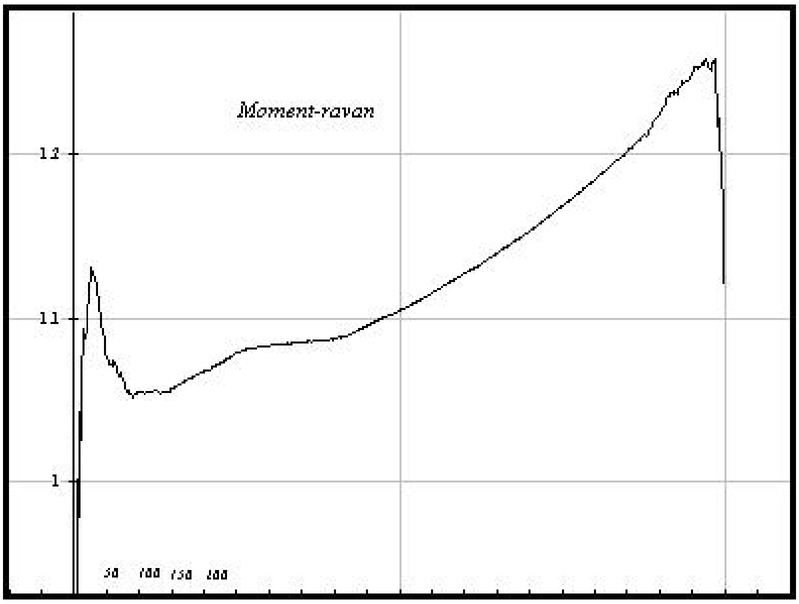
Broj ekstremnih statistika	*Pikandsova ocena*
90	*1.09*
100	*1.211*
105	*1.221*
110	*1.005*
120	*1.083*
130	*1.229*
140	*1.190*
150	*1.108*
160	*1.173*



SLIKA 3.4: Pikandsova ocena

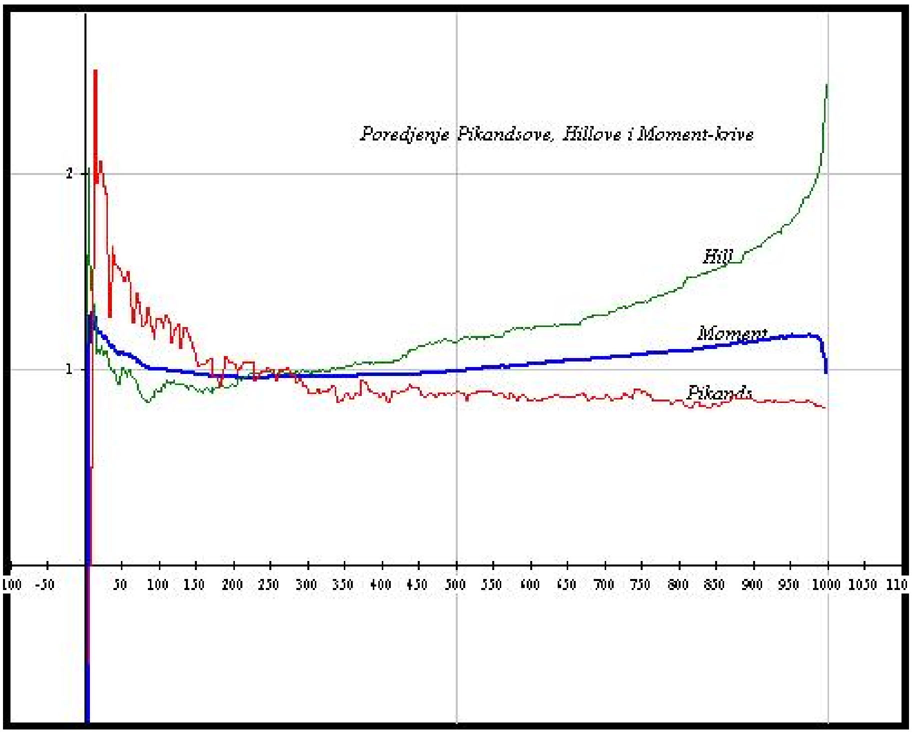
Tabela 3. ***MOMENT OCENA***

Broj ekstremnih statistika	*Moment ocena*
90	*1.052*
100	*1.054*
105	*1.055*
110	*1.055*
120	*1.055*
130	*1.054*
140	*1.055*
150	*1.057*
160	*1.059*



Slika 3.5: Moment ocena

Na grafikonu slike 3.6 možemo uočiti da se sve tri ocene poklapaju u oblasti broja statistika koji pripada intervalu (220, 270) i u tom intervalu su sve tri ocene veoma bliske jedinici. U praktičnim aplikacijama obično se i koristi više različitih ocena da bi se potvrdila pretpostavljena vrednost traženog indeksa. Dakle, u našem primeru sve tri ocene pokazuju da je najpribližnija ocenjena vrednost parametra upravo $\gamma = 1$.



SLIKA 3.6: Poredjenje posmatranih ocena

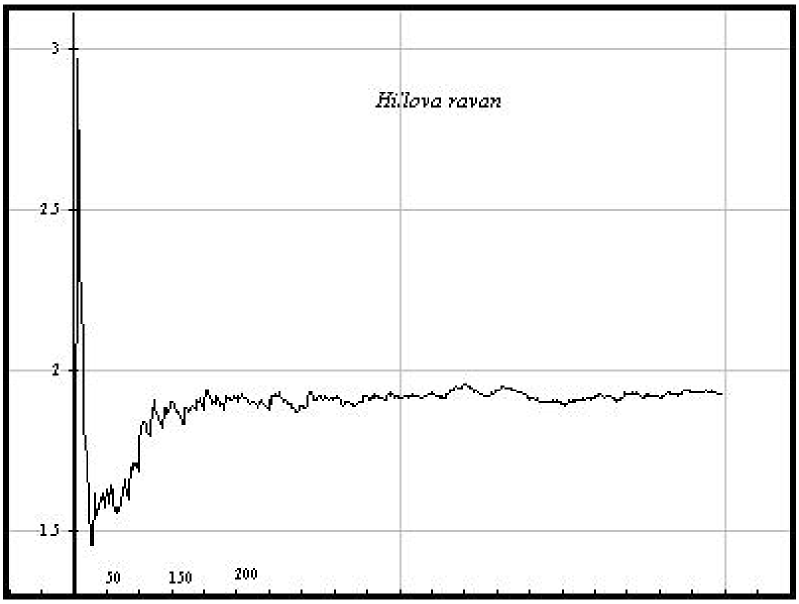
Primer 3.2. ***[Simulacija različitih ocena Paretovog indeksa]***

U ovoj simulaciji koristili smo 1000 podataka sa Paretovom funkcijom raspodele sa parametrom debljine repa $\gamma = 2$, *parametrom položaja* $\mu = 0$ *i parametrom razmere* $\sigma = 1$. *Posmatrali smo tri ocene parametra koje smo u prethodnom tekstu opisali.*

Svaka ocena opisana je tabelom sa brojem korišćenih ekstremnih statistika, kao i odgovarajućim grafikonom (slika 3.7, slika 3.8 i slika 3.9). I u ovom slučaju, za statističku obradu korišćen je program Xtremes.

Tabela 4. ***HILOVA OCENA***

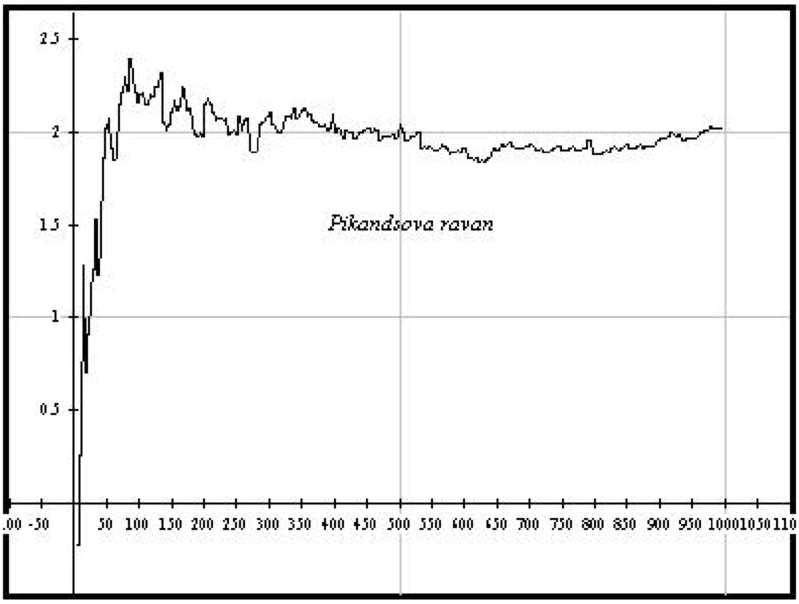
Broj ekstremnih statistika	*Hilova ocena*
100	*1.91*
105	*1.88*
110	*1.90*
115	*1.89*
120	*1.89*
135	*1.84*
140	*1.92*
145	*1.94*
150	*1.93*
155	*1.91*
160	*1.93*



SLIKA 3.7: Hilova ocena

Tabela 5. ***PIKANDSOVA OCENA***

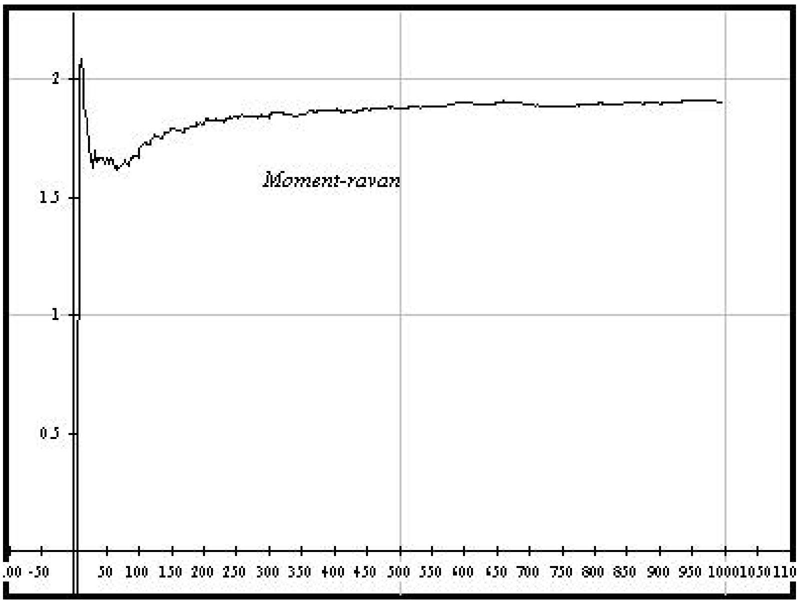
Broj ekstremnih statistika	*Pikandsova ocena*
100	*2.07*
105	*2.11*
110	*2.09*
115	*2.09*
120	*2.20*
130	*2.02*
135	*2.03*
140	*2.02*
145	*2.02*
155	*1.92*
160	*1.89*



Slika 3.8: Pikandsova ocena

Tabela 6. ***MOMENT OCENA***

Broj ekstremnih statistika	*Moment ocena*
115	*1.86*
120	*1.86*
145	*1.89*
150	*1.89*
155	*1.88*
160	*1.89*



SLIKA 3.9: Moment ocena

Napomena: Sem slike 1.1, slike 1.2 i slike 2.3 koje su kao ilustracija preuzete sa interneta od nepoznatog autora, sve ostale slike su delo autora disertacije.

Poglavlje 4

NEKOMPLETNI UZORCI

Nepotpuni podaci u datom uzorku su česta pojava i mogu značajno uticati na donošenje statističkih zaključaka. Razlozi zbog kojih se može javiti prazan odgovor u upitniku mogu biti različiti: zbog grešaka samih istraživača u sakupljnju podataka, ukoliko nije dobro vodjeno istraživanje ili se desila greška prilikom unosa podataka, zbog prostog odbijanja subjekta da odgovori na izvesna pitanja, kao i iz odredjenih političkih razloga kada vlada ne objavljuje sve podatke naročito u istraživanjima vezanim za oblast ekonomije, sociologije i političkih nauka. Takodje, često se nekompletnost javlja prilikom ponovljenih merenja, ako se vrši testiranje nekog parametra kroz duži vremenski period na istom uzorku. Tada može doći do nestajanja samih subjekata iz analize iz raznih pojedinačnih razloga. S obzirom da nekompletni podaci smanjuju reprezentativnost uzorka i remete zaključke vezane za populaciju, važno je mogućnost njihovog pojavljivanja svesti na minimum. U tom cilju, osmišljeni su kompjuterski upitnici kod kojih se na svako pitanje mora odgovoriti, inače se ne može preći na sledeće. Na taj način se kao uzrok moguće greške potpuno eliminišu istraživači. Naravno, u situacijama gde se pretpostavlja ili gde je već utvrdjeno da će se pojaviti nedostajući podaci, istraživač treba da planira korišćenje onih metoda koje su otpornije na nekompletnost. Metoda se smatra otpornom na nekompletnost ukoliko mala narušavanja osnovnih premisa neće dovesti do neželjene pristrasnosti i pogrešnih zaključaka o populaciji.

4.1 Tipovi nekompletnih uzoraka

4.1.1 Podaci nedostaju na potpuno slučajan način (MCAR-Missing completely at random)

Kada kažemo da podaci nedostaju na potpuno slučajan način, odnosno da su tipa MCAR, smatramo da verovatnoća da observacija X_i nedostaje nije povezana sa njenom vrednošću ili sa vrednošću bilo koje druge slučajne veličine. Kao primer ovog tipa može se navesti recimo neispravna oprema, loše vreme, pogrešan unos podataka. Medjutim, nedostajući podaci vezani za prihode ne mogu biti smatrani potpuno slučajnom pojavom, odnosno nisu tipa MCAR, ukoliko ispitanici sa nižim zaradama ne odgovaraju na to pitanje u sličnom broju kao oni sa visokim prihodima. S druge strane treba imati u vidu da je u odredjivanju tipa strukture nekompletnosti važna vrednost nedostajuće observacije. U slučaju prethodnog primera, ako ispitanici koji nisu odgovorili na pitanje o sopstvenoj zaradi nisu odgovorili ni na pitanje o celokupnim porodičnim prihodima, podaci još uvek mogu biti smatrani MCAR, ukoliko nepružanje informacija nije povezano sa vrednošću samih prihoda. Važna prednost ovakve strukture nekompletnosti je u tome što se analiza izvodi bez pristrasnosti. Naravno može se smanjiti moć statističkog testa ali ocenjeni parametri nemaju pristrasnost uzrokovanu nedostatkom podataka.

4.1.2 Podaci nedostaju na slučajan način (MAR-Missing at random)

Često u praktičnim situacijama podaci nisu MCAR, ali se mogu klasifikovati kao MAR, odnosno za date podatke se može reći da se pojavljuju na slučajan način. Za podatke koji su MCAR, zaključili smo da ne postoji korelacija izmedju verovatnoće da X_i nedostaje i njene i bilo koje druge vrdenosti ostalih slučajnih veličina. U slučaju MAR podataka, posmatra se verovatnoća da podatak X_i nedostaje ali u korelaciji sa nekim faktorom. Recimo, može se desiti da depresivni ljudi imaju malo motiva da odgovore na pitanje zarade, pa će postojati zavisnost izmedju izračunatog prihoda i depresije. Srednja vrednost prihoda kod depresivnih osoba biće na ovaj način značajno manja nego u slučaju kompletnog uzorka. U ovom slučaju, iako smo daklarisali ovaj tip strukture sa nedostajanjem na slučajan način, moramo se potruditi da pogodnim metodama i tehnikama koje su na raspolaganju obezbedimo što preciznije i značajne ocene sa što manjom pristrasnošću.

4.1.3 Podaci nedostaju na neslučajan način (MNAR-Missing not at random)

Ako podaci ne pripadaju ni jednom od prethodne dve strukture, onda možemo reći da se nedostajući podaci pojavljuju na neslučajan način, odnosno da pripadaju tipu MNAR. Na primer, ako se ispituje mentalno zdravlje i ako ispitanici koji imaju dijagnozu depresije manje od drugih odgovaraju na pitanja vezana za njihov mentalni status, onda podaci nedostaju na neslučajan način. Jasno je da prosečni rezultat mentalnog statusa dostupnih podataka neće biti postojana ocena mentalnog statusa koji bi se dobio u slučaju kompletnih podaka. Ovakva struktura podataka predstavlja najveći izazov jer se moraju modelirati nedostajući podaci, odnosno mora se napraviti model koji uzima u obzir podatke koji nisu dostupni, ali to nimalo nije lak zadatak.

4.2 Metodi koji se primenjuju u slučajevima kada uzorak nije kompletan

4.2.1 Pametno brisanje sa liste (listwise deletion)

Ovaj način rešavanja problema je najjednostavniji i najčešći. U ovom slučaju se jednostavno zaobilaze podaci koji nedostaju i statistička analiza se izvodi samo sa raspoloživim podacima. Iako u ovom slučaju može doći do situacije da se značajno smanjila veličina uzorka potrebnog za analizu, ovaj pristup ipak ima odredjene prednosti. Konkretno, pod pretpostavkom da su podaci koji nedostaju tipa MCAR (missing completely at random), tj. da, kao što smo definisali u prethodnoj sekciji, nedostaju na potpuno slučajan način, ova metoda nas dovodi do ocene parametara bez pristrasnosti. Na žalost, postoji gubitak moći testa, u ovom slučaju upravo zbog toga što se smanjuje veličina uzorka, čak i pod pretpostavkom da su podaci tipa MCAR, s obzirom da su t-testovi funkcije veličine uzorka. U slučaju da podaci nisu MCAR, pojavljuje se pristrasnost. Alternativni pristupi koje ćemo objasniti u narednim paragrafima mogu biti u odredjenim situacijama pogodna zamena za pametno brisanje sa liste.

4.2.2 Brisanje u parovima (pairwise deletion)

Druga metoda ima izvesne nedostatke u odnosu na prethodno opisanu i obično se naziva "*unwise*" (neinteligentna) metoda. Mnogi kompjuterski paketi nude opciju koja se naziva *pairwise deletion.* Ona se zasniva na tome da se ocenjuje svaki element u interkorelacionoj matrici pri čemu se koriste svi raspoloživi podaci. Problem ovog pristupa je da će ocenjeni parametri modela biti zasnovani na različitim skupovima podataka, sa različitim veličinama uzorka i različitim standardnim greškama. Čak je sasvim moguće generisati interkorelacionu matricu koja nije pozitivno definitna, što može lako dovesti čitavu analizu u pitanje. Ova se metoda može upotrebiti u slučaju da nedostaje vrlo malo podataka, medjutim u tom slučaju se preporučuje da se koristi prethodna metoda tako što će se elementi koji nedostaju prosto ukloniti iz analize. Ukoliko nedostaje veći broj podataka onda se ovim dvema opisanim metodama može značajno ugroziti analiza. U takvim slučajevima upotrebljavaju se drugačiji pristupi opisani u daljem tekstu.

4.2.3 Zamena srednjom vrednošću (Mean Substitution)

Zamena srednjom vrednošću je postupak koji pri kome se vrši zamena svih nedostajućih podataka date veličine srednjom vrednošću te veličine. Ova metoda je dobar izbor u slučaju da se radi o normalno raspodeljenim podacima tipa MAR. Takodje, ovaj postupak proizvodi mnogo konzistentniji skup korelacionih matrica nego metodom brisanja po parovima. Kao u slučaju *pametnog brisanja sa liste*, kada je proporcija nedostajućih podataka veća, javiće se pristrasnost, čak i kad su u pitanju podaci tipa MAR.

4.2.4 Imputacija pomoću regresije (Imputation by Regression)

Za predvidjanje nedostajućih podataka ova metoda koristi regresionu jednačinu u kojoj se kao nezavisne veličine pojavljuju svi ostali relevantni podaci kao prediktori. Prednost ove metode je što zadržava varijansu i kovarijansu slučajnih veličina sa nedostajućim podacima. Medjutim, ukoliko se zanemare standardne greške, može se značajno oslabiti predvidjajuća moć modela s obzirom da se za vrednosti zavisne veličine predpostavlja da su prognozirane sa savršenom tačnošću.

4.2.5 Slučajna imputacija (Hot Deck Imputation)

Ova metoda predstavlja zamenu nedostajućih podataka slučajno izabranim vrednostima iz nekog sličnog kompletnog uzorka. S obzirom sa se zamenjujuće vrednosti biraju na slučajan način, slučajna imputacija dovodi do većih varijacija od srednje vrednosti. Ovo je metoda koja se zasniva na zamenjivanju nedostajućih vrednosti na slučajan način onim vrednostima koji potiču iz sličnog slupa podataka. Ovaj termin "*hot deck*" datira još iz perioda kada su podaci skladišteni na bušenim karticama i indikuje da podaci supstitucije potiču iz istog skupa kao i neregistrovani podaci. Reč "*hot*" je iz razloga što se trenutno procesuiraju dati podaci. Nasuprot ovoj metodi, postoji i *Cold Deck Imputation* metoda pri kojoj se podaci substituišu podacima iz nekog drugog skupa. Osnovni problem koji se nameće prilikom korišćenja ove metode je kako odabrati reprezentativan skup i identifikovati odgovarajuće vrednosti koje će se koristiti u substituciji i proizvesti najmanju varijansu. Ova tehnika se intenzivno koristi od strane vladinih agencija i u velikoj meri je zadovoljavajuća s obzirom da uspešno proizvodi reprezentativne uzorke populacije koju ispituje.

4.2.6 EM Algoritam (Expectation Maximization Algorithm)

Značenje ove metode je maksimizacija očekivanja, iterativni proces koji ocenjuje parametre modela polazeći od neke početne vrednosti. Svaka iteracija se sastoji iz dva koraka:

1. korak očekivanja, pri kome se identifikuje raspodela nedostajućih podataka na osnovu poznatih vrednosti opserviranih podataka i na osnovu trenutne ocene parametara
2. korak maksimizacije koji substituiše nedostajuće podatke očekivanom vrednošću.

Metoda predstavlja elegantan i moćan pristup, medjutim zahteva specijalizovan i efikasan softver s obzirom da može biti veoma vremenski zahtevna.

4.2.7 FIML Metod (Raw Maximum Likelihood or Full Information Maximum Likelihood)

FIML metod je obično predstavljen kao kovarijansna matrica promenljive i vektora očekivanja. Tehnika koristi sve raspoložive informacije o posmatranim podacima, kao i očekivanje i varijansu na osnovu raspoloživih podataka za svaku promenljivu. Prednost u odnosu na EM metode je u tome što omogućava direktno izračunavanje odgovarajućih standardnih grešaka i testiranje statistike. FIML metod, koji se još naziva i "direktni metod maksimalne verodostojnosti", "grubi metod maksimalne verodostojnosti" ili samo "ML", je trenutno dostupan u svim većim statističkim paketima koji se koriste u imputaciji nedostajućih vrednosti. Postupak zahteva da podaci budu bar tipa MAR ili MCAR. Vrši se ocenjivanje funkcije verodostojnosti za svakog ispitanika na osnovu registrovanih varijabli ali tako da se iskoriste svi raspoloživi podaci. Ovaj postupak takodje zahteva specijalizovani program i može biti veoma vremenski intenzivan.

4.2.8 Višestruke imputacije (Multiple Imputations)

Višestruke imputacije - slične su EM algoritmu, generišu maksimalnu verodostojnost zasnovanu na kovarijansnoj matrici vektora srednjih vrednosti. Razlika je u tome što se zahteva izgradnja pet do deset baza podataka sa odgovarajućim vrednostima, od kojih svaka mora biti pojedinacno analizirana. Rezultati se zatim kombinuju u jednom skraćenom skupu vrednosti. Postupak je veoma moćan, medjutim vremenski jako zahtevan. Višestruka imputacija je vrlo korisna i primenljiva strategija koja se koristi u radu sa nekompletnim uzorcima. To je metoda koja potiče od Rubina (Rubin (1987)), koi umesto popunjavanja nedostajućeg podatka jednom vrednošću, predlaže zamenu skupom odgovarajućih vrednosti. Ovi višestruko umetnuti skupovi podataka se zatim analiziraju standardnim procedurama za kompletne podatke i kombinuju sa već utvrdjenim rezultatima statističke analize. Proces kombinovanja rezultata iz različitih skupova je u suštini isti, bez obzira o kojoj se statističkoj analizi za kompletne podatke radi.

4.3 Istorijski razvoj i literatura

Sve do 70-tih godina prošlog veka se nije ozbiljno pristupalo problemu nekompletnosti uzoraka. Rubin (1976) formira odredjenu vrstu algoritma za tretiranje nedostajućih podataka koja se na izvestan način zadržala do danas.

Formulacija EM algoritma (Dempster, Laird, Rubin (1977)) je omogućila izračunavanje ocena maksimalne verodostojnosti u različitim problemima nekompletnih uzoraka. Mnogi primeri EM algoritama su opisani u knjizi Little, Rubin (1987). Takodje, pogledati i Rubin (1976, 1987).

Od 1990-te godine kreće razvoj novih i modernijih pravaca u tretiranju problema nedostajućih podataka (Ibrahim (1990)). Praktične aplikacije i analiza softvera za obradu ovakvih tipova podataka može se pogledati u Schafer (1997, 1999) i Schafer, Olsen (1998). Videti takodje Robins, Rotnitzky, Zhao (1994). Nove metode se naročito primenjuju u biostatistici i javnom zdravlju (Little (1995)). Pogledati još:

Dunsmuir, Robinson (1981), Kline, Santos (2010), Mladenović, Petrović (2010), Koopman i dr. (2007), Verbeke, Molenberghs (2000), Graham, Hofer (2000), Little, Rubin (2002), Schafer, Graham (2002), Graham i dr. (2003), Graham (2009) i Wang, Luo (2011).

Poglavlje 5

OCENJIVANJE INDEKSA REPA RASPODELE NA NEKOMPLETNOM UZORKU

U ovom poglavlju bavimo se ocenjivanjem parametra oblika α koristeći niz zavisnih slučajnih veličina sa zajedničkom funkcijom raspodele teškog repa.

Pod pretpostavkom " *ekstremalne zavisnosti*" posmatra se asimptotsko ponašanje Hilove ocene i ocene geometrijskog tipa koju su predložili Brito i Freitas (2003), a koja se pokazala manje osetljivom na devijacije sporo promenljive funkcije od konstante (pogledati Csorgo i Viharos (1998)). Uz korišćenje nekompletnog uzorka, dokazuje se konzistentnost ocene geometrijskog tipa, kao i konzistentnost i asimptotska normalnost Hilove ocene. Kao osnova dokaza, koristi se model koji je prezentovan u radu Mladenović i Piterbarg (2008) u kome je na nepotpuno registrovanom uzorku dokazana konzistentnost Hilove ocene uz pretpostavku stroge stacionarnosti. Autori su eksploatisali Hsinga (1991) koji je posmatrao asimptotsko ponašanje Hilove ocene na stacionarnom uzorku. Generalno, u uslovima teorema ovog poglavlja ne zahteva se stacionarnost, s obzirom da pretpostavka zavisnosti u domenu visokih kvantila pokriva široki spektar procesa. Posmatraju se procesi čije su ekstremne veličine

NED (near epoch dependent) na nekom proizvoljnom *mešajućem* funkcionalu (detaljnije objašnjenje dato je u prvoj sekciji ovog poglavlja). Mnoge vremenske serije u finansijama, makroekonomiji i meteorologiji imaju ekstremne vrednosti koje se pojavljuju u klasterima. Ovaj uslov takodje ima i praktičnih prednosti, s obzirom da se lako verifikuje na osnovu uslovnog očekivanja. U poslednjoj sekciji poglavlja dati su brojni primeri slučajnih nizova koji zadovoljavaju ovaj uslov i na kojima su teoreme primenljive.

5.1 EKSTREMALNA ZAVISNOST

Na početku ovog paragrafa upoznaćemo se sa definicijom odredjene vrste asimptotske nezavisnosti, odnosno sa pojmom jake pomešanosti ili α-pomešanosti.

Definicija 5.1 (α**-pomešanost**)**.** Neka je dat niz $\{X_t\}$ slučajnih veličina na nekom prostoru verovatnoća $(\Omega,\, \mathcal{F},\, \mathcal{P})$. Neka za $-\infty \leq \jmath \leq \ell \leq \infty$ $\mathcal{F}_\jmath^\ell$ predstavlja σ-polje indukovano slučajnim veličinama X_k, gde je $\jmath \leq k \leq \ell$ i $k \in Z$.

Niz slučajnih veličina $\{X_t\}$ zadovoljava uslov α-pomešanosti ukoliko

$$\alpha(d) = \sup_{t \in Z} \sup_{A \in \mathcal{F}_{-\infty}^t, B \in \mathcal{F}_{t+d}^{+\infty}} |P(A \cap B) - P(A)(B)| \to 0,$$

kad $d \to \infty$.

Neka je $\Im_t := \sigma(\epsilon_\tau : \tau \leq t)$ σ-polje indukovano nekom bazom ϵ_t koja zadovoljava uslov α-pomešanosti. Dalje, neka je q_n proizvoljan niz celih brojeva koji zadovoljava $1 \leq q_n < n$ i $q_n \to \infty$, kad $n \to \infty$. Na primer, možemo pretpostaviti da je $\epsilon_t = I(X_t > b_{k_n} e^u)$ gde ϵ_t zadovoljava uslov α-promešanosti i gde je $b_{k_n} := F^{-1}(1 - k_n/n)$. Ovaj primer koji smo naveli predstavlja specijalan slučaj koji se pojavljuje u Hsingovom radu (1991) i njegovi uslovi pomešanosti odnose se na ovakve tipove ϵ_t baze. Medjutim, dokazi ovog poglavlja se intenzivno baziraju na Hilovom radu (Hill 2010, Teorema 1.) koji pretpostavlja nove uslove zavisnosti u repu raspodele na sličnoj bazi $\epsilon_t = I(X_t > b_{k_n} e^u)$, koji pokrivaju i značajno uopštavaju Hsingove pretpostavke mešanja (pogledati takodje i Hill (2011)).

L_p-Extremal-Near Epoch Dependence (L_p-E-NED) svojstvo. $\{X_t\}$ predstavlja L_p-E-NED na $\{\Im_t\}$, $p > 0$, veličine $\lambda > 0$ ukoliko važi

$$\|I(X_t > b_{k_n} e^u) - P(X_t > b_{k_n} e^u | \Im_{t-q_n}^{t+q_n})\|_p \leq f_{nt}(u) \times \psi_{q_n},$$

gde je $\psi_{q_n} = o(q_n^{-\lambda})$ i $f_{nt} : R_+ \to R_+$ je merljiva u Lebegovom smislu.

$\sup_{1 \leq t \leq n} \sup_{u \geq 0} f_{nt}(u) = O((k_n/n)^{1/p})$ (pogledati Hill (2010), p.1402).

Pretpostavka A. $\{X_t\}$ je L_2-E-NED na $\{\Im_t\}$, sa koeficijentima ψ_{q_n} veličine 1/2 i konstantama $f_{nt}(u)$ gde je $f_{nt} : R_+ \to R_+$ merljiva smislu Lebega i

$\sup_{1 \leq t \leq n} \int_0^\infty f_{nt}(u) du = O((k_n/n)^{1/2})$.

E-NED pokriva NED i $\alpha-$zavisnost, kao i nelinearne ARCH i GARCH procese. (videti Hill (2010, 2011)).

5.2 MODEL NEKOMPLETNOG UZORKA

Ukoliko je uslov (1.1) zadovoljen, može se lako pokazati (videti Leadbetter i dr. (1983), Teorema 1.5.1 i 1.7.3) da je

$$\lim_{x \to \infty} \frac{1 - F(x)}{1 - F(x - 0)} = 1, \tag{5.1}$$

i

$$1 - F(F^{-1}(1 - \frac{1}{t})) \sim \frac{1}{t} \quad \text{as} \quad t \to \infty, \tag{5.2}$$

gde je $F^{-1}(y) := \inf\{x : F(x) \geq y\}$ levi neprekidni inverz funkcije F. S obzirom da nam je u zaključivanjima i procenama parametra α neophodan deo repa empirijske funkcije raspodele uzorka, bez gubljenja opštosti pretpostavićemo da je F definisana na intervalu $(0, \infty)$.

Sada pretpostavimo da su nam dostupne samo observacije u nekim tačkama. Sada označimo registrovane slučajne promenljive medju $\{X_1, \ldots, X_n\}$ sa $\widetilde{X}_1$, ..., $\widetilde{X}_{S_n}$. Ovde slučajna veličina S_n predstavlja broj registrovanih slučajnih promenljivih medju prvih n članova niza $\{X_t\}$.

Nepotpun uzorak se može dobiti ukoliko je verovatnoća observacije svakog člana niza $\{X_t\}$ jednaka p, nezavisno od ostalih elemenata niza. U ovom slučaju S_n ima binomnu raspodelu, tj. predstavlja binomnu slučajnu varijablu. Medjutim, pretpostavićemo da su registrovane veličine definisane nekim tačkastim procesom i samo ćemo zahtevati uslove koji se odnose na slučajnu veličinu S_n. Ovakav model definisan je u radu Mladenovic i Piterbarg (2008) gde je dokazana konzistencija Hilove ocene na zavisnom nekompletnom uzorku i on najviše odgovara tipu MCAR nedostajućih podataka o kome smo govorili u paragrafu 4.1.1 prethodnog poglavlja.

Pretpostavka B. Niz X_1, X_2, $\ldots$ ne zavisi od S_n i

$$\frac{S_n}{n} \xrightarrow{p} c_0 > 0 \quad \text{as } n \to +\infty.$$

Neka je β_n niz realnih brojeva takav da $\lim_{n\to\infty} \beta_n = \infty$ i $\lim_{n\to\infty} \beta_n/n = 0$. Definišimo još dve slučajne veličine

$$M_n = \left[\frac{S_n}{\beta_n}\right] \quad \text{i } B_n = \begin{cases} 0, & S_n = 0 \\ \frac{M_n}{S_n}, & S_n \geq 1, \end{cases}$$

gde funkcija [.] označava ceo deo datog broja. Interesuje nas ocena za parametar debljine repa raspodele α koju ćemo dobiti uz pomoć odredjene porcije uzorka.

Neka je $\widetilde{X}_{(1)} \geq \widetilde{X}_{(2)} \geq ... \geq \widetilde{X}_{(S_n)}$ niz statistika poretka od S_n registrovanih slučajnih veličina.

Označimo

$$x_+ = max(x, 0),\ x_- = max(-x, 0). \tag{5.3}$$

Hilova ocena je data sa:

$$H_{S_n} = I\{S_n \geq \beta_n\} \frac{1}{M_n} \sum_{t=1}^{M_n} \ln \widetilde{X}_{(t)} - \ln \widetilde{X}_{(M_n+1)},$$

a ocena geometrijskog tipa sa

$$\widehat{R}(S_n) = I\{S_n \geq \beta_n\} \sqrt{\frac{\sum_{t=1}^{M_n} \ln^2(S_n/t) - \frac{1}{M_n}(\sum_{t=1}^{M_n} \ln(S_n/t))^2}{\sum_{t=1}^{M_n} \ln^2 \widetilde{X}_{(t)} - \frac{1}{M_n}(\sum_{t=1}^{M_n} \ln \widetilde{X}_{(t)})^2}}. \tag{5.4}$$

Definišimo još dve slučajne veličine u skladu sa gore opisanim modelom:

$$\widetilde{H}_{S_n} = I\{S_n \geq \beta_n\} \frac{1}{M_n} \sum_{t=1}^{M_n} \ln \widetilde{X}_{(t)} - \ln F^{-1}(1 - B_n)$$

i

$$H^+_{S_n} = I\{S_n \geq \beta_n\} \frac{1}{M_n} \sum_{t=1}^{S_n} (\ln \widetilde{X}_t - \ln F^{-1}(1 - B_n))_+.$$

Prema rezultatima iz Mladenovic i Piterbarg (2008) sve tri veličine H_{S_n}, $\widetilde{H}_{S_n}$ i $H^+_{S_n}$ imaju isto asimptotsko ponašanje u raspodeli.

5.3 REZULTATI

Označimo

$$\widetilde{Y}_{nt} = (\ln \widetilde{X}_t - \ln F^{-1}(1 - B_n))_+$$

i

$$I_{nt} = I\left\{\ln \widetilde{X}_t - \ln F^{-1}(1 - \rho B_n) > \varepsilon\right\},$$

gde $\varepsilon \in R$ i $\rho \in J$, gde J označava neku okolinu broja 1. $I\{\}$ predstavlja oznaku funkcije indikatora.

Lema 5.2. *Ukoliko važi* Pretpostavka B*, onda za bilo koji ceo broj $k > 0$ važi*

$$\lim_{n\to\infty} \beta_n E(\ln \widetilde{X}_t - \ln F^{-1}(1 - B_n))_+^k = \frac{k!}{\alpha^k}.$$

Primedba 5.3. Primetimo da, ukoliko je $B_n = 0$, onda je $\ln F^{-1}(1 - B_n) = +\infty$ i prema (5.3) slučajna veličina $\widetilde{Y}_{nt}$ dobija vrednost 0.

Teorema 5.4. *Pretpostavimo da F zadovoljava (1.1). Neka takodje važe* Pretpostavke A *i* B*. Tada sve tri veličine H_{S_n}, $H_{S_n}^+$ i $\widetilde{H}_{S_n}$ konvergiraju ka α u verovatnoći.*

Primedba 5.5. Primetimo da, ukoliko niz $\{X_t\}$ zadovoljava uslov α-pomešanosti, tada taj uslov zadovoljava i niz $\{\widetilde{X}_t\}$, s obzirom da važi

$$\sup_{s\in Z} \sup_{A\in\widetilde{\mathfrak{F}}_{-\infty}^{s}, B\in\widetilde{\mathfrak{F}}_{s+d}^{+\infty}} |P(A\cap B) - P(A)(B)| \leq \sup_{t\in Z} \sup_{A\in\Im_{-\infty}^{t}, B\in\Im_{t+d}^{+\infty}} |P(A\cap B) - P(A)(B)|,$$

gde je $\widetilde{\mathfrak{F}}_s := \sigma(\widetilde{\epsilon}_\tau : \tau \leq s)$ i $\widetilde{\epsilon}_s$ proizvoljan funkcional od $\widetilde{X}_s$.

Lema 5.6. *Posmatrajmo niz sledećih slučajnih veličina:*

$$t(M_n) := \frac{1}{M_n}\sum_{t=1}^{M_n} ln^2(S_n/t) - (\frac{1}{M_n}\sum_{t=1}^{M_n} ln(S_n/t))^2,$$

gde je M_n definisano u prethodnoj sekciji. Tada $E(t(M_n)) \to 1$, kad $n \to \infty$.

Teorema 5.7. *Pretpostavimo da F zadovoljava (1.1). Neka takodje važe* Pretpostavke A *i* B*. Tada $\widehat{R}(S_n)$ konvergira ka α^{-1} u verovatnoći.*

Pretpostavka C. Postoji pozitivna merljiva funkcija g definisana na intervalu $(0, \infty)$ takva da za proizvoljno $\lambda > 0$ važi:

$$L(\lambda x)/L(x) - 1 = O(g(x)),$$

kad $x \to \infty$. Takodje, postoje konstante $D > 0$, $z_0 < \infty$ i $\tau \leq 0$ takve da važi $g(\lambda z)/g(z) \leq D\lambda^{\tau}$ za neko $\lambda \geq 1$ i neko $z \geq z_0$. Potrebno je da k_n, b_{k_n} i g zadovoljavaju da $k_n^{1/2} g(b_{k_n}) \to 0$.

Teorema 5.8. *Neka važe* Pretpostavke A, B *i* C. *Tada*

$$M_n^{1/2}(H_{S_n} - \alpha^{-1})/\sigma_{M_n} \xrightarrow{d} N(0,1),$$

gde je $\sigma_{M_n}^2 = E(M_n^{1/2}(H_{S_n} - \alpha^{-1}))^2 = O(1)$. *Takodje,*

$$|\sigma_{M_n}^2 - E(I\{S_n \geq \beta_n\} \frac{1}{M_n^{1/2}} [\sum_{t=1}^{S_n} \{\widetilde{Y}_t - E(\widetilde{Y}_t) - \alpha^{-1}(\widetilde{Y}_t^{\zeta} - E(\widetilde{Y}_t^{\zeta}))\}])^2| \to 0,$$

gde je $\widetilde{Y}_t^{\zeta} = I\left\{\ln \widetilde{X}_t - \ln F^{-1}(1 - B_n) > \frac{\zeta}{\sqrt{M_n}}\right\}$ *i* $\zeta \in R$.

Primer 5.1. *Model koji se često koristi u aplikacijama je GARCH (1,1) proces. Pod nekim veoma opštim uslovima koji se odnose na šum* (Z_t) *GARCH (1,1) proces ima marginalnu raspodelu Paretovog tipa i kao takav predstavlja veoma atraktivan alat u modelovanju finansijskih podataka čija je empirijska raspodela u repu deblja od normalne raspodele. Ovaj proces se definiše odredjivanjem* σ_t *na sledeći način:*

$$\sigma_t^2 = \alpha_0 + \beta_1\sigma_{t-1}^2 + \alpha_1 X_{t-1}^2 = \alpha_0 + \sigma_{t-1}^2(\beta_1 + \alpha_1 Z_{t-1}^2),$$

gde $t \in Z$, *a parametri* α_0, α_1 *i* β_1 *su nenegativni. GARCH (1,1) proces je* L_2*-NED ako je*

$$\beta_1^2 + 2\alpha_1\beta_1 + 3\alpha_1^2 < 1$$

kad su Z_t *nezavisne slučajne veličine sa istom normalnom funkcijom raspodele (videti Davidson (2004)). Takodje Hill (2005, Lema 7) dokazuje da procesi sa pravilno promenljivim repom koji su* L_2*-NED takodje imaju i* L_2*-E-NED osobinu. Takodje, pošto su GARCH (1,1) procesi jako mešajući oni su automatski E-NED na sopstvenom funkcionalu.*

Primer 5.2. *Drugi model koji se takodje koristi u aplikacijama je nelinearni autoregresivni model (AR). Za nelinearne modele*

$$y_t = f(y_{t-1}) + \varepsilon_t$$

je moguće pokazati L_2*-NED svojstvo, za bilo koji niz slučajnih veličina koje zadovoljavaju model za* $t = 1, ..., n$ *i imaju proizvoljnu početnu vrednost* y_0*, takvu da važi* $E(y_0)^2 < \infty$*. Takodje, mora biti zadovoljeno da*

$$|f(x) - f(y)| \leq L|x - y|,$$

gde je $f(\cdot)$ *Borel merljiva i* $L < 1$ *(videti Tjostheim (1990)).*

5.4 DOKAZI

Dokaz Leme 5.2. Primetimo da važe sledeće jednakosti

$$\begin{aligned} E(\ln \widetilde{X}_t - \ln F^{-1}(1 - B_n))_+^k &= \int_0^\infty P\{(\ln \widetilde{X}_t - \ln F^{-1}(1 - B_n))^k > u\} \, du \\ &= \int_0^\infty P\{(\ln \frac{\widetilde{X}_t}{F^{-1}(1 - B_n)})^k > u\} \, du \\ &= \int_0^\infty P\{\frac{\widetilde{X}_t}{F^{-1}(1 - B_n)} > e^{u^{1/k}}\} \, du. \end{aligned}$$

Pošto je

$$\left\{e^{u^{1/k}} < \frac{\widetilde{X}_t}{F^{-1}(1 - B_n)}\right\} \subset \left\{e^{u^{1/k}} < \frac{\widetilde{X}_t}{F^{-1}(1 - 1/\beta_n)}\right\},$$

dobijamo da je

$$\begin{aligned}
&E(\ln \widetilde{X}_t - \ln F^{-1}(1-B_n))_+^k \\
&\leq \int_0^\infty P\left\{\frac{\widetilde{X}_t}{F^{-1}(1-1/\beta_n)} > e^{u^{1/k}}\right\} du \\
&= \int_0^\infty P\left\{\widetilde{X}_t > e^{u^{1/k}} F^{-1}(1-1/\beta_n)\right\} du \\
&= \int_0^\infty \left\{1 - F(e^{u^{1/k}} F^{-1}(1-1/\beta_n))\right\} du \\
&= \{1 - F(F^{-1}(1-1/\beta_n))\} \int_0^\infty \frac{1 - F(e^{u^{1/k}} F^{-1}(1-1/\beta_n))}{1 - F(F^{-1}(1-1/\beta_n))} du \\
&\sim \frac{1}{\beta_n} \int_0^\infty e^{-\alpha u^{1/k}} = \frac{1}{\beta_n} \frac{k!}{\alpha^k}.
\end{aligned}$$

Označimo $J_\varepsilon = [n(c_0 - \varepsilon), n]$ i $k_n = [n\beta_n^{-1}(c_0 - \varepsilon)]$ gde je $0 < \varepsilon < c_0$. Važi da je

$$\{S_n \geq n(c_0 - \varepsilon)\} \subset \{S_n \geq k_n \beta_n\} \subset \left\{B_n > \frac{k_n}{k_n+1}\frac{1}{\beta_n}\right\}$$

i

$$\begin{aligned}
\{S_n \in J_\varepsilon\} \quad &\subset \quad \left\{B_n > \frac{k_n}{k_n+1}\frac{1}{\beta_n}\right\} \\
&= \quad \left\{F^{-1}(1-B_n) \leq F^{-1}\left(1 - \frac{k_n}{k_n+1}\frac{1}{\beta_n}\right)\right\} \\
&= \quad \left\{\frac{1}{F^{-1}(1-B_n)} \geq \frac{1}{F^{-1}\left(1 - \frac{k_n}{k_n+1}\frac{1}{\beta_n}\right)}\right\}.
\end{aligned}$$

Kao posledicu dobijamo

$$\left\{\frac{\widetilde{X}_t}{F^{-1}(1-B_n)} > e^{u^{1/k}},\, S_n \in J_\varepsilon\right\} \supset \left\{\frac{\widetilde{X}_t}{F^{-1}\left(1 - \frac{k_n}{k_n+1}\frac{1}{\beta_n}\right)} > e^{u^{1/k}},\, S_n \in J_\varepsilon\right\}.$$

Konačno sledi da je

$$
\begin{aligned}
&E(\ln \widetilde{X}_t - \ln F^{-1}(1-B_n))_+^k \\
&\geq P\{S_n \in J_\varepsilon\} \int_0^\infty P\left\{ \frac{\widetilde{X}_t}{F^{-1}\left(1 - \frac{k_n}{k_n+1}\frac{1}{\beta_n}\right)} > e^{u^{1/k}} \right\} du \\
&\sim P\{S_n \in J_\varepsilon\} \frac{k_n}{k_n+1} \frac{1}{\beta_n} \frac{k!}{\alpha^k} \\
&\sim \frac{k_n}{k_n+1} \frac{1}{\beta_n} \frac{k!}{\alpha^k} \\
&\sim \frac{1}{\beta_n} \frac{k!}{\alpha^k}, \quad n \to \infty. \quad \blacksquare
\end{aligned}
$$

Pre nego što formulišemo dokaz *Teoreme 5.4.*, uvodimo pojam Extremal-Mixingale (E-MIXL) procesa.

L_p-Extremal-Mixingale (L_p-E-MIXL) svojstvo. $\{X_t, \Im_t\}$ predstavlja niz sa L_p-E-MIXL svojstvom gde je $p > 0$, $\lambda > 0$ ukoliko važi

$$
\| P(X_t > b_{k_n} e^u) - P(X_t > b_{k_n} e^u | \Im_{t-q_n}) \|_p \leq e_{nt}(u) \times \phi_{q_n}
$$

i

$$
\| I(X_t > b_{k_n} e^u) - P(X_t > b_{k_n} e^u | \Im_{t+q_n}) \|_p \leq e_{nt}(u) \times \phi_{q_n+1},
$$

gde je $\phi_{q_n} = o(q_n^{-\lambda})$, $e_{nt} : R_+ \to R_+$ je merljiva u Lebegovom smislu i

$\sup_{1 \leq t \leq n} \sup_{u \geq 0} e_{nt}(u) = O((k_n/n)^{1/p})$.

Dokaz Teoreme 5.4. Prema identičnom argumentu iz Dejvidsonovog rada (videti Davidson (1994), *Teorema 17.5*) lako je pokazati da L_2-E-NED pretpostavka povlači L_2-E-MIXL pretpostavku, što u našem slučaju znači da je niz $\{X_t\}$ L_2-E-MIXL na $\{\Im_t\}$ sa koeficijentima ψ_{q_n} veličine $1/2$ i konstantama $e_{nt}(u)$ gde je $e_{nt} : R_+ \to R_+$ Lebeg merljiva i $\sup_{1 \leq t \leq n} \int_0^\infty e_{nt}(u) du = O((k_n/n)^{1/2})$.

Prema sličnim argumentima kao kod Hila (Hill (2010), *Lemma B.1.*) i prema *Pretpostavci A* važi da $\{(\widetilde{Y}_{nt} - E(\widetilde{Y}_{nt})), \mathfrak{F}_t\}$ i $\{(I_{nt} - E(I_{nt})), \mathfrak{F}_t\}$ za svako ρ u proizvoljnoj okolini broja 1 formiraju nizove sa L_2-E-MIXL svojstvom veličine 1/2 i nekim konstantama $\{\widetilde{e}^*_{nt}, \widetilde{e}_{nt}(u)\}$. Tada, prateći dokaz iz Hila (Hill (2010), *Lema 1.*) i koristeći posledicu iz Dejvidsonovog rada (Davidson (1994), Corollary 20.16) dobijamo da

$$I\{S_n \geq \beta_n\}\frac{1}{M_n}\sum_{t=1}^{S_n}(\widetilde{Y}_{nt} - E(\widetilde{Y}_{nt})) \xrightarrow{P} 0$$

i

$$I\{S_n \geq \beta_n\}\frac{1}{M_n}\sum_{t=1}^{S_n}(I_{nt} - E(I_{nt})) \xrightarrow{P} 0$$

za svako $\varepsilon \in R$ i $\rho \in J$, gde je J neka okolina broja 1. Konačno, zaključak teoreme dobijamo kao posledicu *Teoreme 1.* iz rada Mladenović i Piterbarg (2008). ■

Dokaz Leme 5.5. Primetimo da je

$$t(M_n) := \frac{1}{M_n}\sum_{t=1}^{M_n}\ln^2(t) - (\frac{1}{M_n}\sum_{t=1}^{M_n}\ln(t))^2.$$

Pošto važi da je

$$\begin{aligned}
E(\frac{1}{M_n}\sum_{t=1}^{M_n}\ln^2(t)) &= E(E(\frac{1}{M_n}\sum_{t=1}^{M_n}\ln^2(t)|M_n = m_n)) \\
&= \sum_{m_n} E(\frac{1}{M_n}\sum_{t=1}^{M_n}\ln^2(t)|M_n = m_n)P\{M_n = m_n\} \\
&= \sum_{m_n} E(\frac{1}{m_n}\sum_{t=1}^{m_n}\ln^2(t))P\{M_n = m_n\} \\
&= \sum_{m_n} \frac{1}{m_n}\sum_{t=1}^{m_n}\ln^2(t)P\{M_n = m_n\} \\
&\leq \sum_{m_n} \frac{1}{m_n}\int_1^{m_n+1}\ln^2(t)dtP\{M_n = m_n\}
\end{aligned}$$

i

$$
\begin{aligned}
E(\frac{1}{M_n}\sum_{t=1}^{M_n}\ln(t)) &= \sum_{m_n}\frac{1}{m_n}\sum_{t=1}^{m_n}\ln(t)P\{M_n=m_n\} \\
&= \sum_{m_n}\frac{1}{m_n}\sum_{t=2}^{m_n}\ln(t)P\{M_n=m_n\} \\
&\geq \sum_{m_n}\frac{1}{m_n}\int_1^{m_n}\ln(t)dtP\{M_n=m_n\},
\end{aligned}
$$

jednostavnom računicom dobijamo da je

$$
\begin{aligned}
t(M_n) &\leq \sum_{m_n}[\frac{1}{m_n}\int_1^{m_n+1}\ln^2(t)dt-\frac{1}{m_n^2}(\int_1^{m_n}\ln(t)dt)^2]P\{M_n=m_n\} \\
&\sim \sum_{m_n}P\{M_n=m_n\}=1,
\end{aligned}
$$

kad $n\to\infty$. Videti Brito i Freitas (2003, *Lema 2.*). ■

Primetimo da se $\widehat{R}(S_n)$ može napisati u sledećem obliku:

$$
\widehat{R}(S_n)=I\{S_n\geq\beta_n\}\sqrt{\frac{t(M_n)}{\frac{1}{M_n}\sum_{t=1}^{M_n}\ln^2\widetilde{X}_{(t)}-\frac{1}{M_n^2}(\sum_{t=1}^{M_n}\ln\widetilde{X}_{(t)})^2}}. \tag{5.5}
$$

Prema *Lemi 5.5.*, da bismo dokazali konzistenciju $\widehat{R}(S_n)$ potrebno je da pokažemo da niz $\widehat{N}_{S_n}$ koji je definisan sa

$$
\begin{aligned}
\widehat{N}_{S_n} &:= \frac{1}{M_n}\sum_{t=1}^{M_n}\ln^2\widetilde{X}_{(t)}-\frac{1}{M_n^2}(\sum_{t=1}^{M_n}\ln\widetilde{X}_{(t)})^2 \\
&= \frac{1}{M_n}\sum_{t=1}^{M_n}(\ln\widetilde{X}_{(t)}-\ln F^{-1}(1-B_n))^2-\frac{1}{M_n^2}(\sum_{t=1}^{M_n}(\ln\widetilde{X}_{(t)}-\ln F^{-1}(1-B_n)))^2,
\end{aligned}
$$

konvergira ka $1/\alpha^2$ u verovatnoći.

Dokaz Teoreme 5.6. Označimo

$$N^+_{S_n} := \frac{1}{M_n}\sum_{t=1}^{S_n}(\ln\widetilde{X}_{(t)} - \ln F^{-1}(1-B_n))^2_+ - \frac{1}{M_n^2}(\sum_{t=1}^{S_n}(\ln\widetilde{X}_{(t)} - \ln F^{-1}(1-B_n))_+)^2.$$

Koristeći *Teoremu 5.4.* i uslove *Teoreme 5.6.*, zaključujemo da $(H^+_{S_n})^2$ konvergira u verovatnoći ka $1/\alpha^2$. Prema *Lemi 5.2.* dobijamo

$$\frac{1}{\beta_n}E(\ln\widetilde{X}_{(t)} - \ln F^{-1}(1-B_n))^2_+ \to 2/\alpha^2,$$

kad $n \to \infty$. Prema sličnom rezonovanju kao u *Teoremi 1.* iz Dejvidsona (Davidson (1994), Corollary 20.16) zaključujemo da

$$\frac{1}{M_n}\sum_{t=1}^{S_n}(\ln\widetilde{X}_{(t)} - \ln F^{-1}(1-B_n))^2_+ \xrightarrow{P} 2/\alpha^2.$$

Na taj način dobijamo da

$$N^+_{S_n} := \frac{1}{M_n}\sum_{t=1}^{S_n}(\ln\widetilde{X}_{(t)} - \ln F^{-1}(1-B_n))^2_+ - (H^+_{S_n})^2 \xrightarrow{P} 1/\alpha^2.$$

Da bi se dokaz kompletirao, potrebno je još pokazati da $N^+_{S_n} - \widehat{N}_{S_n} \xrightarrow{P} 0$. Primetimo da je:

$$N^+_{S_n} - \widehat{N}_{S_n} = A_{S_n} + B_{S_n} + (\widetilde{H}_{S_n})^2 - (H^+_{S_n})^2,$$

gde je

$$A_{S_n} = -\frac{1}{M_n}\sum_{t=1}^{M_n}(\ln\widetilde{X}_{(t)} - \ln F^{-1}(1-B_n))^2_-$$

i

$$B_{S_n} = \frac{1}{M_n}\sum_{t=M_n+1}^{S_n}(\ln\widetilde{X}_{(t)} - \ln F^{-1}(1-B_n))^2_+.$$

Prema *Teoremi 5.4.* i uslovima *Teoreme 5.6.* dolazimo do zaključka da $(\widetilde{H}_{S_n})^2 - (H^+_{S_n})^2 \xrightarrow{P} 0$. Sada samo preostaje da pokažemo da $A_{S_n} \xrightarrow{P} 0$ i $B_{S_n} \xrightarrow{P} 0$.

Najpre, pokažimo da za svako $\rho \in J$, gde je J neka okolina broja 1 važi:

$$\ln\widetilde{X}_{([\rho M_n])} - \ln F^{-1}(1-\rho B_n) \xrightarrow{P} 0, \tag{5.6}$$

kad $n \to \infty$. U tom cilju dovoljno je pokazati da za svako $\rho \in J$ i $\varepsilon > 0$

$$P\{\ln\widetilde{X}_{([\rho M_n])} - \ln F^{-1}(1-\rho B_n) > +\varepsilon\} \to 0 \tag{5.7}$$

i

$$P\{\ln\widetilde{X}_{([\rho M_n])} - \ln F^{-1}(1-\rho B_n) < -\varepsilon\} \to 0 \tag{5.8}$$

kad $n \to \infty$. Da bi dokazali (5.7) možemo da napišemo

$$\begin{aligned} P\{\ln\widetilde{X}_{([\rho M_n])} - \ln F^{-1}(1-\rho B_n) > +\varepsilon\} &= P\{\sum_{t=1}^{S_n} I_{ni} \geq [\rho M_n]\} \\ = P\{\frac{1}{M_n}\sum_{t=1}^{S_n}(I_{ni} - E(I_{ni})) &\geq \frac{1}{M_n}([\rho M_n] - \sum_{t=1}^{S_n} E(I_{ni}))\}. \end{aligned} \tag{5.9}$$

Važi da $[\rho M_n]/M_n \xrightarrow{P} \rho$ kad $n \to \infty$. Na sličan način kao u dokazu *Leme 5.2.*, može se lako pokazati da $E(I_{nt}) \sim \frac{1}{\beta_n}\rho e^{-\alpha\varepsilon}$ kad $n \to \infty$.

Kao posledicu dobijamo: $\sum_{t=1}^{S_n} E(I_{ni}) \sim S_n/\beta_n \rho e^{-\alpha\varepsilon}$ kad $n \to \infty$ i

$$\frac{1}{M_n}([\rho M_n] - \sum_{t=1}^{S_n} E(I_{ni})) \xrightarrow{P} \rho(1 - e^{-\alpha\varepsilon}) > 0. \tag{5.10}$$

Koristeći relacije (5.9) i (5.10) dobijamo (5.7). Dokaz je vrlo sličan u slučaju relacije (5.8).

Lako se može primetiti da:

$$|A_{S_n}| \leq (\ln\widetilde{X}_{(M_n)} - \ln F^{-1}(1-B_n))^2_-$$

i za neko $\varepsilon > 0$, imamo:

$$P\{|A_{S_n}| > \varepsilon\} \leq P\{(\ln\widetilde{X}_{(M_n)} - \ln F^{-1}(1-B_n))^2_- > \varepsilon\}.$$

Ako sada iskoristimo relaciju (5.6) za $\rho = 1$, dobijamo da $A_{S_n} \xrightarrow{P} 0$.

Sada pokazujemo da $B_{S_n} \xrightarrow{P} 0$.

Za neko $\delta \in R^+$ takvo da $(1-\delta, 1+\delta) \subset J$, možemo da zapišemo

$$B_{S_n} = C_{S_n} + D_{S_n},$$

gde je

$$C_{S_n} = \frac{1}{M_n} \sum_{t=M_n+1}^{[(1+\delta)M_n]} (\ln\widetilde{X}_{(t)} - \ln F^{-1}(1-B_n))^2_+$$

i

$$D_{S_n} = \frac{1}{M_n} \sum_{t=[(1+\delta)M_n]+1}^{S_n} (\ln\widetilde{X}_{(t)} - \ln F^{-1}(1-B_n))^2_+.$$

Može se pokazati da oba niza C_n i D_n konvergiraju u verovatnoći ka 0. Iz implikacije

$$D_{S_n} > 0 \Rightarrow \ln\widetilde{X}_{([(1+\delta)M_n])} - \ln F^{-1}(1-B_n) > 0$$

sledi da

$$P\{D_{S_n} > 0\} \leq P\{\ln\widetilde{X}_{([(1+\delta)M_n])} - \ln F^{-1}(1-B_n) > 0\}.$$

Dalje važi da je

$$\begin{aligned}
&P\{\ln\widetilde{X}_{([(1+\delta)M_n])} - \ln F^{-1}(1-B_n) > 0\} \\
&\quad = P\{\ln\widetilde{X}_{([(1+\delta)M_n])} - \ln F^{-1}(1-(1+\delta)B_n) \\
&\quad > \ln F^{-1}(1-B_n) - \ln F^{-1}(1-(1+\delta)B_n)\}.
\end{aligned}$$

Prema (5.6) zaključujemo da

$$\ln\widetilde{X}_{([(1+\delta)M_n])} - \ln F^{-1}(1-(1+\delta)B_n) \xrightarrow{P} 0.$$

Takodje, pošto je F^{-1} pravilno veličina u ∞ sa parametrom $1/\alpha$, sledi da je

$$\lim_{n\to\infty} \ln\frac{F^{-1}(1-B_n)}{F^{-1}(1-(1+\delta)B_n)} = \frac{1}{\alpha}\ln(1+\delta) > 0.$$

Na taj način dobijamo da $P\{D_{S_n} > 0\} \to 0$ kad $n \to \infty$.

Uzimajući u obzir C_{S_n}, dobijamo da

$$\begin{aligned}
C_{S_n} &\leq \frac{1}{M_n}([(1+\delta)M_n] - M_n)(\ln\widetilde{X}_{(M_n+1)} - \ln F^{-1}(1-B_n))_+^2 \\
&\leq \delta(\ln\widetilde{X}_{(M_n+1)} - \ln F^{-1}(1-B_n))_+^2.
\end{aligned}$$

Odatle za neko $\varepsilon > 0$ sledi da je

$$\begin{aligned}
P\{|C_{S_n}| > \varepsilon\} &\leq P\{(\ln\widetilde{X}_{(M_n+1)} - \ln F^{-1}(1-B_n))_+^2 > \delta^{-1}\varepsilon\} \\
&\leq P\{(\ln\widetilde{X}_{(M_n)} - \ln F^{-1}(1-B_n))_+ > \sqrt{\delta^{-1}\varepsilon}\}.
\end{aligned}$$

Koristeći ponovo (5.6) za $\rho = 1$, dobijamo da $C_{S_n} \xrightarrow{P} 0$. ■

Dokaz Teoreme 5.7. Koristeći Cramér-Wold aparat, prateći dokaz iz Hila (Hill (2010), dokaz *Teoreme 2.*) i uzimajući u obzir činjenicu da, pod navedenim uslovima teoreme važi da

$$\ln \widetilde{X}_{([\rho M_n])} - \ln F^{-1}(1-\rho B_n) \xrightarrow{P} 0,$$

(videti dokaz *Teoreme 5.6.*) zaključujemo da:

$$\sqrt{M_n}(H^+_{S_n} - EH^+_{S_n} - \alpha^{-1}(\widetilde{S}^{(\zeta)}_n - E\widetilde{S}^{(\zeta)}_n))/\sigma^*_{M_n} \xrightarrow{d} N(0,1),$$

gde je $\sigma^*_{M_n} = E(I\{S_n \geq \beta_n\}\frac{1}{M_n^{1/2}}[\sum_{t=1}^{S_n}\{\widetilde{Y}_t - E(\widetilde{Y}_t) - \alpha^{-1}(\widetilde{Y}^{\zeta}_t - E(\widetilde{Y}^{\zeta}_t))\}])^2$.

Sledeći zaključak direktno sledi iz Hsinga (Hsing (1991), *Teorema 2.4*):

$$M_n^{1/2}(H_{S_n} - \alpha^{-1})/\sigma^*_{M_n} \xrightarrow{d} N(0,1).$$

Pošto je $\sigma^2_{M_n} = E(M_n^{1/2}(H_{S_n} - \alpha^{-1}))^2$, sledi da $|\sigma^*_{M_n} - \sigma_{M_n}| \xrightarrow{p} 0$. ■

Poglavlje 6

ZAKLJUČAK

Dobro je poznato da je u slučaju identično raspodeljenih medjusobno nezavisnih slučajnih promenljivih Hilova ocena indeksa debljine repa raspodele α^{-1} konzistentna i asimptotski normalna. Svrha ove teze bila je da se pokaže da se slična svojstva ocene zadržavaju i uz značajno slabljenje početnih uslova, konkretno, uz uslove nekompletnosti i ekstremalne zavisnosti. Takodje, sem ove ocene posmatrana je još jedna ocena parametra α, ocena geometrijskog tipa, kod koje je pod prethodno navedenim pretpostavkama pokazana konzistentnost.

Dokazi teze su intenzivno bazirani na radovima: Hill (2010), Breito i Freitas (2003) i Mladenović i Piterbarg (2008). Hill (2010) zasniva svoj rad na pojmu ekstremalne zavisnosti i značajno proširuje rezultate iz Hsingovog rada (Hsing (1991)) u kome je dokazana konzistentnost i asimptotska normalnost Hilove ocene jako mešajućih procesa pod veoma složenim pretpostavkama. Hill (2010) pokazuje da slični rezultati važe pod znatno pojednostavljenim uslovima koji su podesniji za praktičnu primenu. U radu Breito i Freitas (2003) analiziraju se svojstva ocene geometrijskog tipa, koju smo posmatrali kroz prizmu izmenjenih početnih uslova. Dokazali smo da se osobine te ocene ne menjaju ukoliko se oslabe i pojednostave početni uslovi. U radu Mladenović i Piterbarg (2008) definisan je model nepotpunog uzorka koji nam je poslužio kao osnova dokaza, zbog svoje jednostavne i jasne strukture.

Centralna tema teze je ocenjivanje indeksa pravilne promenljivosti α koji figuriše u izrazu funkcije raspodele teškog repa. Ocenjivanje se vrši pod pretpostavkom da nam nisu dostupni svi podaci i formira se model nekompletnog uzorka. Takodje, zahteva se da važi *ekstremalna zavisnost* medju elementima uzorka koja predstavlja tip zavisnosti u domenu visokih kvantila. Kritična pitanja o kojima smo govorili u uvodnom poglavlju a koja se odnose na verovatnoću dogadjanja kraha na berzi, ekonomske krize ili bilo kakvog kolapsa u različitim domenima, podrazumevaju poznavanje očekivanog ponašanja na krajevima funkcije raspodele u statističkom smislu. Odredjivanjem što približnije vrednosti parmetra α i pretpostavljanjem da važe gore navedene pretpostavke (nekompletnost i ekstremalna zavisnost) koje su u praktičnim situacijama veoma realne, može se simulirati funkcija raspodele teškog repa i u velikoj meri poboljšati prognoziranje teško predvidivih i nestabilnih procesa u ekonomiji, osiguranju, finansijamna, hidrologiji i drugim oblastima od interesa. Takodje, različiti slučajevi finansijskih turbulencija u novije vreme dokazuju da se dogadjaji za koje se smatralo da se mogu pojaviti "jednom u hiljadu godina" ipak pojavljuju mnogo frekventnije. Ova činjenica pokazuje aktuelnost i važnost odredjivanja indeksa debljine repa raspodele u cilju procene rizika i posmatranje različitih ocena pod izmenjenim početnim uslovima.

Postoji još jedna važna komponenta ove teze, a to je praktična primenljivost dobijenih rezultata. Početni uslovi se lako proveravaju, s obzirom da je za njihovu verifikaciju potrebno samo izračunavanje uslovnog očekivanja. Pod pretpostavkom stroge stacionarnosti, početni uslovi teorema se znatno usložnjavaju. Takodje, postoji veliki broj procesa za koje se može pokazati da zadovoljavaju zahteve teorema a koji se sa velikom frekvencijom pojavljuju u oblastima od interesa kao što su finansije i ekonomija, dakle u uzorcima sa empirijskim raspodelama teških repova. Neki od primera tih procesa su detaljno objašnjeni i analizirani u petom poglavlju, ali treba još napomenuti da je spektar procesa koji zadovoljavaju početne uslove zavisnosti veoma širok i obuhvata, pored pomenutih primera i: konačno zavisne procese, jako mešajuće GARCH procese, procese koji zadovoljavaju Hsingove uslove mešanja, ARFIMA procese i eksplozivne GARCH procese (pogledati prethodno poglavlje za primere procesa na kojima je pokazano da zadovoljavaju početne uslove dokazanih teorema, a koji su veoma česti u praktičnim aplikacijama).

S obzirom na sve veći interes za što boljom procenom rizika, nameću se i pitanja asimptotskog ponašanja i nekih drugih ocena parametra oblika repa raspodele pod istim početnim premisama. Takodje, prirodno se postavlja pitanje pod kojim uslovima važi asimptotska normalnost i stroga konzistentnost ocene geometrijskog tipa, kao i jaka konzistentnost Hilove ocene u slučaju nekompletnih podataka koji zadovoljavaju odredjene uslove zavisnosti.

Literatura

Anderson J. A. (1984): Regression and Ordered Categorical Variables. *Journal of the Royal Statistical Society.* Vol. 46, No. 1, p. 1-30

Beirlant J., Vynckier P., Teugels J. L. (1996): Tail Index Estimation, Pareto Quantile Plots, and Regression Diagnostics. *Journal of the American Statistical Association*, Vol. 91, No. 436, p. 1659-1667

Bacro, J. N., Brito, M. (1995): Weak limiting behavior of a simple tail Pareto-index estimator. *J. Statist. Plann. Inference*, Vol. 45, p. 7-19.

Beirlant J., Dierckx G., Goegebeur Y., Matthys G. (1999): Tail Index Estimation and an Exponential Regression Model. *Extremes*, Vol. 2, No. 2, p. 177-200

Brito M., Freitas A. C. M. (2003): Limiting behavior of a geometric-type estimator for tail Indices. *Insurance Math. Econom.*, Vol. 33, p. 211-226

Blattberg R. C., Nicholas J. G. (1974): A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices. *The Journal of Business,* Vol. 47, No. 2, p. 244-280

Cshorgo S., Viharos L. (1998): Estimating the tail index. Asymptotic methods in probability and statistics. *North Holland Amsterdam*, p. 833-881

Cshorgo S., Deheuvels P., Mason D. M. (1985): Kernel estimates of the tail index of a distribution. *Ann. Statist.*, Vol. 13, p. 1050-1077

Chorgo S., Mason, D. M. (1985): Central limit theorems for sums of extreme values. *Mathematical Proceedings of the Cambridge Philosophical Society*, p. 547-558

Davis R. A. and Resnick S. T. (1984): Tail estimates motivated by extreme value theory. *Ann. Statist.* Vol. 12, p. 1467-1487.

Davidson, J. (1994): Stochastic Limit Theory. *Oxford University Press.*

Davidson, J. (2004): Moment and memory properties of linear conditional heteroscedasticity models, and a new model. *Journal of Business and Economics Statistics,* Vol. 22, p. 16-29

Danielsson J., de Vries C. G. (1997): Beyond the Sample: Extreme Quantile and Probability Estimation. *Journal of Empirical Finance*, Vol. 4, p. 241–257

Drees H., Kaufmann E., (1998): Selection of the optimal sample fraction in univariate extreme value estimation. *Stochast. Process. Astrl.*, Vol. 75, p. 149–172

Danielson J., de Haan L., Peng L., de Vries C. G., (1996): Using a bootstrap method to choose the sample fraction in tail index estimation. *Preprint, Econometric Institute, Erasmus University Rotterdam, Rotterdam, Netherlands*.

Dacorogna M. M., Muller U. A., Nagler R. J., Olsen R. B., Pictet O. V., (1993): A geographical model for the daily and weekly seasonal volatility in the FX market. *Journal of International Money and Finance*, p. 413–438

De Haan, L. (1971): A form of regular variationand its astrlication to the domain of attraction of the double exponential distribution. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* Vol. 17, p. 241-258

De Haan L., Peng L. (1998): Comparison of tail index estimators. *Statistica Neerlandica*, Vol. 52, p. 60–70

De Haan L. (1981): Estimation of the Minimum of a Function Using Order Statistics. *Journal of the American Statistical Association* Vol. 76, p. 467-469

De Haan L., Resnick S. I. (1980): A Simple Asymptotic Estimate for the Index of a Stable Distribution. *Journal of the Royal Statistical Society.* Vol. 42, p. 83-87

Deheuvels P., Haeusler E., Mason D. M. (1988): Almost sure convergence of the Hill estimator. *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 104, p. 371-381

De Meyer, A., Teugls, J. L (1983): Limit theorems for Pareto-type distributions. *Trans. S. Banach Int. Inst. Warsaw*

Dempster A.P, Laird N. M, Rubin D. B. (1977): Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Stat. Soc.*, p. 39-138

Dekkers A. L. M., De Haan L. (1993): Optimal Choice of Sample Fraction in Extreme-Value Estimation. *Journal of Multivariate Analysis*, Vol. 47, p. 173–195

Dekkers A. L. M., De Haan L. (1989) : On the Estimation of the Extreme-Value Index and Large Quantile Estimation. *The Annals of Statistics,* Vol. 17, p. 1795-1832

Dekkers A. L. M., Einmahl J.H.J. and de Haan, L. (1989): A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* 17, p. 1833-1855

Dekkers M. J. (1989): Magnetic Properties of Natural Goethite. *Trim Behavior During Thermal and Alternating Field Demagnetization and Low-Temperature Treatment,* Vol. 97, p. 341–355.

Drees H. (1995): Refined Pickands estimators of the extreme value index. *The Annals of Statistics,* Vol. 3, p. 2059-2080

Du Mouchel W. H. (1983): Estimating the Stable Index α in Order to Measure Tail Thickness. *Ann. Statist.* Vol. 11, p. 1019-1031

Dunsmuir, W., Robinson, P. M. (1981): Asymptotic theory for time series containing and amplitude modulated observations. *The Indian Journal of Statistics*, Vol. 43, p. 260-281

Embrechts P., Klüstrelberg C., Mikosch T. (1997): Modeling Extremal Events for Insurance and Finance. *Springer*

Fama E. F., Roll R. (1968): Some Properties of Symmetric Stable Distributions. *Journal of the American Statistical Association,* Vol. 63, No. 323, p. 817-836

Feuerverger A., Hall P. (1999): Estimating a tail exponent by modeling departure from a Pareto distribution. *Ann. Statist.* Vol. 27, No. 2, p. 760-781

Fisher R. A., Tistrett L. H. C. (1928): Limiting forms of the frequency distribution of the largest or smallest member of a sample. Mathematical Proceedings of the Cambridge Philosophical Society, Vol. 24, p. 180-190

Gawronski W., Stadtmüller U. (1984): On the zeros of Jonquière's function with a large complex parameter. *Source: Michigan Math. J.* Vol. 31, p. 275-293

Gnedenko B. (1943): Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *The Annals of Mathematics,* Vol. 44, No. 3, p. 423-453

Graham J. W, (2009): Missing data analysis: Making it work in the real world. *Annu. Rev. Psychol*. p. 549-576

Graham J. W, Hofer S. M. (2000): Multiple imputation in multivariate research. In Modeling Longitudinal and Multiple-Group Data. *Practical Issues, Astrlied Astrroaches, and Specific Examples*, *Hillsdale, Erlbaum*.

Graham J. W, Cumsille P. E, Elek-Fisk E. (2003): Methods for handling missing data. *In Research Methods in Psychology, ed. JA Schinka, WF Velicer*, p. 87-114

Goldie C. M, Smith R. L (1987): Slow Variation With Remainder. *Theory and Astrlications*, *p.* 45-71

Guillou A., Hall P. (2001): A diagnostic for selecting the threshold in extreme value analysis. *Journal of the Royal Statistical Society*, p. 293–305

Haeusler E. and Teugels J. L. (1984): On Asymptotic Normality of Hill's Estimator for the Exponent of Regular Variation. *Ann. Statist.,* Vol. 13, No. 2, p. 743-756

Hall P. (1982): On Some Simple Estimates of an Exponent of Regular Variation. *Journal of the Royal Statistical Society. Series B (Methodological),* Vol. 44, No. 1 p. 37-42

Hall P., Welsh A. H. (1984): Best Attainable Rates of Convergence for Estimates of Parameters of Regular Variation *the Annals of Statistics,* Vol. 12, No. 3, p. 1079-1084

Hill B. M. (1975): A simple general approach to inference about the tail of a distribution. *Ann. Statist*. Vol. 3, p. 1163-1174.

Hill J. B. (2005): On tail index estimation using dependent, heterogeneous data. *Working paper*.

Hill J. B. (2010): On tail index estimation for dependent, heterogeneous data. *Econometric Theory,* Vol. 26, p. 1398-1436

Hill, J. B. (2011): Tail and nontail memory with the applications to extreme value and robust statistics. *Econometric Theory*, Vol. 27, str. 844-884

Hsing T. (1991): On tail index estimation using dependent data. *Ann. Statist*. Vol. 19, p. 1547-1569

Huisman R., Koedijk K. G., Kool C. J. M, Palm F. (2001): Tail-Index Estimates in Small Samples. *Journal of Business & Economic Statistics,* Vol. 19, p. 208-216

Ibrahim J. G. (1990): Incomplete Data in Generalized Linear Models. *Journal of the American Statistical Association,* Vol. 85, No. 411, p. 765-769

Ilić I. (2012): On tail index estimation using a sample with missing observations. *Statistics & Probability Letters*, Vol. 82, p. 949–958

Jondeau E., Rockinger M., (1998): Reading the smile: the message conveyed by methods which infer risk neutral density. *CEPR Discussion Paper*, No. 2009.

Kline P., Santos A., (2010): Sensitivity to Missing Data Assumptions: Theory and An Evaluation of the U.S. Wage Structure*. 2010 Seoul Summer Economics Conference.*

Koopman L., Geert J. M. G. van der Heiden, Diederick E. Grobbee, Maroeska M. Rovers (2007): Comparison of Methods of Handling Missing Data in Individual Patient Data Meta-analyses: An Empirical Example on Antibiotics in Children with Acute Otitis Media. *American Journal of Epidemiology*, Vol. 167, No. 5

Koedijk K. G., Schafgans M. M., de Vries C. G. (1990): The Tail Index Of Exchange Rates Returns. *Journal of International Economics*.

Leadbetter M. R., Rootzen H. (1982): Extreme value theory for continuous parameter stationary processes. *Probability Theory and Related Fields,* Vol. 60, No 1, p. 1-20

Leadbetter M. R., Rootzen H. (1988): Extremal Theory for Stochastic Processes *The Annals of Probability,* Vol. 16, No. 2, p. 431-478

Leadbetter M. R., Lindgren G., Rootzen H. (1983): Extremes and related properties of random sequences and processes. *Springer-Verlang New York*,

Ling, S. and Peng, L. (2004) Hill's estimator for the tail index of an ARMA model. *Journal of Statistical Planning and Inference,* Vol. 123, p. 279-293

Little R. J. A, Rubin D.B. (1987): Statistical Analysis with Missing Data. *New York:Wiley*

Little R. J. A, Rubin D.B. (2002): Statistical Analysis with Missing Data. *New York:Wiley. 2nd ed.*

Little R. J. A. (1995): Modeling the Drop-Out Mechanism in Repeated-Measures Studies. *Journal of the American Statistical Association,* Vol. 90, No. 431, p. 1112-1121

Luenberger D. G. (1998): Investment Science. *Oxford University Press: New York*

Lux T. (2000): On moment condition failure in German stock returns: an astrlication of recent advances in extreme value statistics. *Empirical economics, Springer.*

Mason D. M. (1982): Laws of Large Numbers for Sums of Extreme Values. *The Annals of Probability,* Vol. 10, No. 3, p. 754-764

Mladenović, Z., Petrovic, P. (2010): Cagan's paradox and money demand in hyperinflation: Revisited at daily frequency. *Journal of International Money and Finance,* No. 29, 1369-1384

Mladenović P., Piterbarg V. (2008): On estimation of the exponent of regular variation using a sample with missing observations. *Statist. Prob. Letters*. Vol. 78, p. 327-335

Mc Neil A. J. (2000): Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, p. 271–300

Novak S.Y. (2002): Inference on heavy tails from dependent data. *Siberian Adv. Math.*, p. 73–96

Pickands J. (1975): Statistical inference using extreme order statistics. *Ann. Statist.*, Vol.3, p. 119-131

Prevosti F.J., Chemisquy M.A. (2009): The impact of missing data on real morphological phylogenies: influence of the number and distribution of missing entries. *Cladistics,* No. 26, p. 326-339

Resnick, S.I. (1997): Discussion of the Danish Data on Large Fire Insurance Losses. *ASTIN Bulletin,* Vol. 27, p. 139–51

Resnick S. and Starica, C. (1995): Consistency of Hill's estimator for dependent data. *J. Astrl. Probab*. Vol. 32, p. 139-167

Resnick S. and Starica, C. (1997): Asymptotic behavior of Hill's estimator for autoregressive data. *Stochastic Models* Vol. 13, p. 703-723

Resnick S. and Starica, C. (1998): Tail index estimation for dependent data. *Ann. Astrl. Probab.* Vol. 8, p. 1156-1183

Robins M., Rotnitzky A., Zhao L. P. (1994): Estimation of Regression Coefficients When Some Regressors Are Not Always Observed. *Journal of the American Statistical Association,* Vol. 89, No. 427, p. 846-866

Rubin D. B. (1987): The Calculation of Posterior Distributions by Data Augmentation: Comment: A Noniterative Sampling/Importance Resampling Alternative to the Data Augmentation Algorithm for Creating a Few Imputations When Fractions of Missing Information Are Modest: The SIR Algorithm. *Journal of the American Statistical Association,* Vol. 82, No. 398, p. 543-546

Rubin D. B. (1976): Inference and missing data. *Biometrika*, Vol. 63, p. 581-592

Rubinstein M. (1994): Implied Binomial Trees. *The Journal of Finance,* Vol. 49, No. 3, p. 771-818

Schafer J. L. (1997): Analysis of Incomplete Multivariate Data. *New York: Chapman and Hall.*

Schafer J. L. (1999): Multiple Imputations: a primer. *Stat. Methods Med. Res*. p. 8-315

Schafer J. L., Olsen M.K. (1998): Multiple Imputations for multivariate missing data problems: a data analyst's perspective. *Multivar. Behav. Res.* p. 33-54

Schafer J. L., Graham J.W. (2002): Missing data: our view of the state of the art. *Psychol. Methods*, p. 7-14

Schultze, J., Steinebach, J. (1996): On least squares estimates of an exponential tail coefficient. *Statist. Decisions* Vol. 14, p. 353-372

Smith R.L., Weissman I. (1985): Maximum Likelihood Estimation of the Lower Tail of a Probability Distribution. *Journal of the Royal Statistical Society.* Vol. 47, No. 2, p. 285-298

Smith R. L., Weissman I. (1987): Large Deviations of Tail Estimators Based on the Pareto Astrroximation. *Journal of Astride Probability,* Vol. 24, No. 3, p. 619-630

Taleb N. N. (2007): Black Swans and the Domains of Statistics. *The American Statistician*, Vol. 61.

Teugels J. L. (1981a): Limit Theorems on Order Statistics. *The Annals of Probability,* Vol. 9, No. 5, p. 868-880

Teugels J. L. (1981b): Remarks on large claims. *Bull. Inst. Internat. Statist.,* Vol. 49, p. 1490-1500

Tjostheim D. (1990): Non-linear time series and Markov chains. *Adv. Astrl. Prob.* Vol. 22, p. 587–611

Verbeke G. and Molenberghs G. (2003). The use of score tests for inference on variance components. *Biometrics,* Vol. 59, p. 254-262

Wang Q. and Luo R. (2011): Semi-empirical pseudo likelihood for estimating equa-tions in the presence of missing responses. *Journal of statistical planning and inference,* http://dx.doi.org/10.1016/j.jspi.2011.02.009.

Biografija autora

Ivana Ilić rođena je 14. 07. 1975. godine u Beogradu. Osnovnu školu „Sveti Sava“ i gimnaziju „Bora Stanković“ u Nišu, završila je sa odličnim uspehom kao nosilac diplome „Vuk Karadžić“. U toku školovanja učestvovala je na republičkim takmičenjima u oblasti matematike, srpskog i ruskog jezika. Učestvovala je u raznim aktivnostima iz oblasti matematike u istraživačkoj stanici Petnica.

Studije na Prirodno-matematičkom fakultetu u Nišu, na odseku za matematiku i smeru za teorijsku matematiku i primene, upisala je školske 1994/1995. godine, a završila je 2001. godine sa prosečnom ocenom 9,78.

Školske 2001/2002. godine upisala je poslediplomske studije na Prirodno-matematičkom fakultetu u Beogradu, na odseku za verovatnoću i statistiku.

Magistarsku tezu, pod nazivom ,,Ocenjivanje indeksa pravilne promenljivosti korišćenjem ekstremnih statistika poretka“ odbranila je 2006. godine, i time stekla akademski naziv Magistar matematike.

Školske 2003/2004, radi kao asistent pripravnik na predmetu matematika na Elektronskom fakultetu u Nišu. Od 2004. do 2008. godine, radi kao asistent pripravnik na Medicinskom fakultetu na predmetu matematika na odseku farmacija, i na predmetu medicinska informatika na odseku medicina, stomatologija, farmacija i medicinska nega. Godine 2008, stekla je sertifikat iz oblasti Pedagoško-metodičkog usavršavanja fakultetskih saradnika u organizaciji centra za praćenje, obezbeđivanje, unapređivanje i razvoj kvaliteta nastave i naučno-istraživačkog rada Medicinskog fakulteta u Nišu. Godine 2008. birana je u zvanje asistenta na istom fakultetu i ponovo birana u zvanje asistenta godine 2011.

Radila je kao instruktor informatike u školi za talente ,,Energea“.

Bila je saradnik na dva projekta ministarstva za nauku: na Prirodno-matematičkom fakultetu u Nišu (rukovodilac projekta Prof. dr Vladimir Rakočević) i na Prirodno-matematičkom fakultetu u Beogradu (rukovodilac projekta Prof. dr Pavle Mladenović). Takođe, bila je učesnik je na projektu *Uloga prosvetnih radnika u opismenjavanju učenika* (akreditovan program za usavršavanje učitelja i nastavnika u katalogu 2008/2009).

Član je naučne asocijacije specijalizovane za verovatnoću i statistiku Bernoulli Society.

Poznavalac je sledećih kompjuterskih programa: MS Word, MS Exel, MS Power Point, MS Access, Tex, SPSS, STATISTICA, Corel Draw, Adobe Photoshop, Xtremes, itd.

Govori engleski, francuski i ruski jezik. Bavi se pevanjem i plesom.

Udata je i ima dve kćerke, Teodoru i Nađu.

Objavila je do sada osam radova računajući magistarsku tezu, i to:

1) *ON SOME ASPECTS OF USING MATHEMATICS IN MEDICAL SCIENCES* Ivana Ilić, Acta Medica Medianae 2008, Vol.47.

2) *NEW ESTIMATION PROCEDURE OF THE SHAPE PARAMETER USING THE EXTREMES*, Ivana Ilić, Pavle Mladenović, International Statistical Institute (ISI) (2007) (Book of abstracts).

3) *56th Congress of the ISI,* Ivana Ilić, Statistical Review 2007, number 3-4.

4) *INCOMPLETE SAMPLES AND TAIL ESTIMATION FOR STATIONARY SEQUENCES*, Ivana Ilić, Pavle Mladenović, Novi Sad J. Math.Vol. 38, No. 3, 2008, str. 97-104.

5) *META ANALYSIS*, Ivana Ilić, Acta Medica Medianae 2009, Vol. 48.

6) *WEAK CONVERGENCE OF PRODUCT OF SUMS OF INDEPENDENT VARIABLES WITH MISSING VALUES*, Ivana Ilić, Filomat (2010), vol. 24 br. 3, str. 73-81.

7) *ON TAIL INDEX ESTIMATION USING A SAMPLE WITH MISSING OBSERVATIONS*, Ivana Ilić, Statistics and Probability Letters (2012), Vol. 82, str. 949–958.

8) *OCENJIVANJE INDEKSA PRAVILNE PROMENLJIVOSTI KORIŠĆENJEM EKSTREMNIH STATISTIKA PORETKA*, Ivana Ilić, Magistarska teza (2006).

Прилог 1.

Изјава о ауторству

Потписани-а Ивана Илић

број уписа

Изјављујем

да је докторска дисертација под насловом

Оцењивање индекса репа расподеле коришћењем неколитетних узорака

- резултат сопственог истраживачког рада,
- да предложена дисертација у целини ни у деловима није била предложена за добијање било које дипломе према студијским програмима других високошколских установа,
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио интелектуалну својину других лица.

Потпис докторанда

У Београду, јун, 2012

Ивана Илић

Прилог 2.

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора Ивана Илић

Број уписа

Студијски програм

Наслов рада Оцењивање индекса репа расподеле коришћењем некомплетних узорака

Ментор Проф. др Павле Младеновић

Потписани

изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла за објављивање на порталу **Дигиталног репозиторијума Универзитета у Београду.**

Дозвољавам да се објаве моји лични подаци везани за добијање академског звања доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис докторанда

У Београду, јун, 2012

Прилог 3.

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

Оцењивање индекса репа расподеле коришћењем некомплетних узорака

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигитални репозиторијум Универзитета у Београду могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство
2. Ауторство - некомерцијално
3. Ауторство – некомерцијално – без прераде
4. Ауторство – некомерцијално – делити под истим условима
5. Ауторство – без прераде
6. Ауторство – делити под истим условима

(Молимо да заокружите само једну од шест понуђених лиценци, кратак опис лиценци дат је на полеђини листа).

Потпис докторанда

У Београду, јун, 2012

1. Ауторство - Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.

2. Ауторство – некомерцијално. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.

3. Ауторство - некомерцијално – без прераде. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.

4. Ауторство - некомерцијално – делити под истим условима. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.

5. Ауторство – без прераде. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.

6. Ауторство - делити под истим условима. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.