

MISKOLCI EGYETEM  
GÉPÉSZMÉRNÖKI ÉS INFORMATIKAI KAR



# **Szimmetrikus stabil eloszlások paramétereinek egy robusztus becslési eljárása és alkalmazása**

Doktori (PhD) értekezés

*Készítette:*

**Csendes Csilla**

okleveles közgazdasági programozó matematikus

HATVANY JÓZSEF INFORMATIKAI TUDOMÁNYOK  
DOKTORI ISKOLA

*Tudományos témavezető:*

Dr. Fegyverneki Sándor

*A doktori iskola vezetője:*

Prof. Dr. Szigeti Jenő

A matematikai tudomány kandidátusa

**Miskolc  
2014**

## Köszönetnyilvánítás

Szeretnék köszönetet mondani tudományos vezetőmnek, Dr. Fegyverneki Sándornak, akinek szakmai irányítása alatt 2008-tól fogva először a Kockázatkezelés statisztikai módszerei című szabadon választható egyetemi kurzus, majd tudományos diákköri munka és a diplomamunka készítés keretében dolgoztam a doktori értekezésem tématerületén. A PhD képzésre témavezetőm hívta fel a figyelmemet és az ő bátorítására fogtam bele az egyetemi oklevél megszerzése után a kutatómunkába. Köszönöm szakmai segítségét, a publikációk elkészítésénél adott hasznos tanácsait, és a kutatási tervet illető határozottságát, mellyel az eredményre vezető úton tartott.

Szeretném megköszönni a Miskolci Egyetem Alkalmazott Matematikai Tanszékén és Analízis Tanszékén dolgozó valamennyi kolléga segítőkészségét és útmutatását, mellyel a doktoranduszi éveim alatt támogatták szakmai fejlődésem.

Hálás vagyok a Budapesti Corvinus Egyetem Biometria és Agrárinformatika Tanszékén dolgozó volt kollégáimnak, akikhez bármikor fordulhattam kérdéseimmel és problémáimmal. Köszönöm a támogatásukat, és a rengeteg biztatást, amelyet tőlük kaptam. Szeretném külön is megköszönni Dr. Ladányi Mártának, hogy kikezdhetetlen életszeretetével és optimizmusával mindig új lendületet tudott adni a nehézségek leküzdésére.

Köszönöm páromnak, Bélteky Attilának a végtelen türelmét és a mindennapokhoz adott erőt, amely nélkül a doktori értekezésem nem készülhetett volna el. Köszönöm, hogy mellettem állt, és támaszkodhattam a segítségére.

Köszönöm Nagymamámnak a gondoskodását, és hogy egy csöndes, nyugodt hely biztosításával lehetővé tette számomra a vizsgákra való zavartalan felkészüléseket és az elmélyedt munkát. Szüleimnek köszönöm az értem hozott áldozataikat és a biztatásukat.

Végül, de nem utolsó sorban köszönöm barátaimnak a kikapcsolódás vidám perceit.

# Tartalomjegyzék

<b>1. Bevezetés</b>	<b>1</b>
<b>2. A stabil eloszlások</b>	<b>6</b>
2.1. Egyváltozós stabil eloszlások . . . . .	6
2.2. Többváltozós stabil eloszlások . . . . .	10
<b>3. Portfólió modellek, kockázatkezelés</b>	<b>18</b>
<b>4. A PIT paraméterbecslési eljárás</b>	<b>26</b>
4.1. A stabil paraméterek becslése . . . . .	26
4.2. A robusztus statisztika . . . . .	28
4.3. A PIT paraméterbecslési eljárás ismert eloszlástípus esetén . . . . .	36
4.4. PIT paraméterbecslés nem ismert eloszlástípus esetén . . . . .	39
4.5. A becslések kiszámításának algoritmusai . . . . .	43
4.6. A $\mathcal{B}$ függvények közelítése . . . . .	48
<b>5. Statisztikai vizsgálatok</b>	<b>57</b>
5.1. Stabil eloszlású véletlen számok generálása . . . . .	57
5.2. A PIT paraméterbecslési módszer statisztikai vizsgálata . . . . .	64
5.3. A becslések normalitásának vizsgálata . . . . .	69
5.4. Becslési eljárások összehasonlítása . . . . .	82
<b>6. A BÉT részvényeinek modellezése a PIT módszerrel</b>	<b>86</b>
6.1. A hozambecslés modelljei . . . . .	86
6.2. A BÉT hozamainak PIT becsléssel számított paraméterei . . . . .	89
6.3. A hozamok eloszlásának vizsgálata mozgóablakok segítségével . . . . .	94
<b>7. A PIT becslés implementációja és a segédprogramok</b>	<b>97</b>
7.1. A PIT paraméterbecslést kiszámító MATLAB függvények . . . . .	97
7.2. A törtefüggvény közelítés meghatározásának segédfüggvényei . . . . .	99
7.3. Függvények véletlenszámok generálásához . . . . .	99
7.4. A valós árfolyamok elemzésére készített MATLAB függvények . . . . .	101
7.5. A statisztikai vizsgálatokhoz, szimulációkhoz készített MATLAB függvények . . . . .	101
<b>8. Összefoglalás, tézisek</b>	<b>104</b>
8.1. Összefoglalás és javasolt kutatási irányok . . . . .	104
8.2. Tézisek . . . . .	106
<b>9 Summary</b>	<b>107</b>
9.1 Introduction . . . . .	107
9.2 Main results . . . . .	109
<b>10 Függelék</b>	<b>110</b>

## Ábrák jegyzéke

1.	Az értékpapír-piaci egyenes. . . . .	23
2.	A hatékony portfóliók halmaza kockázatmentes befektetés lehetősége esetén. . . . .	24
3.	A $\psi$ és $\chi$ súlyfüggvények, $\alpha = 1$ és $\alpha = 2$ . . . . .	40
4.	A ping-pong módszerrel meghatározott $S_1(\alpha)$ és $S_2(\alpha)$ skálaparaméter görbék $n = 50$ elemből, $\alpha = 1.2$ (felül) és $\alpha = 1.8$ (alul) esetén . . . . .	42
5.	A ping-pong módszerrel meghatározott $S_1(\alpha)$ és $S_2(\alpha)$ skálaparaméter görbék $n = 500$ elemből, $\alpha = 1.3$ (felül) és $\alpha = 1.7$ (alul) esetén . . . . .	44
6.	A ping-pong módszerrel meghatározott $S_1(\alpha)$ és $S_2(\alpha)$ skálaparaméter görbék $n = 5000$ elemből, $\alpha = 1.4$ (felül) és $\alpha = 1.6$ (alul) esetén . . . . .	46
7.	A ping-pong módszerrel meghatározott $S_1(\alpha)$ és $S_2(\alpha)$ skálaparaméter görbék $n = 500$ elemből, $\alpha = 1.5$ , $\gamma = 5$ (felül) és $\gamma = 0.5$ (alul) esetén . . . . .	47
8.	A közelítő függvényértékek szórása a $\mathcal{B}_1$ függvény esetén . . . . .	50
9.	A közelítő függvényértékek szórása a $\mathcal{B}_2$ függvény esetén . . . . .	50
10.	A $\mathcal{B}_1(\alpha)$ függvény közelített értékei . . . . .	53
11.	A $\mathcal{B}_2(\alpha)$ függvény közelített értékei . . . . .	53
12.	A $\mathcal{B}_1(\alpha)$ hibás közelítő függvénye, ( $m = 5$ , $n = 4$ ) eset . . . . .	56
13.	$\alpha = 1.5$ paraméterű stabil eloszlású három dimenziós generált minta. . . . .	61
14.	$\alpha = 1.5$ paraméterű két dimenziós gömbszimmetrikus generált minta. . . . .	62
15.	$\alpha = 1.5$ paraméterű egy dimenziós generált minta. . . . .	62
16.	$\alpha = 1.5$ paraméterű két dimenziós generált minta. . . . .	63
17.	$\alpha = 1.3$ paraméterű három dimenziós generált minta. . . . .	63
18.	Normalitásvizsgálat az $\alpha$ paraméterre (felső sor: $n = 50$ , alsó sor: $n = 100$ ) . . . . .	73
19.	Normalitásvizsgálat az $\alpha$ paraméterre (felső sor: $n = 400$ , alsó sor $n = 2500$ ) . . . . .	74
20.	A vizsgált részvények napi záróárainak alakulása 2004.01.01. és 2012. 12. 31. között . . . . .	87
21.	A vizsgált részvények logaritmikus hozamai . . . . .	88
22.	A vizsgált részvények logaritmikus hozamainak gyakorisági hisztogramjai . . . . .	88
23.	A becsült alak- és skálaparaméterek közötti kapcsolat . . . . .	89
24.	Q-Q ábrák a logaritmikus hozamok empirikus eloszlása és a normális eloszlás között . . . . .	93
25.	Q-Q ábrák a logaritmikus hozamok empirikus eloszlása és a becsült stabil eloszlás között . . . . .	93
26.	Konfidencia intervallum normális és stabil eloszlás alapján - $n = 50$ , BUX, 2004.01.01. - 2012. 12. 31. . . . .	96

## Táblázatok jegyzéke

1.	A $\mathcal{B}_1$ és $\mathcal{B}_2$ függvények értékei az $\alpha = 1, 1.05, \dots, 2$ alappontokban . . . . .	52
2.	A legmegfelelőbb törtfüggvény közelítés keresése során vizsgált fokszámok és alappontok . . . . .	54
3.	A közelítések legnagyobb abszolút eltérései a szimulált függvényértékektől . . . . .	55
4.	A $\mathcal{B}_1$ és $\mathcal{B}_2$ legmegfelelőbb racionális törtfüggvény közelítésének együtthatói . . . . .	56
5.	A PIT becslés performanciájának jellemzői, $r = 2500$ . . . . .	66
6.	MSE értékek $r=2500$ . . . . .	67
7.	Kovariancia mátrix a három becsült paraméterre - $n = 50$ . . . . .	68
8.	Kovariancia mátrix a három becsült paraméterre - $n = 100$ . . . . .	68
9.	Kovariancia mátrix a három becsült paraméterre - $n = 400$ . . . . .	68
10.	Kovariancia mátrix a három becsült paraméterre - $n = 2500$ . . . . .	68
11.	Az érvényes becslések száma a generált minták elemszáma ( $n$ ) és az $\alpha$ paraméter alapján . . . . .	68
12.	A meghatározott p-értékek és a szabadsági fokok a $\gamma$ paraméter egyváltozós normalitásának tesztelése során $\chi^2$ próbával . . . . .	70
13.	A meghatározott p-értékek és a szabadsági fokok a $\delta$ paraméter egyváltozós normalitásának tesztelése során $\chi^2$ próbával . . . . .	71
14.	A meghatározott p-értékek és a szabadsági fokok az $\alpha$ paraméter egyváltozós normalitásának tesztelése során $\chi^2$ próbával . . . . .	72
15.	A meghatározott p-értékek az $\alpha$ paraméter normalitásának vizsgálatára Kolmogorov - Smirnov próbával . . . . .	75
16.	A meghatározott p-értékek az $\alpha$ paraméter normalitásának vizsgálatára Sarkadi próbával . . . . .	77
17.	A Mardia-féle többváltozós ferdeség és lapultság teszt döntései(D), a tesztstatisztika értéke és a p-érték . . . . .	79
18.	Doornik és Hansen féle többváltozós teszt eredményei (D: döntés) . . . . .	81
19.	Becslési eljárások összehasonlítása 1. . . . .	82
20.	Becslési eljárások összehasonlítása 2. . . . .	83
21.	Becslési eljárások összehasonlítása 3. . . . .	84
22.	Becslési eljárások összehasonlítása 4. . . . .	85
23.	A PIT módszerrel a logaritmikus hozamokból számított $\hat{\alpha}$ alak-, $\hat{\delta}$ hely- és $\hat{\gamma}$ skálaparaméter becslések . . . . .	89
24.	A PIT módszerrel a százalékos hozamokból számított $\hat{\alpha}$ alak-, $\hat{\delta}$ hely- és $\hat{\gamma}$ skálaparaméter becslések . . . . .	90
25.	A STABLE programmal a logaritmikus hozamokból számított alak- $\hat{\alpha}$ , ferdeségi- $\hat{\beta}$ , hely- $\hat{\delta}$ és skála- $\hat{\gamma}$ paraméter becslések . . . . .	90
26.	A hely- és skálaparaméter robusztus és normális eloszlás szerinti becslései a logaritmikus hozamokból . . . . .	91
27.	A logaritmikus hozamok illeszkedésvizsgálatára végzett $\chi^2$ próba p-értékei, tesztstatisztika értékei és szabadsági fokai . . . . .	91
28.	A logaritmikus hozamok illeszkedésvizsgálatára végzett Kolmogorov-Smirnov próba p-értékei és tesztstatisztika értékei . . . . .	92
29.	A PIT algoritmus implementációjának függvényei . . . . .	98
30.	A törtfüggvény közelítés meghatározásának segédfüggvényei . . . . .	99
31.	Függvények véletlenszám generáláshoz (.m fájlok) . . . . .	100
32.	A valós árfolyamok elemzésére készített MATLAB függvények . . . . .	102
33.	A PIT becslés normalitásvizsgálatának néhány MATLAB függvénye . . . . .	103

## Jelölések listája

A dolgozatban használt jelölések.

- $\stackrel{D}{=}$  : eloszlásban való egyezés (egyenlőség)
- $a^T$  :  $T$  transzponálás
- $\langle u, v \rangle$ : az  $u$  és  $v$  vektorok skaláris szorzata, belső szorzat
- $f'$ : az  $f$  függvény deriváltja
- $*$ : konvolúció operátor
- $\|x\|$ : az  $x$  vektor normálja

## 1. Bevezetés

A pénzügyi modellezés központi kérdései a portfólió optimalizálás, a kockázatkezelés, valamint a különböző pénzügyi termékek (derivatívák) árazása. A portfólió kiválasztás alapmodellje (Markowitz [48], 1952) szerint a portfólió hozama a portfólióban szereplő eszközök hozamainak súlyozott összege, a portfólió kockázata pedig az eszközök varianciájával mérhető. Az optimális portfólió kialakításához szükséges a hozamok eloszlásának, valamint az eszközök hozamai közötti összefüggési struktúrájának az ismerete. A gyakorlatban ez azt jelenti, hogy a hozamok eloszlását a rendelkezésre álló historikus adatokból becsüljük, és ezek alapján megpróbáljuk a hozamokat modellezni, valamint előrejelezni. A kockázatkezelésben valamely kockázatos eszközbe történő befektetés és az eszköz tartásának kockázatát mérjük, és általában csak az extrém veszteségek előfordulásának valószínűségére fókuszálunk. Ekkor a hozamok eloszlásának teljes ismerete nem szükséges, a hozameloszlás farkának, azaz a szélső kvantiliseinek minél pontosabb becslése viszont annál inkább.

A pénzügyi modellezési problémákban tehát a parametrikus modellek használatához elengedhetetlen, hogy a hozamok eloszlását megfelelő pontossággal, könnyen és egyszerűen használható statisztikai eszközökkel becsülhessük. A Markowitz modell a hozamok logaritmusainak eltérését vizsgálja, amelyek eloszlását normális eloszlásúnak tételezi fel. A logaritmikus hozamok eloszlásának alakulását sok, egymástól független, véges szórású tényező együttes hatásainak tulajdonítja. Ekkor a központi határeloszlás tétel értelmében a normális eloszlás adódik határeloszlásként. A normális eloszlás használata mellett szól, hogy az eloszlás összes jellemzője jól ismert, rengeteg statisztikai módszer került kidolgozásra a normális eloszláshoz, és az összes statisztikai szoftvercsomag tartalmaz rutinokat az eloszlás kezelésére. Az empirikus tanulmányok azonban azt mutatják, hogy a hozamok eloszlásának farka vastagabb, azaz nagyobb valószínűséggel következnek be extrém kilengések az árfolyamokban, mint azt a normális eloszlás alapján várnánk, valamint a hozameloszlás csúcsosabb is.

Ennek a jelenségnek a megragadására az 1960-as évektől számos vastag-farkú (heavy-tailed) eloszlást vizsgáltak meg, köztük a stabil eloszlást (Mandelbrot [45], Fama [14], [15], [16]). A kutatások alapján a stabil eloszlások megfelelően leírják az extrém esetek magasabb valószínűségű bekövetkezését, illetve a hozamok aszimmetriája (ferdesége) is modellezhető segítségükkel. A stabil eloszlások családját Levy [40] írta le független valószínűségi változók összegzéséből adódó változók határeloszlásait vizsgáló tanulmányában a XX. század elején. Az eloszláscsalád elnevezése onnan ered, hogy stabil valószínűségi változókat összegezve egy skálázó és egy centráló konstansból eltekintve ismét stabil valószínűségi változót kapunk. A stabil eloszlások természetes általánosítását adják a normális eloszlásnak (mely önmaga is a stabil eloszláscsalád tagja) az általánosított centrális határeloszlás tétel alapján, melyben elhagyva az összegzendő változók létező véges szórására vonatkozó feltételt, határeloszlásként a stabil eloszláscsalád adódik. Ez a probléma a vonzási tartomány probléma, mely részletes leírása található Gnedenko és Kolmogorov [25], Petrov[65] és Feller [21] munkáiban.

A stabil eloszláscsalád tehát fontos matematikai statisztikai szereppel rendelkezik, viszont gyakorlati alkalmazásokban nehezen használható. Bár már az 1960-as években felmerült a lehetősége az eloszláscsalád használatának, mégsem tudott széles körben elterjedni, részben az alkalmazást nehezítő tulajdonságok, részben a végtelen szórásnégyzet tulajdonsága miatt. Számos szerző azért veti el a stabil eloszlásokat, mert a gyakorlatban nehezen értelmezhető a hozamok eloszlásának végtelen szórása. A másik tényező, ami problémát okoz, az ismeretlen zárt alakú általános sűrűségfüggvény és eloszlásfüggvény. A stabil eloszlásokat a karakterisztikus

függvény segítségével lehetséges leírni, amely különböző parametrizációkban adott (Uchaikin és Zolotarev [79]). A parametrizációkból adódó eltérések korábban félreértéseket is eredményeztek. A parametrizációkban közös, hogy egy általános stabil eloszlás négy paraméter ( $\alpha$  karakterisztikus kitevő,  $\beta$  aszimmetria vagy ferdeségi,  $\gamma$  skála-, és  $\delta$  helyparaméter) segítségével írható le.

Matematikai statisztikai szempontból a nem létező szórásnégyzet, a nem létező magasabb rendű momentumok, valamint az ismeretlen zárt alakú sűrűségfüggvény azt eredményezi, hogy a paraméterbecslésre leggyakrabban használt momentumok módszere és a maximum likelihood módszer nem alkalmazható közvetlenül. A sűrűségfüggvény numerikusan közelíthető, de ennek számítási igénye nagyon jelentős, az eljárások implementálása bonyolult. A sorfejtéssel adott ún. integrál reprezentációban felírt eloszlás és sűrűségfüggvény értékeinek pontos kiszámításához a formula nagy számú tagjának kiértékelésére van szükség, azaz a formula konvergenciája nagyon lassú.

Kutatásom célja olyan numerikus statisztikai eljárás kifejlesztése volt, amelynek segítségével a stabil eloszlások paraméterbecslési feladata nagy pontossággal, számítási igényt tekintve gyorsan és egyszerűen megoldható. A stabil eloszlásból származó minták paraméterbecsléséhez a fenti tulajdonságok miatt merőben új módszerekre van szükség. Az általam bemutatott paraméterbecslési eljárás a robusztus statisztikában használt M-becslések (maximum likelihood típusú becslések) közé tartozik. Az új eljárás a hely- és skálaparaméter együttes M-becslésén alapszik. A szakirodalomban ismert számos paraméterbecslési eljárással szemben az új módszer megbízható eredményt ad, gyors, és egyszerűen implementálható.

A robusztus statisztika célja olyan eljárások létrehozása, amelyek kevésbé érzékenyek az adatok kisebb mértékű kerekítési hibáiból, vagy akár a feltételezett modell-eloszlástól való jelentős, durva hibákból eredő eltérésekre. A robusztus eljárások segítségével csökkenthető az outlierok, azaz kiugró értékek hatása a becslésekre, valamint a hatásfüggvény koncepciójával lehetővé válik, hogy egy-egy kiugró érték becslésre gyakorolt hatását mérjük. Továbbá a katasztrófa fogalmának bevezetésével meghatározható, hogy egy adott becslési eljárás mekkora mértékű hibás megfigyelés esetén tud még mindig megbízható eredményt adni. A robusztus statisztika a paraméteres és a nemparaméteres statisztikai eljárások között helyezhető el, mivel nem egy konkrét paraméteres modellben gondolkodik, hanem a modell eloszlás egy tetszőlegesen kicsi környezetét vizsgálja. A robusztus statisztikában egy eljárás jóságát különböző robusztus (V-robusztus, B-robusztus, kvalitatív robusztus) tulajdonságokkal mérik. A legfontosabb eredményeket a robusztus statisztika témakörében Huber [32] és Hampel et al. [27] foglalták össze 1981-ben, illetve 1986-ban megjelent könyveikben.

Az általam bemutatott becslési eljárás rendelkezik a V-robusztus, B-robusztus, kvalitatív robusztus tulajdonságokkal. A becslési eljárás támaszkodik bizonyos függvények ismeretére, amely függvényeket előzetesen racionális törtfüggvénnyel közelítettem. Ezeknek a függvényeknek a közelítése időigényes számításokon, nagy mintás véletlenszám generáláson alapszik, viszont elegendő volt csak egyszer elvégezni a közelítést. A közelítés használata lehetővé teszi az algoritmus gyorsítását is. A becslési eljárás a bemutatott algoritmus és közelítő függvények alapján tetszőleges programozási nyelven implementálható.

A feltételezeten stabil eloszlású hozamok paraméterbecslésével lehetővé válik a kockázatkezelésben az 1990-es évektől kezdődően előszeretettel használt kockázatot érték (Value at Risk, VaR) és várható veszteség (Expected Shortfall, ES vagy feltételes VaR, CVaR) mutatók meghatározása. Valamint a paraméterek becslésével konfidenciaintervallumot konstruálhatunk



a hozamok eloszlására, amely korlátozott előrejelzési lehetőséget is biztosít. A portfólió optimalizálási feladat megoldásában is hasznos eszköz lehet a bemutatott eljárás, valamint természetesen a pénzügyi területeken kívül is használható stabil eloszlású minták vizsgálatára.

A stabil eloszlások pénzügyi modellezésben való alkalmazása mára széles körben elfogadottá vált a gyakorlati szakemberek számára. A számítógépek növekvő számítási kapacitásai, valamint az újabb és újabb algoritmusok révén valódi alternatívát jelenthet a stabil eloszláscsalád használata a normális eloszlással szemben. A szakterület korai eredményeinek összefoglalása található Press [66] 1972-es könyvében, a témában született újabb eredményeket is összefoglaló kézikönyvek, tanulmányok a 2000-es évekből többek között Adler, Feldman és Taqqu [3], Bradley és Taqqu [6], Rachev [68], Rachev és Mittnik [69].

A dolgozatom felépítése a következő. A 2. fejezetben a stabil eloszlások alapvető jellemzőiről, statisztikai modellezésben játszott szerepéről, kedvező és kedvezőtlen tulajdonságairól szólok részletesebben. A stabil eloszlásokról szóló rész kidolgozásában nagyon hasznos könyvek és monográfiák voltak Zolotarev [84], Uchaikin és Zolotarev [79], valamint Samorodnitsky és Taqqu [71].

A 3. fejezet röviden áttekinti a legfontosabb pénzügyi, közgazdaságtani elméleteket, amelyek a modern pénzügytan alapjait képezik, és amely modelleknek mindegyikének van stabil eloszlásokat feltételező változata, kiterjesztése. A dolgozatom ezen áttekintő fejezetét részben a Stabil portfólió analízis című Tudományos Diákköri dolgozatom [S12] alapján készítettem, amelyben a hatékony portfóliók kiválasztásának modelljei szakirodalmát dolgoztam fel, és amelyért 2009-ben az Országos Tudományos Diákköri Konferencián az Informatika tudományág, Gazdaságtudományi alkalmazások szekciójában Debrecenben Különdíjat nyertem.

A 4. fejezet 4.1 szakaszában a szakirodalomban ismert paraméterbecslési eljárásokat mutatom be különös tekintettel a módszerek hiányosságaira, hibáira. Az eloszlás farkának aszimptotikus Pareto tulajdonságára épülnek a farokindex becslések (Csörgő [9], Csörgő és Viharos [10], Hall [26], Hill [31], Szeidl [77], Viharos [80]), kvantiliseken alapuló becsléseket dolgozott ki Fama és Roll [17], [18], illetve McCulloch [50], a sűrűségfüggvény numerikus Fast Fourier Transform algoritmussal történő integrálásán keresztül a Maximum likelihood módszert alkalmazza Nolan [55], a karakterisztikus függvényen alapuló becsléseket mutatott be Press [66], a karakterisztikus függvény segítségével definiált regressziós módszert mutatott be Koutrouvelis [39], majd a módszer javítását Kogon és Williams [38]. A legfrissebb tanulmányok közül Bayes statisztikai megközelítést alkalmazott Garcia et. al [24], valamint paraméterbecsléshez és előrejelzéshez vezette le stabil eloszlások törtmomentumainak formuláit Matsui és Pawlas [49].

A robusztus statisztika azon eredményeiről szól a 4.2. szakasz, amelyekre támaszkodtam az új módszer kidolgozásánál. Ebben a fejezetben kerülnek kifejtésre a hatásfüggvényen alapuló megközelítés robusztus statisztikai mérőszámai és a katasztrófa pont fogalma. Továbbá összefoglaltam az M-becslésekkel kapcsolatos tételeket, ismert eredményeket. Az M-becslésekkel kapcsolatos fontos eredmény, hogy aszimptotikusan normális eloszlásúak. A hely- és skálaparaméter együttes M-becslésének aszimptotikus eloszlásának kovariancia mátrixa szintén megadható (Hampel et al. [27], Fegyverneki [19]).

A 4.3. szakasz a hely- és skálaparaméter együttes M-becslését írja le abban az esetben, ha a modelleloszlás típusa ismert, azaz stabil eloszlások esetében az  $\alpha$  alakparamétert ismerjük. A 4.4. szakaszban részletesen bemutatásra kerül a kidolgozott új statisztikai módszer, amely a stabil  $\alpha$ ,  $\gamma$  és  $\delta$  paraméterek együttes (egyidejű) becslését teszi lehetővé. Az eljárás a Probability Integral Transformation (PIT) technikát használja. A módszer szimmetrikus ( $\beta = 0$ ) stabil

eloszlás feltételezése mellett használható. A szakaszban megadom az eljárás algoritmusát, és az eljárásra vonatkozó tulajdonságokat. Az új eredményt a saját munkák között hivatkozott Csenedes [S1] dolgozat alapján mutatom be.

A 4.6. szakasz a becslési eljárás alkalmazását lehetővé tevő numerikus közelítések meghatározásának lépéseit mutatja be. A racionális törtfüggvény közelítés meghatározásához nagy számú stabil eloszlásból származó véletlen szám generálására volt szükség. Tetszőleges paraméterű stabil eloszlású változók generálásához használható a Zolotarev [84] által bemutatott formula. A generált stabil véletlenszámok segítségével a közelítendő függvények értékeit bizonyos alappontokban meghatároztam, és az alappontokból lineáris egyenletrendszerként írtam fel. A lineáris egyenletrendszer megoldásával megkaptam a racionális törtfüggvények együtthatóit.

Az 5. fejezet összefoglalja a statisztikai vizsgálat eredményeit, amelyet az új módszer hatékonyságának, megbízhatóságának értékelésére végeztem. A stabil eloszlású véletlenszám generálás formuláit a fejezet 5.1. szakasza tartalmazza. Egy általános stabil eloszlású véletlenszám generálására az eloszláscsalád XX. század elején történt definiálása után még hosszú ideig nem állt rendelkezésre megfelelő képlet. Az eloszlásfüggvény inverzén alapuló klasszikus módszer nem használható, mert hasonlóan az eloszlásfüggvényhez, annak inverze sem ismert zárt alakban. A problémára Chambers et al. [8] adtak 1976-ban először formulát. Ennek a formulának egy módosított változatát mutatta be Zolotarev [84]. Ez utóbbi formulát használtam a statisztikai vizsgálatokban véletlen számok generálására.

A bemutatott PIT becslési módszer hatékonyságát, pontosságát, az egyes paraméterek becslései közötti összefüggést Monte-Carlo szimulációs vizsgálat segítségével értékeltem. Az 5.2. szakaszban a szimuláció sorozat jellemzőit, és a vizsgálatok eredményeit foglaltam össze.

A PIT becslés aszimptotikus eloszlásának tesztelésére a statisztikai vizsgálat szimulációi során kapott becslésekből mintákat állítottam össze, és a becslések egyváltozós, valamint többváltozós, együttes normalitását teszteltem különböző illeszkedésvizsgálat (normalitás) tesztekkel. A tesztek eredményeit és a következtetéseimet az 5.3. szakaszban foglaltam össze. A hely- és skálaparaméter együttes becslésére ismert elméleti eredmény, amely szerint az együttes eloszlás normális, ismert kovariancia mátrixszal. A normalitásvizsgálat célja tehát hasonló eredmény empirikus kimutatása volt a három stabil paraméter becslése esetén.

Az új módszer pontosságát a szakirodalomban ismert paraméterbecslési eljárásokkal is összehasonlítottam a Weron [81] szimulációs tanulmányában közölt eredmények alapján. Az ismert módszerek és az általam bemutatott paraméterbecslési eljárás pontosságában nem mutatkozott jelentős eltérés, nagyjából azonosnak tekinthető a módszerek pontossága. A statisztikai vizsgálat eredményeit a dolgozatom 5.4. szakasza tartalmazza, amelyeket a Csenedes és Fegyverneki [S2] tanulmány alapján mutatok be.

A 6. fejezet az új paraméterbecslési eljárás valós adatokon történő alkalmazásával kapott eredményeket tartalmazza. Az elemzésben a Budapesti Értéktőzsde (BÉT) legjelentősebb, vezető részvényeinek napi záróárainak árfolyamváltozásait modelleztem. A becslött paraméterek alapján illeszkedés vizsgálatot végeztem a becslött stabil eloszláshoz képest és a normális eloszláshoz képest Kolmogorov-Smirnov, valamint  $\chi^2$  tesztekkel. Az eredmények alapján a napi záróár adatokat egyértelműen jobban leírta a becslött paraméterekkel rendelkező stabil eloszlás, mint a normális. Meghatároztam a paraméterek alapján a 95%-os megbízhatósági szinthez tartozó konfidenciaintervallumokat. Az elemzés eredményeit a Szigma című folyóiratban magyar nyelven publikált Csenedes [S3] dolgozatban foglaltam össze.

A hazai szakirodalomban is megtalálhatóak azok a dolgozatok, amelyek stabil eloszlások

kockázatkezelésben, árfolyamváltások modellezésében való alkalmazásával foglalkoznak, például a Palágyi[63], Palágyi [64], Lukács [43], valamint Lux és Varga [44] publikációk.

A 7. fejezetben rövid áttekintést adok a PIT becslési eljárást megvalósító programról, és a szimulációkat, a törzfüggvény közelítés elvégzését, a részvény árfolyamok elemzését lehetővé tevő segédprogramokról is. A programokat a MATLAB szoftvercsomag segítségével készítettem. Az elkészült programokat MATLAB fájlkként (.m fájlok) dolgozatomhoz mellékeltem.

A 8. fejezet a dolgozat eredményeit összegzi, és kijelöli a további kutatási irányokat, bemutatja a lehetséges továbbfejlesztési lehetőségeket.

## 2. A stabil eloszlások

### 2.1. Egyváltozós stabil eloszlások

A stabil eloszlások családjának pontos definiálását P. Levy [40] végezte el. Dolgozatában független, azonos eloszlású valószínűségi változók összegeit tanulmányozta, és konvergencia tételeket bizonyított az összegekre. Az eloszláscsalád matematikai statisztikai jelentőségét az adja, hogy a centrális határeloszlás tétel általánosításaként adódó vonzási tartomány (domain of attraction) probléma lehetséges megoldását kizárólag ez a család tartalmazhatja. A vonzási tartomány problémában az összegzett változók függetlenek, azonos eloszlásúak, de a szórásnégyzetük nem véges. Ezért a stabil eloszlások az általánosított centrális határeloszlás tétel alapján a normális eloszlás általánosítását adják.

A stabil eloszlások részletes leírása található Gnedenko és Kolmogorov [25], Feller [21], Uchaikin és Zolotarev [79], valamint Samorodnitsky és Taqqu [71] munkáiban. A vonzási tartomány problémát a Statisztikai Enciklopédia alapján ismertetem.

**1. Definíció.** *Legyenek  $X_1, X_2, \dots$  független, azonos eloszlású valószínűségi változók. Legyen  $X$  valószínűségi változó. Ekkor ha léteznek  $b(n)$  és  $g(n)$  konstansok úgy hogy*

$$\frac{\sum_{i=1}^n X_i}{b(n)} - g(n) \rightarrow X, \quad (1)$$

*ahol  $\rightarrow$  eloszlásbeli konvergenciát jelöl, akkor azt mondjuk,  $X$  vonzza  $X_1$ -et.*

Azoknak a valószínűségi változóknak az összességét, amelyeket  $X$  vonz,  $X$  vonzási tartományának nevezzük. Legyen  $X_1$  és  $X$  eloszlásfüggvénye  $F$  és  $G$ . Ekkor gyakran azt mondjuk,  $G$  vonzza  $F$ -et, és azoknak az eloszlásfüggvényeknek az összességét, amelyeket  $G$  vonz,  $G$  vonzási tartományának nevezzük. A stabil eloszlások elmélete lehetővé teszi, hogy leírjuk azokat az eloszlásokat, amelyek nem üres vonzási tartománnyal rendelkeznek.

**2. Definíció.** *Egy  $X$  valószínűségi változót stabilnak nevezünk, ha minden  $n$ -re léteznek olyan  $X_1, X_2, \dots, X_n$  független valószínűségi változók, melyeknek közös az eloszlása, és amely eloszlás megegyezik  $X$  eloszlásával, továbbá léteznek olyan  $e(n)$  és  $a(n)$  konstansok, úgy hogy*

$$\frac{\sum_{i=1}^n X_i}{a(n)} - e(n) \quad (2)$$

*eloszlása megegyezik  $X$  eloszlásával.*

A definícióban  $e(n)$  centráló,  $a(n)$  skálázó szerepet tölt be. Az alapvető eredmény a stabil eloszlásokról a következő.

**1. Tétel.** *Egy  $X$  valószínűségi változó vonzási tartománya akkor és csak akkor nem üres, ha  $X$  stabil.*

A centrális határeloszlás tételből tudjuk, hogy egy  $X_1$  valószínűségi változó, amely véges szórásnégyzettel rendelkezik, a normális eloszlás vonzási tartományába tartozik. Ha az 1. definícióban  $b(n) = n^{1/2}$ -nek választjuk, akkor a  $b(n)$ -nel való normalizálást feltételezve vizsgálhatjuk azt a kérdést, hogy mely valószínűségi változók tartoznak a normális eloszlás vonzási tartományába. Ekkor arra az eredményre jutunk, hogy csakis a véges szórásnégyzettel rendelkező eloszlások tartoznak a tartományba. Másrészt ha  $b(n)$ -et szabadabban választhatjuk meg, akkor a normális eloszláshoz tartó eloszlások köre szélesebb.

**2. Tétel.** Az  $X_1, X_2, \dots$  független, azonos eloszlású változók

$$Z_n = \frac{\sum_{i=1}^n X_i - a_n}{b_n}$$

normalizált összegének  $F(x)$  eloszlásfüggvénye valamilyen  $a_n$  és  $b_n > 0$  esetén a normális eloszláshoz tart akkor és csak akkor, ha

$$\lim_{x \rightarrow \infty} \frac{x^2 \int_{|y|>x} dF(y)}{\int_{|y|\leq x} y^2 dF(y)} = 0.$$

Ha megvizsgáljuk a (2) formulát, azt kapjuk, hogy  $a(n)$  csakis  $a(n) = cn^\alpha$  alakú lehet, ahol  $0 < \alpha \leq 2$ ,  $c > 0$ . Itt  $\alpha$  a stabil eloszlás stabilitási indexét jelöli. Ez azt mutatja, hogy  $b(n) = n^\alpha$  megfelelő normalizálás lehet az  $\alpha$  indexű stabil eloszlások esetében az (1) formula szerint. Ha (1) teljesül ilyen  $b(n)$ -re, akkor azt mondjuk,  $X_1$  a stabil eloszlás normális vonzási tartományában van. Már láttuk a normális eloszlás vonzási tartományánál, hogy egy  $X_1$  valószínűségi változó benne lehet egy  $X$  változó vonzási tartományában anélkül, hogy benne lenne az  $X$  változó normális vonzási tartományában.

**3. Definíció.** Egy  $L : (0, \infty) \rightarrow (0, \infty)$  függvényt lassan változó függvénynek nevezzük ( $\infty$ -ben), ha minden  $a > 0$ -ra

$$\lim_{x \rightarrow \infty} \frac{L(ax)}{L(x)} = 1.$$

**3. Tétel.** Az  $F$  eloszlás az  $\alpha$ -stabil eloszlás vonzási tartományához tartozik  $0 \leq \alpha < 2$  esetén akkor és csak akkor, ha léteznek olyan  $C_- > 0$  és  $C_+ > 0$  számok, hogy fennálljon valamilyen  $L(u)$ ,  $u > 0$  lassan változó függvény és  $u \rightarrow \infty$  mellett

$$F(-u) = (C_- + o(1))u^{-\alpha}L(u),$$

és

$$1 - F(u) = (C_+ + o(1))u^{-\alpha}L(u).$$

Elképzelhető, hogy a két eloszlásfarcok közül csak az egyik létezik.

A vonzási tartományok elmélete fontos gyakorlati alkalmazási feladatokban is szerepet kap. Egy véletlen jelenséget gyakran modellezünk normális valószínűségi változókkal, azzal a megfontolással, hogy a jelenség amelyet vizsgálunk, sok független, közel azonos eloszlású mennyiség összegzéséből adódik. Azonban ez az érvelés nem igaz, ha nem akarjuk feltenni, hogy a mennyiségek eloszlása véges szórásnégyzettel rendelkezik. Ekkor a vonzási tartományok elmélete szerint a jelenséget modellezhetjük nem normális stabil változóval. Ha feltételezzük egy ésszerű közelítésként, hogy az összegzett mennyiségek azonos eloszlásúak, akkor az egyetlen megengedhető eloszlás amelyet ebből a fajta összegzésből származó jelenség leírására használhatunk, egy stabil eloszlás. A gazdasági jelenségek stabil eloszlásokkal való modellezésének emiatt bőséges szakirodalma van.

A stabil eloszlásoknak egy másfajta definíciója is adható, amely a stabilitási tulajdonságot írja le: ha stabil eloszlású változókat összegzünk, az ismét stabil eloszlású változó lesz, tehát az eloszlás típusa változatlan marad. Innen ered a stabil elnevezés.

**4. Definíció.** Két valószínűségi változó,  $X$  és  $Y$  azonos eloszlástípusú, ha léteznek  $A > 0$  és  $B \in \mathbb{R}$  konstansok, hogy

$$X \stackrel{D}{=} AY + B,$$

és  $\stackrel{D}{=}$  eloszlásban való egyenlőséget jelöli.

**5. Definíció.** (Tágabb értelmezés) Egy  $X$  valószínűségi változót stabilnak nevezünk, ha  $X_1$  és  $X_2$  független,  $X$  eloszlásával azonos eloszlású valószínűségi változók, és tetszőleges  $A, B$  konstansra fennáll

$$AX_1 + BX_2 \stackrel{D}{=} CX + D, \quad (3)$$

ahol  $C = C(A, B)$  és  $D = D(A, B)$  valós számok.

**6. Definíció.** Egy valószínűségi változó szigorúan stabil, vagy szűkebb értelemben stabil, ha  $\forall A, B$  esetén  $D = 0$ .

**7. Definíció.** Egy valószínűségi változó szimmetrikus stabil, ha stabil, és szimmetrikus az eloszlása 0 körül, azaz  $X \stackrel{D}{=} -X$ .

Az  $\alpha = 2$  stabilitási indexű  $\alpha$ -stabil eloszlás a normális eloszlás, míg az  $\alpha = 1$  stabilitási indexű eloszlás a Cauchy eloszlás. Ezek a speciális eloszlások tehát a stabil eloszlások családjának tagjai. Az általános stabil sűrűségfüggvénynek és az eloszlásfüggvénynek ezekről a speciális esetektől<sup>1</sup> eltekintve nem ismert zárt alakja. A sűrűségfüggvény integrál reprezentációval adható meg, amely a karakterisztikus függvényből levezethető.

A karakterisztikus függvény négy paraméter segítségével ír le egy általános stabil eloszlást, ezek

- a  $0 < \alpha \leq 2$  karakterisztikus kitevő (stabilitási index, farok index),
- a  $-1 \leq \beta \leq 1$  ferdeségi (aszimmetria) paraméter
- a  $\gamma > 0$  skálaparaméter
- a  $\delta \in \mathbb{R}$  helyparaméter.

A karakterisztikus függvény segítségével megfelelően jól leírhatók az eloszláscsalád tagjai. A stabil eloszlásoknak különféle parametrizációi léteznek, azaz a négy paraméter segítségével többféleképpen is felírható a karakterisztikus függvény. A szakirodalomban korábban a nem egységes megadás miatt több félreértés is adódott. A különböző parametrizációk azért alakultak ki, mert míg egyes parametrizációk algebrailag egyszerűbbek, jobban kezelhetők elméleti bizonyításoknál, addig más parametrizációk az alkalmazások szempontjából rendelkeznek jobb tulajdonságokkal. Az itt megadott karakterisztikus függvény előnye, hogy mind a négy paraméterben folytonos.

Legyen  $X \sim S(\alpha, \beta, \gamma, \delta)$  stabil eloszlású valószínűségi változó a fenti paraméterekkel. Ekkor az  $X$  változó karakterisztikus függvénye:

$$E \exp(itX) = \begin{cases} \exp(-\gamma^\alpha |t|^\alpha [1 - i\beta \tan(\frac{\pi\alpha}{2})(\text{sgnt})((\gamma|t|)^{1-\alpha} - 1)] + i\delta t), & \alpha \neq 1, \\ \exp(-\gamma |t| [1 - i\beta \frac{2}{\pi}(\text{sgnt})(\ln |t| + \ln \gamma)] + i\delta t), & \alpha = 1. \end{cases} \quad (4)$$

<sup>1</sup>Ismert továbbá az  $\alpha = 1.5$  indexű, aszimmetrikus Levy eloszlás sűrűség- és eloszlásfüggvénye.

A stabil eloszlások szimmetrikusak a nulla körül, ha  $\beta = 0$  és  $\delta = 0$ . Ebben az esetben  $X \sim S(\alpha, 0, \gamma, 0)$  karakterisztikus függvényének egyszerűbb alakja:

$$\phi(t) = e^{-\gamma^\alpha |t|^\alpha}. \quad (5)$$

Uchaikin és Zolotarev [79] megadja különböző stabil eloszlások (szigorúan stabil, ferde, szimmetrikus, stb.) esetén a sűrűségfüggvényt, amely a karakterisztikus függvényből vezethető le. Szigorúan stabil eloszlások esetén a karakterisztikus függvény logaritmusának alakja az ún. C parametrizáció szerint:

$$\ln g^C(k; \alpha, \delta, \lambda) = -\lambda |k|^\alpha \exp\{-i\alpha\delta(\pi/2)\text{sign}k\},$$

ahol

$$0 < \alpha \leq 2, |\delta| \leq \delta_\alpha = \min\{\alpha, 2 - \alpha\}, \quad \lambda > 0.$$

Ha  $\lambda = 1$ , akkor az egyszerűbb  $g^C(k; \alpha, \delta)$  jelölést használjuk. A sűrűségfüggvény ekkor

$$q(x; \alpha, \delta) = (2\pi)^{-1} \int_{-\infty}^{\infty} e^{-ikx} g(k; \alpha, \delta) dk = \pi^{-1} \Re \int_0^{\infty} \exp\{-ikx\} \exp\{-k^\alpha e^{-i\delta\pi/2}\} dk.$$

A  $\delta = 0$  behelyettesítésével kapjuk a szimmetrikus esetet:

$$q(x; \alpha, 0) = \frac{1}{\pi} \int_0^{\infty} e^{-k^\alpha} \cos(kx) dk, \quad (6)$$

Vegyük a  $\cos$  függvény sorfejtését, és a  $t = k^\alpha$  helyettesítés után

$$q(x; \alpha, 0) = \pi^{-1} \int_0^{\infty} \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m)!} (kx)^{2m} e^{-k^\alpha} dk = \quad (7)$$

$$= (\pi\alpha)^{-1} \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m)!} x^{2m} \int_0^{\infty} e^{-t} t^{(2m+1)/\alpha-1} dt = \quad (8)$$

$$= (\pi\alpha)^{-1} \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m)!} \Gamma\left(\frac{2m+1}{\alpha}\right) x^{2m} \quad (9)$$

A sor konvergencia  $\alpha \geq 1$  esetén.  $\alpha = 2$  esetén

$$q(x; 2, 0) = (2\sqrt{\pi})^{-1} \exp\{-x^2/4\},$$

azaz a normális eloszlást kapjuk, míg  $\alpha = 1$  esetén

$$q(x; 1, 0) = \pi^{-1} \sum_{m=0}^{\infty} (-x^2)^m = \frac{1}{\pi(1+x^2)},$$

azaz a Cauchy eloszlást kapjuk.

A szimmetrikus stabil eloszlásfüggvény a (7) formula integrálásával megadható:

$$G(x; \alpha, 0) = \frac{1}{2} + \frac{1}{\pi\alpha} \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m+1)!} \Gamma\left(\frac{2m+1}{\alpha}\right) x^{2m+1}, \quad (10)$$

ha  $\alpha \geq 1$ .

A következőkben a stabil eloszlások néhány fontos tulajdonságáról lesz szó, ha  $\alpha < 2$ . A stabil eloszlások szórásnégyzete ekkor nem létezik, ugyanis a második momentumot definiáló integrál nem véges az eloszlás farkának vastagsága miatt. Általánosságban igaz, hogy az  $E(|X|^p)$  momentumok

$$E(|X|^p) = \int_{-\infty}^{\infty} |x|^p f(x) dx, \quad p \in \mathbb{R}, \quad (11)$$

nem végesek, ha  $p \geq \alpha$ , ahol  $0 < \alpha < 2$  a karakterisztikus kitevő. Ebből következik, hogy a magasabb rendű momentumok sem végesek. A Cauchy eloszlásnak ( $\alpha = 1$ ) mint ismeretes, a várható értéke sem létezik.

*Standardizált* eloszlás esetén a skálaparaméter  $\gamma = 1$ , a helyparaméter  $\delta = 0$ . A standard eloszlás többféleképpen is kiválasztható az eloszláscsalád tagjai közül. Habár a sűrűségfüggvénynek nem ismert zárt alakja, ismert az a tulajdonság, hogy minden stabil eloszlás folytonos eloszlás, és létezik végtelenül sokszor differenciálható sűrűségfüggvénye. A stabil eloszlások unimodálisak, azaz a sűrűségfüggvényüknek egy lokális maximumpontja van. A módusz értékére sincsen ismert formula, de numerikus módszerekkel kiszámítható. A karakterisztikus függvények, a sűrűség- és eloszlásfüggvények mind a négy paraméterben folytonosak.

## 2.2. Többváltozós stabil eloszlások

Többváltozós stabil eloszlású változók megadása az egyváltozós eloszlásokhoz hasonlóan szintén a karakterisztikus függvény segítségével lehetséges. Viszont a többváltozós eset leírása korántsem olyan egységes, és jól kutatott terület. A következő fejezetben összefoglalom a legfontosabb megközelítéseket, amelyek a többváltozós eset leírásával foglalkoznak. A többváltozós stabil eloszlások leírása főként Nolan munkái alapján [57], [58], [59], valamint Samorodnitsky és Taquu [71] alapján készült. Szintén fontos a Press [66] által bemutatott karakterisztikus függvény reprezentáció, mert Press ezt az alakot használja a stabil portfólió modelljének felírásában. Dolgozatom ezen szakaszát publikációim közül a Csendes [S13] dolgozatra támaszkodva készítettem.

A többváltozós stabil eloszlások definiálhatók az egyváltozós eloszlásokhoz hasonlóan. Jelölje  $\mathbf{X} = (X_1, X_2, \dots, X_d)$  a  $d$ -dimenziós véletlen vektort.

**8. Definíció.** Az  $\mathbf{X}$  véletlen vektor többváltozós stabil eloszlású, ha az  $\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2, \dots$  független, azonos eloszlású véletlen vektorok esetén létezik  $a_n > 0$  és  $b_n \in \mathbb{R}^d$  úgy, hogy

$$\mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_n \stackrel{D}{=} a_n \mathbf{X} + b_n.$$

A többváltozós stabil eloszlások karakterisztikus függvényét általában a spektrálmérték segítségével definiáljuk. Ez a mérték képes megadni az egyváltozós eloszlások közötti összetett összefüggési struktúrát, és az egyváltozós határeloszlások vastagfarkú tulajdonságát is megőrzi. Legyen

$$\mathbb{S} = \{u \in \mathbb{R}^d : |u| = 1\}$$

az egység sugarú gömb, vagy egységgömb  $\mathbb{R}^d$ -ben. Feldheim [20] megmutatta, hogy minden stabil véletlen vektor felírható egy véges, az  $\mathbb{S}$  egységgömbön értelmezett  $\Gamma$  mérték, az ún. spektrálmérték és egy  $\delta \in \mathbb{R}^d$  eltolásvektor segítségével, amelyek egyértelműen meghatározzák a többváltozós karakterisztikus függvényt (Bradley és Taquu [6] alapján).



**4. Tétel.** Legyen  $0 < \alpha < 2$ , és  $X = (X_1, X_2, \dots, X_d)$  véletlen vektor. Az  $X$  véletlen vektor akkor és csak akkor stabil  $\alpha$  stabilitási indexszel, ha létezik egy véges  $\Gamma_X$  mérték az  $\mathbb{S}$  gömbön értelmezve, és egy  $\mu \in \mathbb{R}^d$  vektor, hogy  $X$  karakterisztikus függvénye

$$\phi_\alpha(t) = \begin{cases} \exp(-\int_{\mathbb{S}} |(t^T s)|^\alpha [1 - i(\text{sign}((t^T s)) \tan \frac{\pi\alpha}{2})] \Gamma_X(ds) + i(t^T \mu)), & \text{ha } \alpha \neq 1, \\ \exp(-\int_{\mathbb{S}} |(t^T s)| [1 + i\frac{2}{\pi} \text{sign}((t^T s)) \ln |(t^T s)|] \Gamma_X(ds) + i(t^T \mu)), & \text{ha } \alpha = 1. \end{cases} \quad (12)$$

A  $(\Gamma_X, \mu)$  pár egyértelmű.

Ha  $X$  szimmetrikus stabil eloszlású  $\alpha$  kitevővel  $\mathbb{R}^d$ -ben, akkor a karakterisztikus függvény a

$$\phi_\alpha(t) = \exp\left(-\int_{S_n} |(t^T s)|^\alpha \Gamma_X(ds)\right)$$

egyszerűbb alakban írható fel, ahol  $\Gamma_X$  az egyértelmű szimmetrikus spektrálmérték. Ha  $X$  többváltozós stabil eloszlású  $0 < \alpha < 2$  stabilitási indexszel, akkor  $X$  komponenseinek összes lineáris kombinációja is stabil eloszlást követ ugyanazzal az  $\alpha$ -val, azaz a stabilitási tulajdonság eszerint a definíció szerint is igaz.

Nolan a többváltozós problémát az ún. projekciós módszerrel közelítette meg, [57], [58], [59]. Ha  $X$  stabil véletlen vektor, akkor  $X$  minden egy dimenziós projekciója

$$u \cdot X = \sum u_i X_i, \quad \forall u \in \mathbb{R}^d$$

egyváltozós stabil eloszlású  $\alpha$  karakterisztikus kitevővel. Tehát együttes stabil eloszlás esetén minden  $X_j$  komponens azonos  $\alpha$ -val rendelkezik. Nolan olyan megadást javasol, amelyben az egyváltozós stabilitásokból következik a többváltozós stabilitás. Ha feltesszük, hogy  $X$  egy véletlen vektor, amelyre igaz, hogy minden egy dimenziós  $u \cdot X$  projekciója egy dimenziós stabil eloszlású, azaz

$$u \cdot X \sim S(\alpha(u), \beta(u), \gamma(u), \delta(u)),$$

akkor létezik egy olyan  $\alpha$ , amely mindegyik projekció karakterisztikus kitevője, azaz  $\alpha(u) = \alpha$  konstans. Mivel ismerjük az  $u \cdot X$  karakterisztikus függvényt minden  $u$ -ra, így ismerjük az  $X$  együttes karakterisztikus függvényét is. Tehát  $\alpha$  és a  $\beta(u), \gamma(u), \delta(u)$  függvények együtt teljesen karakterizálják az együttes eloszlást.

A projekciós módszer segítségével, azaz a  $\beta(u), \gamma(u), \delta(u)$  projekciós függvények használatával Nolan megadja a többváltozós stabil karakterisztikus függvényt. A következőkben jelölje  $\langle \cdot, \cdot \rangle$  a skaláris szorzást.

**9. Definíció.** Az  $X$  véletlen vektor stabil eloszlású  $X \sim S(\alpha, \Lambda, \delta)$ , ha  $X$  együttes karakterisztikus függvénye:

$$E \exp(i\langle u, X \rangle) = \exp\left(-\int_{\mathbb{S}} [|\langle u, s \rangle|^\alpha + i\eta(\langle u, s \rangle, \alpha)] \Lambda(ds) + i\langle u, \delta \rangle\right).$$

A spektrálmérték segítségével meghatározhatóak a projekciós függvények:

$$\gamma(u) = \left(\int_{\mathbb{S}} |\langle u, s \rangle|^\alpha \Lambda(ds)\right)^{1/\alpha},$$

$$\beta(u) = \frac{\int_{\mathbb{S}} |\langle u, s \rangle|^\alpha \text{sign}(\langle u, s \rangle) \Lambda(ds)}{\gamma(u)^\alpha},$$

$$\delta(u) = \begin{cases} \langle u, \delta \rangle, & \alpha \neq 1, \\ \langle u, \delta \rangle - \frac{2}{\pi} \int_{\mathbb{S}} \langle u, s \rangle \ln |\langle u, s \rangle| \Lambda(ds), & \alpha = 1. \end{cases}$$

Nem csak a spektrálmérték segítségével írhatók fel a projekciós függvények, hanem fordítva is igaz, azaz a  $\beta(\cdot)$ ,  $\gamma(\cdot)$ ,  $\delta(\cdot)$  projekciós függvények meghatározzák a  $\Lambda$  spektrálmértéket. Általában erre nem létezik egyszerű képlet, ezért Nolan et al. [60] numerikus közelítési eljárást ad a spektrálmérték meghatározására a projekciós függvények segítségével.

A projekciós módszer használatának előnye, hogy a spektrálmértékes forma matematikailag egyszerűbb, segítségével a sűrűségfüggvény és az eloszlásfüggvény felírható. Ehhez Nolan [58] speciális függvényeket definiál, amelyekkel a függvényeket ún. polár koordinátás reprezentációban írja fel. Abdul-Hamid és Nolan [2] eredménye alapján a sűrűségfüggvény a polár koordinátákkal adott függvények segítségével meghatározható. Az eredmények sajnos csak korlátozottan használhatóak, mert a formulák kiszámításánál  $(d - 1)$ -szeres integrált kellene vizsgálni. Ez  $d = 2$ ,  $d = 3$  esetben még kiszámítható, de magasabb dimenzióban már nagyon bonyolult és számításigényes.

Nolan [57] megadja a diszkrét spektrálmérték esetén érvényes formulát is a karakterisztikus függvényre. Tételezzünk fel véges számú pontot és pont tömeget. Ekkor a spektrálmérték

$$\Gamma(A) = \sum_{j=1}^k \gamma_j \delta_{s_j}(A),$$

ahol  $\gamma_j > 0$  súlyokat,  $\delta_{s_j}$  ponttömegeket jelöl az  $s_j \in \mathbb{S}$ ,  $j = 1, \dots, k$  pontokhoz rendelve.

$$\delta_s(A) = \begin{cases} 1, & \text{ha } s \in A, \\ 0, & \text{egyébként.} \end{cases}$$

Az ilyen típusú spektrálmérték sok esetben előfordul, például amikor az  $X$  komponensek függetlenek vagy mintából becsüljük a spektrálmértéket. A diszkrét spektrálmértékkel rendelkező eloszlások egy speciális családot alkotnak. A karakterisztikus függvény alakja diszkrét spektrálmérték esetén

$$\phi^*(t) = \exp \left( - \sum_{j=1}^k \psi_\alpha(\langle t, s_j \rangle) \gamma_j \right), \quad (13)$$

ahol

$$\psi_\alpha(u) = \begin{cases} |u|^\alpha \left( 1 - i(\text{sign}(u) \tan \frac{\pi\alpha}{2}) \right), & \text{ha } \alpha \neq 1, \\ |u| \left( 1 + i \frac{2}{\pi} \text{sign}(u) \ln |u| \right), & \text{ha } \alpha = 1. \end{cases}$$

A  $\phi^*(t)$  egyszerűen kiszámítható, numerikusan könnyen kezelhető, míg a  $\phi(t)$  általános karakterisztikus függvény nem az.

Legyen  $p$  egy (12) típusú karakterisztikus függvényhez, és legyen  $p^*$  egy (13) típusú karakterisztikus függvényhez tartozó sűrűségfüggvény.

**5. Tétel (Byczkowski, Nolan és Rajput (1993)).** Adott  $\epsilon > 0$  esetén létezik  $n = n(d, \alpha, \epsilon, \Gamma)$ , és  $s_1, s_2, \dots, s_n$  valamint  $\gamma_1, \gamma_2, \dots, \gamma_n$  értékek, úgy hogy

$$\sup_{x \in \mathbb{R}^d} |p(x) - p^*(x)| < \epsilon.$$

Tehát az általános spektrálmértékkel rendelkező eloszlások tetszőlegesen pontosan közelíthetők diszkrét spektrálmértékkel. A tétel megadja a pontok kiválasztásának módját is. A diszkrét spektrálmértékkel rendelkező eloszlások családja sűrű a stabil eloszlások között.

A következő tétel, amely az eloszlás farkának viselkedéséről szól, a  $\Gamma$  spektrálmérték és az  $X$  eloszlása közötti kapcsolatot írja le. Definiáljunk az  $A \subset \mathbb{S}$  halmazhoz egy kúpot:

$$\text{Cone}(A) = \left\{ x \in \mathbb{R}^d : \|x\| > 0, \frac{x}{\|x\|} \in A \right\} = \{ra : r > 0, a \in A\}.$$

**6. Tétel (Corollary 6.20, Araujo és Gine, 1980).**

$$\lim_{r \rightarrow \infty} \frac{P(X \in \text{Cone}(A), \|X\| > r)}{P(\|X\| > r)} = \frac{\Gamma(A)}{\Gamma(S_d)}$$

Tehát az a "tömeg" amelyet a  $\Gamma$  spektrálmérték az  $A$  halmazhoz rendel, meghatározza az  $X$  eloszlás farkának viselkedését az  $A$  "irányában". Az eloszlás középső részét viszont a karakterisztikus függvényben szereplő integrál tag határozza meg, amely nem függ a spektrálmérték forgatásától, így a módusz környéke és az eloszlás farkának viselkedése nagyon eltérő lehet.

Nolan [58] bemutat stabil eloszlások speciális csoportjaira vonatkozó eredményeket is, ezek a sugarasan szimmetrikus (radial symmetric) vagy izotróp eloszlások, valamint az elliptikus eloszlások. Ezeknek a speciális többváltozós eloszlásoknak a sűrűségfüggvénye az  $X$  változó hosszának (normájának) segítségével is megadható. Ezeknek a speciális eloszlások generálásáról a dolgozatom 5.1 szakaszában írok.

Egy másik megközelítést mutatott be Press [67] a többváltozós karakterisztikus függvény levezetésére. Press a szimmetrikus stabil eloszlásokat használó portfólió modelljében a karakterisztikus függvény alábbi reprezentációját használta, ezért foglalkozom ennek a reprezentációnak az előállításával.

A stabil eloszlások definícióját írjuk fel az eloszlásfüggvény segítségével.

**10. Definíció.** Az  $F(y)$  eloszlás egyváltozós stabil eloszlásfüggvény, ha minden  $b_1 > 0$  és  $b_2 > 0$  és valós  $c_1, c_2$  számokhoz tartozik egy pozitív  $b$  és egy  $c$  valós szám, úgy hogy minden  $-\infty < y < \infty$  skálárra

$$F\left(\frac{y - c_1}{b_1}\right) * F\left(\frac{y - c_2}{b_2}\right) = F\left(\frac{y - c}{b}\right),$$

ahol  $*$  jelöli a konvolúció operátort.

Azok az eloszlások, amelyek a fenti egyenletet kielégítik, az alábbi log-karakterisztikus függvénnyel rendelkeznek:

$$\log \phi(v) = ia_0 v - \gamma |v|^\alpha \left[ 1 + i\beta \frac{v}{|v|} \omega(v, \alpha) \right],$$

ahol  $-\infty < v < \infty$ ,  $-\infty < a_0 < \infty$ , a stabil paraméterek a szokásos paramétertérrel adottak:  $\gamma \geq 0$ ,  $-1 < \beta < 1$ ,  $0 < \alpha \leq 2$ . Továbbá,  $[v/|v|] = 0$ , ha  $v = 0$ , és minden  $v$ -re

$$\omega(v, \alpha) = \begin{cases} \tan(\pi\alpha/2), & \alpha \neq 1, \\ (2/\pi) \log |v|, & \alpha = 1. \end{cases} \quad (14)$$

Most jelöljön  $x$  egy  $p \times 1$  vektort, és az egyváltozós esettel analóg módon adjuk meg a többváltozós stabil definíciót.

**11. Definíció.** A  $G(x)$  eloszlásfüggvény többváltozós stabil eloszlású, ha minden skaláris  $b_1 > 0$ ,  $b_2 > 0$  párra és valós  $c_1, c_2$  vektorokra létezik egy olyan pozitív skalár  $b$  és egy valós vektor  $c$ , úgy hogy minden  $x = (x_1, \dots, x_j)^T$ ,  $-\infty < x_j < \infty$ ,  $j = 1, \dots, p$  esetén

$$G\left(\frac{x - c_1}{b_1}\right) * G\left(\frac{x - c_2}{b_2}\right) = G\left(\frac{x - c}{b}\right).$$

Levy és Feldheim bebizonyították, hogy azon eloszlásoknak a családja, amely kielégíti a fenti egyenletet a

$$\log \phi(t) = iP_1(t) - \frac{1}{2}P_2(t) + \int \left[ e^{it^T w} - 1 - \frac{it^T w}{1 + w^T w} \right] \frac{dr}{r^{\alpha+1}} d\Phi(w), \quad (15)$$

többváltozós log - karakterisztikus függvényvel rendelkezik, ahol  $t$  egy  $p \times 1$  dimenziós vektor,  $P_1(t)$  és  $P_2(t)$  első és másodfokú homogén polinomok,

$$r = (w^T w)^{1/2}$$

jelöli a  $w$  vektor hosszát,  $\Phi(w)$  egy véges mérték az  $r = 1$ ,  $p$ -dimenziós gömb felszínén értelmezve, és ezen integrálható is. Az integrál a  $p$  dimenziós egységgömbön polárkoordinátákban történik és  $0 < \alpha \leq 2$ . A fenti eredményt úgy kapjuk, hogy megkeressük a többváltozós korlátlanul osztható eloszlások kanonikus integrál reprezentációját, majd ezt specializáljuk a stabil eloszlásokra.

A (15) formulát írjuk át a

$$\log \phi(t) = ia^T t - I(t)$$

alakba, ahol

$$I(t) = (t^T t)^{\alpha/2} \{ f[t/(t^T t)^{1/2}] + ig[\alpha, t/(t^T t)^{1/2}] \}, \quad (16)$$

és

$$f[t/(t^T t)^{1/2}] = k_\alpha \int |\cos \theta|^\alpha d\Phi(w), \quad k_\alpha > 0,$$

$$g[\alpha, t/(t^T t)^{1/2}] = \begin{cases} -k_\alpha \tan \frac{\pi\alpha}{2} \int |\cos \theta|^{\alpha-1} d\Phi(w), & \text{ha } \alpha \neq 1, 0 < \alpha < 2, \\ \frac{2}{\pi} k_\alpha \int \cos \theta \log [t/(t^T t)^{1/2}] |\cos \theta| d\Phi(w), & \text{ha } \alpha = 1. \end{cases}$$

$$k_\alpha = \begin{cases} -\Gamma(-\alpha) \cos \pi\alpha/2, & \text{ha } \alpha \neq 1, \alpha \neq 2, \\ \pi/2, & \text{ha } \alpha = 1, \end{cases}$$

A képletekben  $\theta$  jelöli a szöveget a  $t$  és  $w$  vektor között,  $a$  bármilyen  $p$  dimenziós vektort jelöl, az integrálást az egység sugarú  $p$  dimenziós gömb felszínén végezzük.

Ha  $\alpha \neq 1$ ,  $0 < \alpha < 2$ , akkor

$$I(t) = k_\alpha (t^T t)^{\alpha/2} \int [\cos^2 \theta]^{\alpha/2} \left[ 1 - i \tan \frac{\pi\alpha}{2} \frac{\cos \theta}{|\cos \theta|} \right] d\Phi(w). \quad (17)$$

Mivel  $\theta$  a  $t$  és  $w$  közötti szög, ezért

$$t^T w = (t^T t)^{1/2} (w^T w)^{1/2} \cos \theta.$$

Helyettesítsük be  $\cos \theta$ -ba, és mivel  $w^T w = 1$ , az integrál a (17) formulában

$$I(t) = k_\alpha \int (t^T w w^T t)^{\alpha/2} \left[ 1 - \frac{i w^T t}{|w^T t|} \tan \frac{\pi\alpha}{2} \right] d\Phi(w)$$

átalakítást eredményezi. Tegyük fel, hogy  $\Phi(w)$  abszolút folytonos a Lebesgue mértékre nézve, és legyen  $d\Phi(w) = \Phi'(w)dw$ . Az integrálszámítás első középérték tétele értelmében

$$I(t) = c_\alpha (t^T w w^T t)^{\alpha/2} \Phi'(w_0) \left[ 1 - \frac{i w_0^T t}{|w_0^T t|} \tan \frac{\pi\alpha}{2} \right],$$

ahol  $c_\alpha = k_\alpha \int dw$ . A  $w_0$  a  $w$  egy értékét jelöli, amely a középérték tétel értelmében létezik. Legyen

$$\Omega = 2^{2/\alpha} c_\alpha^{2/\alpha} [\Phi'(w_0)]^{2/\alpha} w_0 w_0^T.$$

Jegyezzük meg, hogy  $\Omega$  pozitív szemidefinit szimmetrikus mátrix  $1 \leq r \leq p$ . Tehát

$$I(t) = \frac{1}{2} (t^T \Omega t)^{\alpha/2} \left[ 1 - \frac{i w_0^T t}{|w_0^T t|} \tan \frac{\pi\alpha}{2} \right],$$

ahol  $w_0^T w_0 = 1$ . Tegyük fel, hogy  $\Phi(w)$  diszkrét mérték  $m$  atommal,  $m = 1, 2, \dots$ , tehát  $m$  végtelen is lehet, ekkor  $I(t)$  egyszerűen a súlyozott összeg, amelyet  $\Phi(x)$  ugrásaikor értékelünk ki, azaz

$$I(t) = \frac{1}{2} \sum_{j=1}^m \left\{ (t^T \Omega_j t)^{\alpha/2} \left( 1 - \frac{i w_j^T t}{|w_j^T t|} \tan \frac{\pi\alpha}{2} \right) \right\}, \quad (18)$$

ahol

$$\Omega_j = [2k_\alpha \Phi(w_j)]^{2/\alpha} w_j w_j^T,$$

$w_j$  jelöli  $\Phi(w)$   $j$ -edik atomját, és  $w_j^T w_j = 1$  minden  $j$ -re. Ekkor  $\Omega_j$  minden  $j$ -re pozitív szemidefinit szimmetrikus mátrix,  $r_j$  ranggal,  $1 \leq r_j \leq p$ . Feltesszük, hogy az összeg bármely két tagja nem arányos egymással, hogy bizonyíthassuk a reprezentáció egyértelműségét. Ha lenne két ilyen tag, akkor azok kombinálhatók egymással, és ezáltal eggyel kevesebb tagból álló összeg keletkezne. Az általános esethez  $m$ -et végtelennek választjuk úgy, hogy végtelen számú  $\Omega_j$  mátrixunk van, amelyek bizonyos esetekben lehetnek nullák. Ha  $m$  véges, akkor sok érdekes eloszlás adódhat. Az  $m = 1$  eset különösen érdekes.

A (18) formula kifejtésével és egyszerűsítésével azt kapjuk, hogy

$$I(t) = \frac{1}{2} \sum_{j=1}^m (t^T \Omega_j t)^{\alpha/2} \left\{ 1 + i\beta(t) \tan \frac{\pi\alpha}{2} \right\}, \quad (19)$$

ahol

$$\beta(t) = \frac{\sum_{j=1}^m (t^T \Omega_j t)^{\alpha/2} \frac{(-w_j^T t)}{|w_j^T t|}}{\sum_{j=1}^m \{(t^T \Omega_j t)^{\alpha/2}\}} \quad (20)$$

és  $w_j^T w_j = 1$  minden  $j$ -re. Ha (20) formulában abszolút értéket veszünk, megkapjuk, hogy  $-1 \leq \beta \leq 1$ . Jegyezzük meg, hogy minden  $v$  skalárra

$$\beta(vt) = \frac{v}{|v|} \beta(t).$$

Tekintsük az  $\alpha = 1$  esetet. A (16) formula alapján

$$I(t) = k_1 \int [\cos^2 \theta]^{1/2} \left[ 1 + \frac{2i \cos \theta}{\pi |\cos \theta|} \log |(t^T t)^{1/2} \cos \theta| \right] d\Phi(w).$$

Helyettesítsük be  $\cos \theta$  értékét:

$$I(t) = k_1 \int (t^T w w^T t)^{1/2} \left[ 1 + \frac{2i w^T t}{\pi |w^T t|} \log |w^T t| \right] d\Phi(w),$$

az  $\alpha \neq 1$  esetre

$$I(t) = \frac{1}{2} \sum_{j=1}^m (t^T \Omega_j t)^{1/2} \left[ 1 + \frac{2i w_j^T t}{\pi |w_j^T t|} \log |w_j^T t| \right], \quad (21)$$

ahol  $w_j^T w_j = 1$ , minden  $j$ -re, és  $\Omega_j$  pozitív szemidefinit szimmetrikus mátrix  $r_j$  ranggal,  $1 \leq r_j \leq p$ . A (21) formula kifejtésével és egyszerűsítésével azt kapjuk, hogy

$$I(t) = \frac{1}{2} \sum_{j=1}^m (t^T \Omega_j t)^{1/2} \left\{ 1 + \frac{2i}{\pi} \beta_1(t) \right\}, \quad (22)$$

ahol

$$\beta_1(t) = \frac{\sum_{j=1}^m (t^T \Omega_j t)^{1/2} \frac{w_j^T t}{|w_j^T t|} \log |w_j^T t|}{\sum_{j=1}^m (t^T \Omega_j t)^{1/2}}, \quad (23)$$

ahol  $w_j^T w_j = 1$  minden  $j$ -re. Így  $p = 1$ -re

$$w_j^T w_j = w_j^2 = 1,$$

azaz  $|w_j| = 1$ . Mivel

$$\log \phi(t) = ia^T t - I(t),$$

ezért a (19), (20) valamint (22), (23) kombinálásával azt kapjuk, hogy egy eloszlás akkor és csak akkor többváltozós stabil eloszlású, ha a log-karakterisztikus függvénye az alábbi formában reprezentálható. Ha  $\alpha \neq 1$ , akkor

$$\log \phi(t) = ia^T t - \frac{1}{2} \sum_{j=1}^m (t^T \Omega_j t)^{\alpha/2} \left\{ 1 + i\beta(t) \tan \frac{\pi\alpha}{2} \right\},$$

és ha  $\alpha = 1$ , akkor

$$\log \phi(t) = ia^T t - \frac{1}{2} \sum_{j=1}^m (t^T \Omega_j t)^{\alpha/2} \left\{ 1 + \frac{2i}{\pi} \beta_1(t) \right\},$$

ahol  $\beta(t)$  és  $\beta_1(t)$  úgy, mint fent. Az akkor és csak akkor állítás abból következik, hogy a (15) egyenlet, amelyből ezeket az eredményeket levezettük, egyértelműen karakterizálja az egész családot.

Ha  $\alpha = 2$ , akkor  $\tan(\pi\alpha/2) = 0$ , és  $\log \phi(t)$  egyszerűsödik:

$$\log \phi(t) = ia^T t - \frac{1}{2} \sum_{j=1}^m (t^T \Omega_j t)$$

és

$$\log \phi(t) = ia^T t - \frac{1}{2} \sum_{j=1}^m (t^T \Sigma t),$$

amely a többváltozós normális eloszlás log-karakterisztikus függvénye. A karakterisztikus függvények ezen alakjainak segítségével írja fel Press a többváltozós stabil portfóliókiválasztási modellt. A probléma a log-karakterisztikus függvényt tartalmazó optimalizálási feladatként definiálható, melyre Press speciális esetekben egzakt megoldást is ad. A stabil eloszlások portfólió optimalizálási feladatokban betöltött szerepéről dolgozatom 3. fejezetében lesz szó.

### 3. Portfólió modellek, kockázatkezelés

A XX. században számos közgazdaságtudományi Nobel - emlék díjjal jutalmazták a modern pénzügytan alapjait lefektető elméleteket, W. Sharpe-ot a tőkepiaci árfolyamok modelljéért (Capital Asset Pricing Model, CAPM)<sup>2</sup> (1990), H. Markowitz-ot portfólió optimalizálási modelljének kidolgozásáért (1990), R. Mertont és M. Scholes-t a derivatívák (származtatott pénzügyi termékek) értékének meghatározásáért (1997). Ezeknek a meghatározó közgazdaságtani, pénzügyi modelleknek mind létezik stabil eloszlásokra épülő kiterjesztése. Ezekről a stabil eloszlást feltételező modellekről részletes áttekintést ad például Uchaikin és Zolotarev [79] (17. fejezet).

A pénzügytan másik nagy területén, a kockázatkezelésben a várható veszteségek előrejelzése az elsődleges cél, amely a várható hozamok eloszlásának a veszteségeket jelentő oldalának becslését jelenti. A legelterjedtebb kockázati mutatószámok a Value-at-Risk (VaR) és a feltételes Value-at-Risk (CVaR), melyek a kockázatos eszköz hozameloszlásának bizonyos kvantiliseként meghatározott küszöbértéket, illetve a küszöbérték feletti veszteségek várható értékét jelentik. Ezeknek a mutatóknak meghatározó szerepe van a pénzügyi szervezetek számára előírt tőketartalékok meghatározásában, ezért a becslésük minden szervezet számára kiemelt feladat. A VaR és CVaR mutatók meghatározhatók stabil eloszlású hozamok feltételezése mellett (Khindanova, Rachev, és Schwartz [36], Stoyanov [76]).

Dolgozatom ezen fejezete a portfólió optimalizálás alapmodelljét, a többváltozós stabil portfóliókiválasztási modellt (Press [66]), a tőkepiaci árfolyamok modelljét, valamint a VaR és CVaR mutatók meghatározásának módjait tekinti át röviden a következőkben.

A portfólió optimalizálás klasszikus modelljében a portfólió hozama a portfólióban található eszközök hozamainak súlyozott átlaga, a kockázatot pedig a portfólió szórásnégyzetével mérjük. A modellt H. Markowitz fogalmazta meg úttörőnek számító *Portfolio Selection* című munkájában [48] 1952-ben, amely mean-variance (a magyar szakirodalomban átlag-szórás) megközelítés néven vált ismertté. Legyen  $n$  különböző értékpapír, amelyeknek a hozamai

$$X = (X_1, \dots, X_n),$$

a várható értékük

$$\mu = (\mu_1, \dots, \mu_n),$$

$\Sigma$  jelöli a kovariancia mátrixot, és az eszközök súlyai a portfólióban

$$w = (w_1, \dots, w_n).$$

Ha feltételezzük, hogy  $X \sim \mathcal{N}(\mu, \Sigma)$  többváltozós normális eloszlású, akkor a portfólió hozamának eloszlása szintén normális  $X_p \sim \mathcal{N}(\mu_p, \sigma_p^2)$ , ahol

$$\mu_p = w^T \mu$$

és

$$\sigma_p^2 = w^T \Sigma w.$$

---

<sup>2</sup>A CAPM megalkotója W. Sharpe mellett J. Lintner



A portfólió feladat lényege, hogy meghatározott várható hozamszint elérése mellett minimalizáljuk a kockázatot:

$$\begin{aligned} \min_w w^T \Sigma w, \\ w^T \mu \geq a, \\ e^T w = 1, \end{aligned} \tag{24}$$

ahol  $T$  a transzponálást jelenti,  $e = (1, \dots, 1)$  az összegző vektor.

Az optimalizálási probléma célfüggvénye ekkor kvadratikus függvény, és a feladat a kvadratikus programozás ismert algoritmusával megoldható. A portfólió feladat felírható a fenti problémával azonos eredményre vezető lineáris programozási feladatként is, amelynek egy kvadratikus feltétele van. Ekkor a befektető a számára maximálisan elviselhető  $\sigma_{max}^2$  kockázati szint mellett maximalizálja az elérhető hozamot:

$$\begin{aligned} \max_w \mu_p, \\ \sigma_p^2 \leq \sigma_{max}^2, \\ w^T e = 1. \end{aligned} \tag{25}$$

Az optimális portfólióvektor ekkor kvadratikus programozási feladatok sorozatának megoldásával állítható elő.<sup>3</sup> A fenti modellekben a fedezetlen rövidre eladás (short-selling), azaz ha  $w_i < 0$ , nem engedélyezett. A fedezetlen rövidre eladás lehetőségét is megengedő nem korlátozott probléma, valamint további relaxált feltételekkel rendelkező problémák részletes leírása található például Lintner [42] dolgozatában.

Bár a megközelítést sok kritika érte, főként a normális eloszlás feltételezése miatt, a modell jelentősége vitathatatlan. A tapasztalati vizsgálatok hamar rámutattak, hogy a normális eloszlástól csúcsosabb és vastagabb farkú (heavy-tailed) eloszlások sokkal inkább leírják a hozamokat, azaz az árfolyamok extrém kilengésének valószínűsége nagyobb, mint azt a normális eloszlás feltételezése alapján várnánk. Az első tanulmányok, amelyek az árfolyamok változását stabil eloszlásokkal modellezték Mandelbrot [45], illetve Fama [14], [15], [16] voltak. Mandelbrot a gyapot árának változását tanulmányozva jutott arra a következtetésre, hogy a hozamok eloszlásának szórásnégyzete nem véges, valamint az eloszlás jól modellezhető stabil eloszlásokkal.

Később újabb és újabb tanulmányok, empirikus vizsgálatok jelentek meg, és mára már széles körben elterjedt a stabil portfólió modell alkalmazása, például Adler, Feldman és Taqqu [3], Bradley és Taqqu [6], Rachev [68], Rachev és Mittnik [69]. A hazai szakirodalomban is számos tanulmány található, amely stabil eloszlások pénzügyi területen való alkalmazásával foglalkozik, például Palágyi [63], Palágyi [64], Lukács [43], valamint Lux és Varga [44].

A stabil portfólió feladatban azt feltételezzük, hogy a hozamok többváltozós szimmetrikus stabil eloszlást követnek. A stabil modellt Press [66] könyve alapján mutatom be. A többváltozós karakterisztikus függvény (26) alakja hosszú levezetés eredménye (Press [67], illetve Press

<sup>3</sup>A (24) és (25) optimalizálási feladatok megoldásához nincs szükség a normalitás feltételezésére. A hozamok normalitásának feltételezése azt teszi lehetővé, hogy a kockázatot a portfólióban szereplő eszközök hozamainak kovariancia mátrixával mérjük.

[66], 6. fejezet), amelyből a fontosabb lépések megtalálhatók dolgozatom 2.1 szakaszában. A Press által bemutatott modellben vizsgált többváltozós szimmetrikus ( $\beta = 0$ ) eloszláscsalád log-karakterisztikus függvénye:

$$\ln\phi_{\xi}(t) = ia^T t - \frac{1}{2} \sum_{j=1}^m (t^T \Omega_j t)^{\frac{\alpha}{2}}, \quad (26)$$

ahol

- $X$  többváltozós hozameloszlás,
- $\phi_X(t)$  jelöli az  $X$  változó karakterisztikus függvényét,
- $i = \sqrt{-1}$ ,
- $a^T = (a_1, a_2, \dots, a_n)$  vektor az eloszlás helyvektora,  $\alpha > 1$  esetén várhatóérték vektora
- $\Omega_j$  szimmetrikus skálamátrixok a változók függési struktúráját írják le,  $\forall j : \Omega_j \geq 0$  nem-negatív definit mátrixok
- $0 < m \leq \infty$  egész szám a többváltozós stabil karakterisztikus függvény előbbi reprezentációjának előállításánál során bevezetett irányok számát jelenti. Az  $n$  változós, egység sugarú hipergömb felszínén értelmezett integrálás helyett  $m$  diszkrét pont (irány) felvételével és irányonkénti ( $j = 1, \dots, m$ ) összegzéssel határozzuk meg az összefüggési struktúrát.

Feltesszük, hogy  $\sum_{j=1}^m \Omega_j > 0$ , azaz nem degenerált az eloszlás, és hogy a várható érték véges, azaz  $1 < \alpha \leq 2$ . A portfóliót alkotó részvények súlyainak vektorát továbbra is  $w^T = (w_1, w_2, \dots, w_n)$  jelöli. Ekkor a portfólió hozama

$$X_p = \sum_{i=1}^n w_i X_i,$$

a portfólió elvárt hozama

$$E(X_p) = \sum_{i=1}^n w_i E(X_i) = \sum_{i=1}^n w_i a_i = w^T a,$$

ha az  $X$  valószínűségi vektor eloszlása (26) szerinti. Az  $X_p$  portfólió hozam karakterisztikus függvénye:

$$\phi_{X_p}(v) = E(e^{ivX_p}),$$

$v \in \mathbb{R}$  a karakterisztikus függvény függő változója. Az  $X$  többváltozós hozam vektor karakterisztikus függvénye

$$\phi_X(t) = E(e^{it^T X}),$$

$t \in \mathbb{R}^n$  a karakterisztikus függvény függő változója. Legyen  $t = vw$ , ekkor

$$\phi_X(t) = \phi_X(vw) = E(e^{ivw^T X}) = \phi_{X_p}(v),$$

azaz  $X_p$  log-karakterisztikus függvénye

$$\ln\phi_{X_p}(v) = iv(w^T a) - \frac{1}{2} |v|^\alpha \sum_{j=1}^m (w^T \Omega_j w)^{\frac{\alpha}{2}}.$$

Hasonlóan a Markowitz modellhez, a portfólió kockázatot a hozamok eloszlásának valamely szóródási mutatójával, stabil eloszlások esetében a szórás hiányában a skálaparaméterrel mérjük, amely

$$\gamma(x) = \frac{1}{2} \sum_{j=1}^m (w^T \Omega_j w)^{\frac{\alpha}{2}}. \quad (27)$$

A  $w$  portfólió vektort a következő optimalizálási feladat megoldásával kapjuk:

$$\begin{aligned} \max_x \left[ \lambda a^T w - \frac{1}{2} \sum_{j=1}^m (w^T \Omega_j w)^{\frac{\alpha}{2}} \right], \\ w^T e = 1, \\ w \geq 0. \end{aligned}$$

A fenti modellben a részvények hozama közös  $\alpha$  paraméterrel rendelkezik (Model I). Press [66] bemutat egy általánosabb problémát is (Model II), amelyben ez a megkötés már nem szerepel, azaz a részvények különböző  $\alpha$  paraméterű hozameloszlással rendelkezhetnek. Bizonyos esetben explicit megoldás adható a feladatokra a Lagrange-szorzók módszerével, továbbá bemutatja az  $m = 1$  speciális eset megoldását is. A Press által definiált problémáktól általánosabb leírást tesz lehetővé a spektrálmérték használata, amely megadja az összefüggési struktúrát a többváltozós stabil eloszlás komponensei között, lásd 2.2 szakasz.

Legyen a portfólió feladatban most  $X \in \mathbb{R}^n$  hozamvektor többváltozós stabil eloszlású (12) szerint, valamint  $w \in \mathbb{R}^n$  mint korábban. Ekkor

$$X_p = (w^T X) = \sum_{i=1}^n w_i X_i$$

stabil eloszlású  $S_\alpha(\beta_p, \gamma_p, \delta_p)$  paraméterekkel. A (12) karakterisztikus függvény alapján a  $\beta_p, \gamma_p, \delta_p$  paraméterek meghatározhatóak.

A skálaparaméter:

$$\gamma_{X_p} = \left( \int_{S_n} |(w^T s)|^\alpha \Gamma_X(ds) \right)^{\frac{1}{\alpha}}. \quad (28)$$

Legyen az  $X^T = (X_1, X_2, \dots, X_n)$  árfolyamváltozás vektor (12) karakterisztikus függvényű többváltozós stabil eloszlás  $1 < \alpha < 2$  karakterisztikus kitevővel, a jelölések a korábbiaknak megfelelőek. A skálaparamétert (28) szerint definiáljuk. A portfólió optimalizálási probléma ekkor:

$$\begin{aligned} \min_w \left[ \left( \int_{S_n} |(w^T s)|^\alpha \Gamma_X(ds) \right)^{\frac{1}{\alpha}} \right], \\ w^T \mu \geq \lambda, \\ w^T e = 1, \end{aligned} \quad (29)$$

ahol  $e$  az  $n$  dimenziós összegző vektor,  $\lambda$  a befektető által meghatározott elvárt hozamszint. Ha a kockázat  $w$  konvex függvénye, akkor a probléma kvadratikus programozási feladatok egymás utáni végrehajtásával oldható meg.

A fenti portfólió modellek célja a portfólióban szereplő részvények súlyainak megválasztása, ezáltal optimális befektetési struktúra kialakítása. A tőkepiaci árfolyamok modellje (Capital Asset Pricing Model, CAPM) a részvények elvárt hozamát a piacon elérhető kockázatmentes befektetés és a piaci portfólió különbségéből adódó többlethozam függvényében vizsgálja (Sharpe [75] és Lintner [42], J. Treynor<sup>4</sup>). A részvények kockázatosságát ebben a modellben a részvények árának a piaci változásokra adott elmozdulásaival mérjük. A CAPM a piacon egyensúlyi állapotot feltételez, ekkor az értékpapírok iránti kereslet megegyezik a kínálattal.

Jelölje  $X_M$  a piaci portfólió,  $X_i$  az  $i$ -edik részvény hozamát, és  $r$  a kockázatmentes hozamot. Ekkor a CAPM modell formálisan

$$E(X_i - r) = \beta_i E(X_M - r) \quad (30)$$

alakban írható, ahol

$$\beta_i = \frac{\text{Cov}(X_i, X_M)}{\text{Var}X_M}.$$

A részvényekre jellemző  $\beta_i$  érték<sup>5</sup> leírja a részvény és a piac kapcsolatát a piaci változásokra adott árfolyamváltozáson keresztül. A  $\beta_i$  értékek tehát a részvények kockázatosságát fejezik ki. Úgy is fogalmazhatunk, hogy "valamely részvény annyival növeli a jól diverzifikált portfólió kockázatát, amennyire érzékeny a piaci változásokra" (Brealey és Meyers [7]). Ha a  $\beta_i > 1$ , azaz az értékpapír felnagyítja a piaci mozgásokat (ugyanabba az irányba változik az árfolyama, és erőteljesebben mint a piaci portfólióé), akkor az (30) egyenlet értelmében az  $i$ -edik részvény tartása nagyobb hozamot ígér, mint a piaci portfólió. Ha  $0 < \beta_i < 1$ , azaz ugyanabba az irányba mozog a papír, mint a piac, de a részvény árának megváltozása kisebb a piac változásánál, akkor a hozam prémium is kisebb. A  $\beta$  értelmezhető a következőképpen is: ha a piacon átlagosan 1%-kal növekednek az árfolyamok, akkor a vizsgált értékpapír árfolyama  $\beta$  %-kal fog növekedni.

A 3. ábrán különböző befektetések bétája és hozama közötti kapcsolat látható. A kincstári váltó bétája és kockázati díja 0, míg a piaci portfólió bétája 1, várható kockázati díja  $E(X_m - r_f)$ . Az egyenest, mely összeköti a kincstári váltót és a piaci portfóliót, értékpapír-piaci egyenesnek nevezzük. Az összes lehetséges befektetés az értékpapír - piaci egyenesen (SML, Security Market Line) helyezkedik el, vagyis a befektetések várható kockázati díja arányos a befektetés bétájával.

A piaci hozam és az értékpapír hozama közötti kapcsolatot egy  $\beta_i$  meredekségű egyenessel lehet leírni, mivel az (30) egyenlet átalakítható a

$$X_i - r = \beta_i(X_M - r) + \epsilon_i$$

alakba, ahol  $E(\epsilon_i) = 0$  és  $\text{Cov}(\epsilon, X_M) = 0$ . Ezt az egyenletet gyakran a hozamok egytényezős modelljének nevezik. A részvények kockázatára

$$\sigma_{X_i}^2 = \beta_i^2 \sigma_{X_M}^2 + \sigma_{\epsilon_i}^2$$

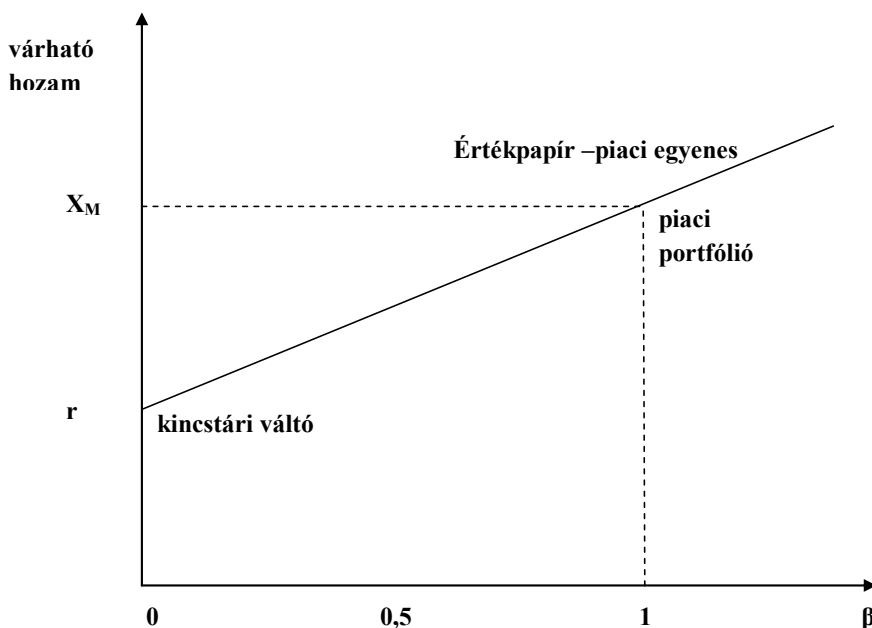
ahol az első tag jelenti a szisztematikus kockázatot, míg a második tag a reziduális kockázat. A portfólió bétáját a részvények bétájának súlyozott átlagával kapjuk, míg a portfólió kockázata hasonlóan az egyes részvényekhez

$$\sigma_{X_p}^2 = \beta_p^2 \sigma_{X_M}^2 + \sigma_{\epsilon_p}^2,$$

<sup>4</sup>J. Treynor nem publikálta az eredményeit

<sup>5</sup>A  $\beta$  jelölés itt a szakirodalomban általánosan használt jelölés, és nincs összefüggésben a stabil  $\beta$  ferdeségi paraméterrel.

1. ábra. Az értékpapír-piaci egyenes.



$$\sigma_{\epsilon_p}^2 = \sum_{i=1}^n w_i^2 \sigma_{\epsilon_i}^2,$$

ha feltételezzük, hogy a reziduálisok páronként függetlenek.

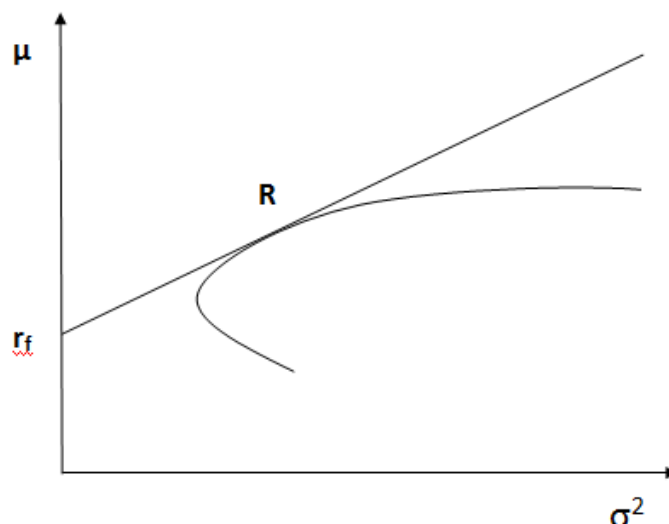
A befektetők tehát a kockázatvállalás fejében magasabb hozamot várnak el a kockázatosabb részvényektől. Ezt az összeget, azaz a piaci portfólió hozamának ( $X_m$ ) és a kockázatmentes kamatlábnak ( $r$ ) a különbségét *piaci kockázati díjnak* nevezzük.

A Markowitz modell optimalizálási feladatának megoldásai, a hatékony portfóliók halmaza ( $\mu, \sigma$ ) párokat jelent. A 3. ábrán a hatékony portfóliók halmaza látható, abban az esetben, ha kockázatmentes befektetést is választhat a befektető. Ha csak kockázatos befektetések lehetségesek, akkor a hatékony portfólió párok ( $\mu, \sigma^2$ ) konvex görbét alkotnak. Ha van kockázatmentes befektetés is, akkor a hatékony portfóliók egy kockázatos ( $R$ ) és egy kockázatmentes ( $r_f$ ) befektetés lineáris kombinációjaként állnak elő, és az  $R$  portfólióhoz tartozó ( $\mu_R, \sigma_R^2$ ) értékpár a hatékony halmazba tartozik.

A portfóliók vagy piaci eszközök kockázatoságának mérésére tehát a szakirodalomban különféle megközelítések adóttak. A Markowitz modell varianciával mért kockázatára általában teljesül, hogy a portfólió varianciája kisebb, mint a portfólióban szereplő eszközök varianciájának összege. Ez a tulajdonság kívánatos a kockázatkezelésben, de a variancia nem használható minden esetben. A tőkepiaci eszközök esetén a hozameloszlások szimmetrikusnak tekinthetők, de vannak olyan eszközök is, ahol erős aszimmetria jellemzi az eloszlást (pl. hitelportfóliók). Ilyen esetben a variancia, mint kockázati mérték nem értelmezhető (Lukács [43]).

A kockázatkezelésben ezért kidolgoztak egy új megközelítést, amely szerint a portfóliók tartásából eredő kockázatot olyan kockázati mértékkel kell mérni, amely rendelkezik a (i) monotonitás, (ii) szubadditivitás, (iii) pozitív homogenitás, (iv) eltolás ekvivalencia tulajdonságával. Ezeknek a tulajdonságoknak a megkövetelése gyakorlati szempontból fontos (Gáll és Pap

2. ábra. A hatékony portfóliók halmaza kockázatmentes befektetés lehetősége esetén.



[22]):

- (i) ha egy portfólió minden esetben többet ígér egy másik portfóliónál, akkor annak ne legyen nagyobb a kockázata;
- (ii) két portfóliót egybetéve ne növekedhessen a kockázat, azaz a portfóliók kockázatának összegét nem haladhatjuk meg;
- (iii) megtöbbszörözve a portfóliót, ám megtartva annak összetételét, a kockázatosság a nagysággal arányosan változzon;
- (iv) ha biztosan realizálunk egy pótlólagos adott összegű pénzáramlást, akkor a portfólió kockázatosságát ennek a pénzáramlásnak a nagyságával csökkenjen.

Ha egy kockázati mérték teljesíti a feltételeket, akkor koherens mértéknek nevezzük. Ezek a feltételek a szórásnégyzet vagy a stabil skálaparaméter esetében nem teljesülnek. A következőkben két kockázati mértéket mutatok be, amelyek alapvető fontosságúak: a Value-at-Risk (VaR, kockázatos érték) mutató, valamint a feltételes VaR (CVaR) vagy Expected Shortfall (ES, várható veszteség) (Bradley és Taqqu [6]).

A VaR 1990-es évek eleje óta a leginkább elterjedt módszer a piaci kockázat mérésére. Legfőbb célja a befektetések kockázatosságából eredő extrém veszteség mértékének becslése. Jelentse  $X$  valószínűségi változó a negatív profitot, azaz veszteséget, és  $F_X$  a veszteségek eloszlását, valamilyen  $t$  időhorizonton. Azaz  $X$  nagy értékei nagy veszteségeket jelölnek. Ekkor az  $0 < a < 1$  konfidencia szinthez tartozó VaR:

$$VaR_a(X) = \inf\{x | F_X(x) \geq a\},$$

$$VaR_a(X) = F_X^{-1}(a),$$

ahol  $F_X^{-1}$  jelöli az  $F_X$  eloszlásfüggvény inverzét.

A VaR hamar népszerűvé vált, mivel egyszerűen használható, és használatához nem szükséges a hozamok teljes eloszlását vagy összefüggési struktúráját ismerni. Hátránya viszont, hogy a VaR nem koherens kockázati mérték, mert nem teljesíti a szubadditivitási feltételt. Ennek ellenére a gyakorlati életben használják, például általában 10 napos 99%-os megbízhatósági szinttel határozzák meg a pénzügyi szervezetek számára az elvárt tőketartalék nagyságát.

A VaR-ral szemben az Expected Shortfall (ES) koherens kockázati mérték, amelyet feltételes VaR-nak is neveznek:

$$ES = E(L|L > VaR),$$

és a VaR fölötti veszteségek (Loss, L) várható értékét jelenti, azaz figyelembe veszi a küszöbérték fölötti lehetséges veszteségek mértékét is.

A VaR kiszámolása a gyakorlatban háromféle módon történhet, ezek: historikus szimuláció, valamely parametrikus modell használata, vagy Monte-Carlo szimuláció.

A historikus VaR-t a vizsgált  $T$  időszak veszteségeinek rendezett sorozatából számíthatjuk. Például ha 1000 időszak megfigyelései állnak rendelkezésünkre és  $X$  jelöli a hozamokat úgy hogy a veszteség pozitív, akkor a rendezett minta  $X^{(1)} \geq X^{(2)} \geq \dots \geq X^{(1000)}$ . A 95%-os VaR ekkor  $VaR_{c=0.95} = X^{(50)}$ . Ennek a módszernek az előnye az egyszerűsége, valamint hogy nem kell a hozamok eloszlására vonatkozó feltevést tennünk. Így elkerülhető a modell kockázat, ami a hibás modellfeltevésből adódna. Ha a portfólió hozamára alkalmazzuk, akkor nem kell a portfólióban szereplő részvények összefüggési struktúráját vizsgálnunk, mert a portfólió hozamában már szerepel ez a tényező. Hátránya, hogy hosszú historikus adatsorra van szükség hozzá, és hogy a múltbeli adatok jelentősen befolyásolják a VaR becslését.

Parametrikus VaR modell esetén valamilyen konkrét hozam eloszlást feltételezünk, és a rendelkezésre álló adatokból becsüljük az eloszlás paramétereit. A parametrikus VaR meghatározásánál statisztikai szempontból a legfontosabb, hogy a hozamok eloszlásának szélső kvantilisait jól becsüljük meg. A normális eloszlás használata több okból sem javasolt (Bradley és Taquu [6]). A dolgozatom 6. fejezetében használt szimmetrikus stabil modellen kívül léteznek más alternatívák is, például a hiperbolikus vagy az inverz normális eloszlás (Weron [83]), valamint a Student, Weibull, és az aszimmetrikus Weibull eloszlások (Rachev, Mitnik és Paolella [52]). Ezek is heavy-tailed eloszlások, amelyek jó eredményekkel használhatók a gyakorlatban.

A Monte - Carlo szimulációval történő VaR számítás a legrugalmasabb megoldás, de a legnagyobb számítási kapacitású is egyben. Tetszőleges összefüggési struktúrával és eloszlással definiálhatók a kockázati faktorok, és minden Monte-Carlo ismétlésben egy portfólió árváltozás kimenetelt határozunk meg. Ezekből a kimenetelekből kapunk egy eloszlást a veszteségekre nézve, amelyből kiszámítható tetszőleges szignifikancia szintű VaR a megfelelő empirikus kvantilis meghatározásával. Hátránya, hogy a módszerben nagyon sok tényezőt kell figyelembe venni, így nagyobb az esélye annak, hogy hibás modellt állítunk fel, és így jelentős lesz a modell kockázat.

Az áttekintett elméleti modellek a modern pénzügytan alapvető fontosságú modelljei. Ezen modellek gyakorlati alkalmazásánál a stabil eloszlások természetes alternatívaként merülnek fel a normális eloszlás mellett az általánosított centrális határeloszlás tétel értelmében. Amennyiben a gyakorlati alkalmazásokban parametrikus modellt szeretnénk használni, akkor az első lépés mindig a modellezett minta eloszlásának meghatározása, illetve a paraméterek megbízható, pontos becslése.

## 4. A PIT paraméterbecslési eljárás

### 4.1. A stabil paraméterek becslése

A stabil eloszlások sűrűség-, eloszlás-, vagy karakterisztikus függvénye ((6), (10), (4) formulák) numerikus számításokban nehezen vagy egyáltalán nem használható. A sorfejtéssel rendelkezésre álló formulák esetén pedig túlságosan lassú a közelítés konvergenciája, így azoknak a használata jelentősen megnövelné a futási időt. Emiatt azok a statisztikai módszerek, amelyek a sűrűség vagy eloszlásfüggvényre épülnek közvetlenül, mint például a maximum likelihood (ML) módszer, nem alkalmazhatók stabil eloszlások paraméterbecslésére. Léteznek numerikus integrálást tartalmazó közelítési eljárások (Nolan [56]), de ezeknek a számítási igénye nagyon jelentős, ezért ezek a módszerek sem ajánlottak olyan gyakorlati problémákban, ahol gyors eredményre van szükség.

A paraméterek becslésére a szakirodalomban számos megközelítés létezik, amelyeket röviden ismertetek. Az ismert eljárásokról részletes leírást ad, és összehasonlító tanulmányt mutat be Weron [81], és Borak *et al.* [4].

Az egyik legegyszerűbb megközelítés lényege, hogy *log-log* skálán ábrázoljuk a megfigyeléseket és a megfigyeléshez tartozó valószínűségeket. Ekkor ha a minta  $\alpha$ -stabil eloszlású, akkor az empirikus eloszlásfüggvény pontjai a farkaknál egy  $-\alpha$  meredekségű egyeneshez illeszkednek. Ez a módszer a stabil eloszlások azon tulajdonságából következik, hogy a fark viselkedése aszimptotikusan Pareto eloszlású, azaz az  $X$  stabil valószínűségi változó eloszlásfüggvényére

$$1 - F(x) \sim (1 + \beta)C_\alpha x^{-\alpha}, \quad (31)$$

ahol

$$C_\alpha = (1/\pi)\Gamma(\alpha) \sin(\alpha\pi/2).$$

Ha  $x$  elég nagy, akkor megközelítőleg igaz, hogy

$$\log P(X > x) = \log C_\alpha(1 + \beta)\gamma^\alpha - \alpha \log x.$$

Ez a grafikus módszer nem biztosít megfelelő paraméterbecslési lehetőséget, mert csak nagy minták esetében használható, és csak az alakparaméter becslésére. Vizsgálatok azt mutatják, hogy az általános Pareto modell alkalmazása  $1 < \alpha < 2$  esetben felülbecsüli az  $\alpha$  paramétert; például egy szimulált  $\alpha = 1.9$  paraméterű mintára  $N = 10^4$  megfigyelésre 4 körüli  $\alpha$  becslés adódott (Borak *et al.* [4]).

Az aszimptotikus Pareto eloszlásra vonatkozó (31) tulajdonságból vezethetők le a farkindex becslések, amelyek az eloszlás parametrikus alakjára nézve semmilyen feltételezéssel nem élnek. Ezek a becslések csak az aszimptotikus viselkedés leírására használhatók, az egész eloszlás alakjának vizsgálatára alkalmatlanok. A farkindex becslések sem megbízhatóak, mert erősen függenek attól, hogy melyik empirikus kvantilis után tekintjük az aszimptotikus viselkedést. Habár a farkindex becslések önmagukban nem megbízható becslései a karakterisztikus kitevőnek, mégis a gyakorlatban egyszerűségük miatt használják őket. A legerjedtebb farkindex becslés a Hill becslés (Hill [31]). A Hill becslésnek számos változata ismert, amelyekről összehasonlító kritikai elemzést mutatott be McCulloch [51].

Kvantiliseken alapuló becsléseket dolgozott ki Fama és Roll [17], [18], illetve McCulloch [50]. Fama és Roll a szimmetrikus esetre adtak kezdetleges becslést empirikus megfigyelések alapján. Becslésük egyszerűen számítható, de torzított. McCulloch kiegészítette Fama és Roll



ötletét az általános esetre, eltüntette a torzítottságot, és mind a négy paraméterre konzisztens becslést adott. Ha az adatok stabil eloszlásból származnak, és a minta kellően nagy, akkor ez a módszer megbízható becslést szolgáltat. A módszer jelentős számítási igénnyel rendelkezik, illetve bizonyos segéd konstansokra is szükség van, amelyeknek a meghatározása nem egyszerű. Szimulációs vizsgálatok alapján a McCulloch-féle becslés pontossága kielégítő (Weron [81]).

A maximum likelihood (ML) becslés a hagyományos módon nem alkalmazható, mert a sűrűségfüggvény nem ismert zárt alakban. A ML módszerrel való becsléhez Nolan [55] dolgozott ki numerikus eljárást. A STABLE program<sup>6</sup> a sűrűség és eloszlásfüggvény közelítésén kívül egyéb diagnosztikai rutinokat is tartalmaz. A paramétervektor legyen  $\vec{\theta} = (\alpha, \beta, \gamma, \delta)$ , a sűrűségfüggvény  $f(x|\vec{\theta})$ , a paramétertér

$$\Theta = (0, 2] \times [-1, 1] \times (0, \infty) \times (-\infty, \infty).$$

Adott egy  $X_1, X_2, \dots, X_n$  minta, a loglikelihood függvény ekkor

$$l(\vec{\theta}) = \sum_{i=1}^n \log f(X_i|\vec{\theta}). \quad (32)$$

Nolan módszere a McCulloch által leírt kvantilis becslést használja közelítő kezdeti értéknek, majd feltételes kvázi-Newton módszerrel maximalizálja a likelihood függvényt. A ML becslés ismert kedvező tulajdonságai, úgy mint a konzisztencia és aszimptotikus normalitás a stabil eloszlások esetében is érvényesülnek. Ellenben számítási igénye jelentős, ahogy Weron [81] fogalmaz, online valós idejű számításokra egyáltalán nem használható.

A karakterisztikus függvény közvetlenül is használható a szimmetrikus esetben, mert ebben az esetben a függvény valós függvény. Általános esetben is, mivel a karakterisztikus függvény ismert, ezért sok szerző használja, kombinálja azt különböző módszerekkel, pl. momentumok módszere, ML módszer. Az empirikus karakterisztikus függvényt használja Press [66] minimális távolságon alapuló becslésekhez. A karakterisztikus függvényen alapuló regressziós módszert mutatott be Koutrouvelis [39], majd Kogon és Williams [38] a javítását.

A momentumok módszere alkalmazható vastag farkú eloszlások paraméterbecslésére az  $r$ -edik törtmomentumok (abszolút törtmomentumok) használatával, ahol  $m_r := E(|X|^r)$ ,  $0 < r < 2$  az  $r$ -edik törtmomentum,  $m_{\mu,r} := E(|X - \mu|^r)$ ,  $0 < r < 2$  az  $r$ -edik  $\mu$ -centrális törtmomentum (Matsui és Pawlas [49]). A törtmomentumok a valószínűségi változó pozitív és negatív részének megfelelő momentumaival adhatók meg. Megjegyezzük, hogy csak az  $\alpha$  karakterisztikus kitevőnél alacsonyabb rendű törtmomentumok léteznek, azaz  $\exists E(|X|^r)$ , ha  $r < \alpha$ . A normális eloszlás ( $\alpha = 2$ ) kivételt képez, amelynek természetesen létezik második momentuma.

Újabb módszerek Bayes statisztikai megközelítéseket használnak. Garcia *et al.* [24] az indirekt következtetés módszerét használja, amelyben egy ferde t-eloszlású segédmodellt ír fel, és a segédmodell paramétereit rendeli hozzá a stabil modelleloszlás paramétereikhez. A paraméterek számítása a segéd-modell pseudo likelihood függvényével történik. Oral and Erdemir [62] a Metropolis random walk chain módszert, és közvetlen numerikus integrálást használ.

<sup>6</sup>A STABLE program elérhető J. P. Nolan stabil eloszlásokkal foglalkozó weboldaláról [61].

## 4.2. A robusztus statisztika

A következő szakaszban a robusztus statisztika alapvető fogalmait ismertetem, amelyek elengedhetetlenek a dolgozatom 4.3. fejezetében ismertetésre kerülő paraméterbecslési eljárás bemutatásához. A robusztus statisztika legfontosabb eredményeit a Huber [32] és Hampel *et al.* [27] könyvek tartalmazzák. A magyar szakirodalomban Kerékfy [35] foglalta össze a cikk megjelenéséig (1978) elért eredményeket a témában. A robusztusság kifejezést (robustness) Box [5] használta először. Itt Kendall és Buckland [34] statisztika szótárának általános értelmezését adom meg. A robusztus statisztika fogalmait Kerékfy [35] dolgozata alapján mutatom be.

**Robusztusság:** *Sok statisztikai módszer, beleértve a valószínűségi szinteket, függ a feltételek pontosságától, pl. a vizsgált változó normális eloszlású-e. Ha a feltételek változására az eredmények csak kissé befolyásolódnak, pl. ha egy próba szignifikancia pontjai csak kissé változnak, ha a populáció lényegesen eltér a normálistól, akkor a próbát robusztusnak nevezzük. Még általánosabb értelemben egy statisztikai eljárás robusztus, ha nem nagyon érzékeny azokra a feltételekre, amelyektől függ.*

A statisztikai becslélméletben gyakran feltesszük, hogy a megfigyelt mennyiségek független valószínűségi változók, közös  $F_{\theta_0}$  eloszlással, amely egy meghatározott

$$\{F_{\theta} : \theta \in \Theta\}$$

eloszláshalmazba tartozik. A paramétertér általában az  $\mathbb{R}^k$   $k$ -dimenziós Euklideszi tér egy részhalmaza. Megpróbáljuk  $\theta_0$  értékét a megfigyelések alapján becsülni, azaz keresünk egy leképezést, amely az összes megfigyelések halmazát  $\Theta$ -ba képezi le és  $\theta_0$ -hoz közeli értéket vesz fel nagy valószínűséggel, ha  $F_{\theta_0}$  a valódi eloszlás. Az ilyen felépítést hívjuk paraméteres modellnek.

Sajnos a paraméteres modellek szinte sohasem igazak, azaz a valódi eloszlás sosem egyezik meg a feltételezett paraméteres modellben szereplő eloszlások egyikével sem. Az eltéréseknek többféle oka lehet, amelyek:

- (1) nagy hibák (gross error): egy értéket nem pontosan másoltak le, rosszul olvastak le a mérőeszköztől,
- (2) a mérések korlátozott pontossága, kerekítések,
- (3) a valódi eloszlás jelentősen különbözik a paraméteres modellben lévőktől, a modell csak közelítőleg érvényes, vagy a paraméter változik az idő során.

A robusztus becsléseket akkor használjuk, ha a kiugró értékeket (outliereket) nem akarjuk vizsgálni, és olyan módszert szeretnénk alkalmazni, amely jól működik kiugró értékek, és más eltérések esetén is.

A robusztus becslések az (1)-(3) hibákat kezelik azzal, hogy nem egy paraméteres modellt használnak, hanem a paraméteres modell egy tetszőlegesen kicsi környezetét. A becslések távolságát a Prohorov-távolság segítségével mérhetjük, amely a fent felsorolt hibák kvantitatív mérését teszi lehetővé. Az (1) eltérés megfelel annak, hogy egy kis tömeg tetszőlegesen helyezkedhet el, a (2) megengedi az egész tömeg elmozdulását egy kis környezetbe. A Prohorov-távolság a gyenge konvergenciára vezet, így az a (3) feltételt is tartalmazza.

Azt fogjuk megkövetelni, hogy a becslés eloszlása (a Prohorov-távolságra vonatkozóan) folytonos funkcionálja legyen a valódi eloszlásnak. Azonban ha becslésről beszélünk, akkor mindig becslések egy egész sorozatára gondolunk. Megtörténhet, hogy a megfigyelések számának növelésével a tényleges eloszlásnak egyre közelebb kell lennie a paraméteres modellhez, hogy a becslés eloszlását közel tartsuk a modellnek megfelelő eloszláshoz. Egy ilyen becsléssorozat igen rosszul fog viselkedni nagy  $n$  esetén, ezért megköveteljük még, hogy a folytonosság  $n$ -ben egyenletes legyen. Ez a robusztusság kvalitatív definíciója.

A kvantitatív vizsgálatok céljából célszerű lineáris megközelítést alkalmazni, azaz a funkcionál deriváltját vizsgálni (*von Mises-derivált*). Ez a derivált a mintatéren értelmezett függvény, és elég jól leírja a becslés lokális viselkedését. Természetes, hogy abszolút értékének felső határát használjuk a robusztusság kvantitatív mértékéül. Ez a szuprémum méri a kilógó értékek lehetséges maximális hatását, és ezt nevezzük a becslés érzékenységének (sensitivity).

Legyen  $(X, \mathbf{A})$  egy mérhető tér, ahol  $X$  szeparábilis, teljes metrikus tér,  $\mathbf{A}$  a topológia által indukált  $\sigma$ -algebra. Jelölje a metrikát  $\rho$ . Legyen  $A \in \mathbf{A}$  esetén

$$A^\varepsilon = \{x \in X \mid \rho(A, x) < \varepsilon\}.$$

**12. Definíció.** Az  $(X, \mathbf{A})$  téren értelmezett  $P$  és  $Q$  valószínűségi mértékek Prohorov-távolsága

$$\pi(P, Q) = \inf\{\varepsilon > 0 \mid \forall A \in \mathbf{A} : P(A) \leq Q(A^\varepsilon) + \varepsilon \text{ és } Q(A) \leq P(A^\varepsilon) + \varepsilon\}.$$

Legyen  $\mathbf{F}$  az  $(X, \mathbf{A})$  mérhető téren értelmezett összes valószínűségi mértékek halmaza. Ha  $n \in \mathbb{N}$ , legyen  $\mathbf{F}_n \subset \mathbf{F}$  azon atomos mértékek halmaza, amelyek atomjainak mértéke  $1/n$ , vagy annak egész többszöröse. Az  $(X_1, X_2, \dots, X_n) \in X^n$  megfigyelések meghatározzák az  $F_n$  empirikus eloszlást, és fordítva,  $F_n$  meghatároz egy sorozatot. A koordináták sorrendjétől most eltekintünk, azaz nem vesszük figyelembe a megfigyelt értékek sorrendjét. Ezt megtehetjük, ha a megfigyelések független, azonos eloszlású valószínűségi változók. Az  $\mathbf{F}_n$  elemei ekkor az  $n$  megfigyelésből álló kísérletek kimeneteleinek tekinthetők. A becsléssorozat tehát egy olyan leképezés, amely minden megfigyelés  $n$ -eshez hozzárendel egy paraméterértéket:

**13. Definíció.** A  $\{\vartheta_n\}$  sorozatot becsléssorozatnak nevezzük, ha

$$\vartheta_n : \mathbf{F}_n \rightarrow \mathbb{R}^k$$

mérhető leképezés minden  $n \in \mathbb{N}$  esetén.

A  $\theta$  paraméter becsléseinek a valós értékű statisztikákat tekintjük:

$$\vartheta_n = \vartheta_n(X_1, \dots, X_n) = \vartheta_n(F_n),$$

tágabb értelemben véve a becslés statisztikáknak egy sorozataként fogható fel,

$$\{\vartheta_n; n \geq 1\},$$

minden  $n$  lehetséges minta elemszámra. Azokat a becsléseket tekintjük, amelyek funkcionálok, azaz

$$\vartheta_n(F_n) = \vartheta(F_n), \text{ minden } F_n\text{-re és } n\text{-re,}$$

vagy aszimptotikusan kicserélhető funkcionálra. Jelölje

$$\text{domain}(\vartheta) = \{F \mid F \in \mathbf{F} \text{ és } \vartheta \text{ definiált } \mathbf{F}\text{-re}\},$$

ekkor a becslés kicserélhető aszimptotikusan funkcionálra, ha létezik  $\vartheta : \text{domain}(\vartheta) \rightarrow \mathbb{R}$  funkcionál, hogy

$$\vartheta_n(X_1, \dots, X_n) \xrightarrow{n \rightarrow \infty} \vartheta(G),$$

ahol a konvergenciát valószínűségben értjük, ha a megfigyelések független, azonos eloszlásúak a valódi  $G$  eloszlás szerint  $\text{domain}(\vartheta)$ -ben. Azt mondjuk, hogy  $\vartheta(G)$  a  $\{\vartheta_n; n \geq 1\}$  aszimptotikus értéke  $G$ -ben.

Legyen  $\xi_1, \xi_2, \dots, \xi_n, \dots$  független, azonos eloszlású valószínűségi változók megszámlálhatóan végtelen sorozata közös  $F$  eloszlással, és legyen  $F_n$  az első  $n$  által meghatározott valószínűségi mérték. A  $\vartheta_n : \mathbf{F}_n \rightarrow \mathbb{R}^k$  leképezés indukál egy valószínűségi mértéket  $\mathbb{R}^k$ -n, ez  $\vartheta_n$  eloszlása  $F$ -re vonatkozóan. Jelölje ezt a valószínűségi mértéket

$$L_F(\vartheta_n) = F^n \vartheta_n^{-1},$$

ahol  $\vartheta_n^{-1}$  a  $\vartheta_n$  teljes inverzét jelöli, azaz egy  $U \in \mathbb{R}^k$  Borel halmazra  $(L_F(\vartheta_n))(U)$  annak a valószínűsége, hogy a  $\vartheta_n$  által szolgáltatott paraméterérték az  $U$  halmazba esik:

$$L_F(\vartheta_n)(U) = F^n(\{x \in X^n \mid \vartheta_n(x) \in U\}).$$

A következő definíció egy becslés kvalitatív robusztus tulajdonságát adja meg, lényegében a  $\vartheta_n$  becsléssorozat eloszlásainak egyenletes folytonosságát (equicontinuity) írja le  $n$ -re nézve.

**14. Definíció.** A  $\{\vartheta_n \mid n \in \mathbb{N}\}$  becsléssorozat kvalitatív robusztus az  $F$  valószínűségi mértéknél, ha

$$\forall \varepsilon > 0 : \exists \delta > 0 : \forall n \in \mathbb{N}, \forall G \in \mathbf{F} : \\ \pi(F, G) < \delta \Rightarrow \pi(L_F(\vartheta_n), L_G(\vartheta_n)) < \varepsilon$$

**15. Definíció.** A  $\{\vartheta_n\}$  becsléssorozat kvalitatív robusztus az eloszlások egy  $\mathbf{G}$  halmazán, ha minden  $G \in \mathbf{G}$ -nél robusztus. Ha  $\mathbf{F} = \mathbf{G}$ , akkor  $\{\vartheta_n\}$  kvalitatív robusztus.

Vizsgáljuk meg, mi az a határ, ameddig a becslés még felhasználható információt ad. Azaz, a minta mekkora része tartalmazhat hibás megfigyeléseket ahhoz, hogy ezek a megfigyelések még ne okozzák a becslés teljes katasztrófáját (használhatatlanságát). A katasztrófpont azt méri, meddig terjed egy becslés robusztussága, azaz milyen messze lehet a valódi eloszlás a paraméteres modelltől.

**16. Definíció.** A  $\{\vartheta_n\}$  becsléssorozat  $\varepsilon^*$  katasztrófpontja  $F$ -ben

$$\varepsilon^* = \sup \{\varepsilon \leq 1; \text{létezik a paramétertérnek olyan } K_\varepsilon \subset \Theta \text{ kompakt részhalmaza,}$$

$$\text{amelyre } \pi(F, G) < \varepsilon \text{ -ből következik, hogy } G(\{\vartheta_n \in K_\varepsilon\}) \xrightarrow{n \rightarrow \infty} 1\}$$

**17. Definíció.** Adott  $\{x_1, \dots, x_n\}$  minta esetén a  $\vartheta_n$  becsléssorozat véges mintás katasztrófpontja

$$\varepsilon_n^*(T_n; x_1, \dots, x_n) = \frac{1}{n} \max\{m; \max_{i_1, \dots, i_m} \sup_{y_1, \dots, y_m} |\vartheta_n(z_1, \dots, z_n)| < \infty\},$$

ahol az  $(z_1, \dots, z_n)$  mintát úgy állítjuk elő, hogy az  $m$  számú adatpont  $x_{i_1}, \dots, x_{i_m}$  értékeit tetszőleges  $y_1, \dots, y_m$  értékekkel helyettesítjük. Ez a katasztrófpont rendszerint nem függ  $(x_1, \dots, x_n)$ -től, és csak kis mértékben függ az  $n$  minta nagyságtól. Sok esetben ha  $\varepsilon_n^*$  határértékét vesszük, azaz ha  $n \rightarrow \infty$ , akkor ez a  $\varepsilon^*$  aszimptotikus katasztrófpontot eredményezi.

A becslések differenciálhatóságával kapcsolatosan számos eredmény született, többféle differenciálhatóságot definiáltak. A következő definícióban az ún. Gateaux differenciálhatóság egy speciális esetét használjuk.

**18. Definíció.** A  $\vartheta$  funkcionál hatásfüggvénye  $F$  eloszlás esetén ( $F$ -ben):

$$IF(x; \vartheta, F) = \lim_{t \downarrow 0} \frac{\vartheta((1-t)F + t\Delta_x) - \vartheta(F)}{t},$$

ahol az  $x \in X$ -re, ahol ez a határérték létezik, és  $\Delta_x$  jelöli az  $x$  pontra koncentrált valószínűségi mértéket.

A hatásfüggvény intuitív értelmezésben az  $x$ -pontban lévő infinitezimális szennyeződés (pl. mérési pontatlanságból adódó hibás megfigyelés) hatását írja le a becslésre, melyet a szennyeződés mértékével standardizálunk. Más szóval a megfigyelésekben jelen lévő szennyeződés becslésre gyakorolt aszimptotikus torzítását méri (Huber [32]). A hatásfüggvény segítségével bizonyítható a becslések aszimptotikus normalitása és a becslések aszimptotikus varianciája is kiszámítható.

A hatásfüggvény segítségével megkapható a becslések aszimptotikus normalitása

$$L_F(\sqrt{n}[\vartheta_n - \vartheta(F)]) \xrightarrow{n \rightarrow \infty} N(0, V(\vartheta, F)),$$

sztochasztikusan, ahol az aszimptotikus variancia

$$V(\vartheta, F) = \int IF(x; \vartheta, F)^2 dF(x).$$

Ennek a formulának a segítségével kiszámítható két becslés aszimptotikus relatív hatékonysága.

**19. Definíció.** A  $\{T_n; n \geq 1\}$  és  $\{S_n; n \geq 1\}$  becslés pár esetén  $T$  és  $S$  aszimptotikus relatív hatékonysága

$$ARE_{T,S} = V(S, F)/V(T, F).$$

**20. Definíció.** Tegyük fel, hogy az  $IF(x; \vartheta, F)$  létezik. A  $\vartheta$  becslés érzékenysége (gross-error sensitivity) a hatásfüggvény abszolút értékének szuprénuma

$$\gamma^*(\vartheta, F) = \sup_x |IF(x; \vartheta, F)|,$$

ahol a szuprénumot minden olyan  $x$ -re tekintjük, ahol az  $IF(x; \vartheta, F)$  létezik.

Az érzékenység azt a legrosszabb (legnagyobb, legerősebb) hatást méri, amelyet egy kicsi, rögzített méretű szennyeződés (lokális zavaró hatás) okoz a becslés értékében. Ezért a becslés standardizált aszimptotikus torzítottságának felső korlátjának tekinthetjük.

**21. Definíció.** A  $\vartheta$  becslés  $B$ -robusztus  $F$ -ben, ha  $\gamma^*(\vartheta, F) < \infty$ .

### Az M-becslések

Legyen adott az  $(X, \mathbf{A}, P_\vartheta)$ ,  $\vartheta \in \Theta \subset \mathbb{R}^k$  statisztikai tér, a  $P_\vartheta$  eloszlássereg dominált a  $\mu$  mértékkel, azaz léteznek a sűrűségfüggvények. Ezenkívül adott a  $\xi_1, \xi_2, \dots, \xi_n$  minta, amely elemeiről tudjuk, hogy függetlenek és azonos eloszlásúak, a közös sűrűségfüggvény  $f(x, \vartheta)$ .

**22. Definíció.** Legyen  $\rho : X \times \Theta \rightarrow \mathbb{R}$ .  $M$ -becslésnek nevezzük azokat a  $T_n$  becsléseket, amelyek minimalizálják a

$$\sum_{i=1}^n \rho(\xi_i; \vartheta)$$

összeget  $\vartheta$ -ra nézve, adott minta esetén.

Az  $M$ -becslések elméleti tulajdonságairól Huber [32], Hampel *et al.*, [27] Davies [11] és Fegyverneki [19] bizonyítottak tételeket. Hampel *et al.* [27] összefoglaló munkájában leírja, hogyan vezethetők le  $M$ -becslések, valamint a robusztus becslések robusztusságának és a robusztusság mérésének különböző koncepcióit is bemutatja.

Nagyon sokszor azonosítják az  $M$ -becsléseket a  $\rho(x, \vartheta)$  függvény deriváltjai alapján (ha léteznek) felírt egyenletekkel. Tegyük fel, hogy a

$$\psi_m(x, \vartheta) = \frac{\partial \rho(x, \vartheta)}{\partial \vartheta_m} \quad (m = 1, 2, \dots, k)$$

parciális deriváltak léteznek, s ekkor az  $M$ -becslésre teljesül, hogy

$$\sum_{i=1}^n \psi_m(\xi_i, \vartheta) = 0 \quad (m = 1, 2, \dots, k).$$

Funkcionálként is definiálhatjuk az  $M$ -becsléseket, azaz  $M$ -becslésnek nevezzük azt a funkcionált, amelyre  $T(F) \in \mathbb{R}^k$  minimalizálja az

$$\int \rho(x, \vartheta) dF(x)$$

integrál értékét  $\vartheta$ -ban, ahol  $\rho : X \times \Theta \rightarrow \mathbb{R}^k$ , vagy megoldása az

$$\int \psi(x, \vartheta) dF(x) = \mathbf{0}$$

vektoregyenletnek, ahol  $\psi : X \times \Theta \rightarrow \mathbb{R}^k$ .

Az  $M$ -becslések hatásfüggvénye a következő:

$$IF(x; T, F) = M(\psi, F)^{-1} \psi(x, T(F)),$$

ahol  $M$   $k \times k$  típusú mátrix, és

$$M(\psi, F) = - \int \left[ \frac{\partial}{\partial \vartheta} \psi(x, \vartheta) \right]_{T(F)} dF(x).$$

Az aszimptotikus kovariancia mátrix

$$V(T, F) = M(\psi, F)^{-1} Q(\psi, F) [M(\psi, F)^{-1}]^T$$

ahol

$$Q(\psi, F) = \int \psi(x, T(F)) \psi(x, T(F))^T dF(x).$$

Ha a becslés konzisztens, azaz

$$\int \psi(x, \vartheta) dF_\vartheta(x) = \mathcal{O}$$

minden  $\vartheta \in \Theta$ -ra, akkor azt kapjuk, hogy

$$M(\psi, F) = \int \psi(x, \vartheta) s(x, \vartheta)^T dF_\vartheta(x),$$

ahol

$$s(x, \vartheta) = \frac{\partial}{\partial \vartheta} \ln f_\vartheta(x) = \left( \frac{\partial}{\partial \vartheta_1} \ln f_\vartheta(x), \dots, \frac{\partial}{\partial \vartheta_k} \ln f_\vartheta(x) \right)$$

ahol  $f_\vartheta$  a  $F_\vartheta$  -hoz tartozó sűrűségfüggvény. Tehát konzisztens esetben  $F_\vartheta$ -nál az M-becslés hatásfüggvényét a  $(\psi(\cdot, \vartheta))$  értékei már meghatározzák és nincs szükség a deriváltakra.

Ha  $k = 1$ , akkor a hatásfüggvény

$$IF(x, \psi, F) = \frac{\psi(x, T(F))}{-\int \frac{\partial}{\partial \vartheta} [\psi(y, \vartheta)]_{T(F)} dF(y)},$$

az aszimptotikus szórásnégyzet pedig

$$V(T, F) = \frac{\int \psi^2(x, T(F)) dF(x)}{\left[ \int \frac{\partial}{\partial T} [\psi(y, \vartheta)]_{T(F)} dF(y) \right]^2}.$$

A következőkben a helyparaméter és a skálaparaméter M-becsléseinek tulajdonságait foglalom össze. Legyen  $X = \mathbb{R}$ ,  $\Theta = \mathbb{R}$ ,  $F_T(x) = F(x - T)$ ,  $T_0 = 0$ , és ekkor  $\psi(x, T) = \psi(x - T)$  típusú  $\psi$ -függvényeket használunk. A következő tétel összefoglalja a helyparaméter M-becsléseinek tulajdonságait.

**7. Tétel (Hampel).** *Legyen  $\psi$  monoton növekvő függvény, amely negatív és pozitív értékeket is felvesz. Továbbá legyen  $T(F)$  az az érték, amelyre*

$$\int \psi(x - T(F)) dF(x) = 0.$$

*Ekkor  $T$  a helyparaméter becslése, B-robosztus, és kvalitatív robusztus  $F_0$ -nál akkor és csak akkor, ha  $\psi$  korlátos és  $T(F_0)$  egyértelmű. A katasztrófpont*

$$\varepsilon^* = \frac{\eta}{1 + \eta},$$

ahol

$$\eta = \min \left\{ -\frac{\psi(-\infty)}{\psi(+\infty)}, -\frac{\psi(+\infty)}{\psi(-\infty)} \right\}.$$

*Ha  $\psi$  nem korlátos, akkor  $T$  se nem B-robosztus se nem kvalitatív robusztus és  $\varepsilon^* = 0$ .*

A skálaparaméter becsléséhez legyen  $X = \mathbb{R}$ ,  $\Theta = (0, +\infty)$ ,  $\vartheta_0 = 1$ ,  $F_{\vartheta}(x) = F\left(\frac{x}{\vartheta}\right)$ , s ekkor  $\psi(x, \vartheta) = \psi\left(\frac{x}{\vartheta}\right)$  alakú a javasolt  $\psi$  függvény. A robustussági tulajdonságok hasonlóak, mint a helyparaméter becslése esetén, azaz ha  $\psi(x)$  páros és szigorúan monoton nő az  $\{x > 0\}$  értékekre, akkor

(1) ha  $\psi$  korlátos, akkor a megfelelő becslés  $B$ -robosztus, kvalitatív robustus és

$$\varepsilon^* = \frac{-\psi(0)}{\psi(+\infty) - \psi(0)} \leq \frac{1}{2};$$

(2) ha  $\psi$  nem korlátos, akkor se nem  $B$ -robosztus, se nem kvalitatív robustus és  $\varepsilon^* = 0$ .

A hatásfüggvény a becslés (az aszimptotikus érték) lokális robustusságára vonatkozik. Hampel [28] az aszimptotikus szórásnégyzetet is megvizsgálta, amely a konfidenciaintervallumok hosszához kapcsolódik. Erre a célra vezette be a szórásnégyzet-változás-függvényt (change-of-variance function, CVF) Rousseeuw [70]. A téma jó összefoglalását adja Hampel *et al.* [27].

Legyen  $\psi$  folytonos, differenciálható, és

$$\begin{aligned} \psi(-x) &= -\psi(x), & \int \psi^2 dF &< +\infty, \\ 0 &< \int \psi' dF &= - \int \psi(x) f'(x) dx &= \int \Lambda \psi dF < +\infty, \end{aligned}$$

ahol  $f$  az  $F$  eloszlásfüggvény sűrűségfüggvénye és szimmetrikus,  $\Lambda = (-\ln f)'$ . Ebben a szakaszban csak az ilyen tulajdonságú függvényeket tekintjük. Ekkor ha alkalmazzuk az  $A(\psi) = \int \psi^2 dF$  és  $B(\psi) = \int \psi' dF$  jelöléseket, akkor

$$\frac{\partial}{\partial t} [V(\psi, (1-t)F + tG)]_{t=0} = \int \left[ \frac{A(\psi)}{B(\psi)^2} \left( 1 + \frac{\psi^2(x)}{A(\psi)} - 2 \frac{\psi'(x)}{B(\psi)} \right) \right] dG(x).$$

**23. Definíció.** A  $\psi$  szórásnégyzet-változás-függvénye (change-of-variance function)  $F$ -nél

$$CVF(x; \psi, F) = \frac{A(\psi)}{B(\psi)^2} \left( 1 + \frac{\psi^2(x)}{A(\psi)} - 2 \frac{\psi'(x)}{B(\psi)} \right).$$

**24. Definíció.** A szórásnégyzet-változás-érzékenység (change-of-variance sensitivity)

$$\kappa^*(\psi, F) = \sup \left\{ \frac{CVF(x; \psi, F)}{V(\psi, F)} \right\}.$$

**25. Definíció.** Az  $M$ -becslés  $V$ -robosztus  $F$ -ben, ha  $\kappa^*(T, F) < \infty$ .

Rousseeuw [70] és Hampel bizonyították, hogy a  $V$ -robosztusságból következik a  $B$ -robosztusság, valamint nem csökkenő  $\psi$  függvény esetén a  $V$ -robosztusság és a  $B$ -robosztusság ekvivalens.

A hely- és skálaparaméter együttes  $M$ -becslésének nevezünk minden olyan  $(T_n, S_n)$  statisztika párt, amelyet a következő két egyenlet határoz meg:

$$\sum_{i=1}^n \psi\left(\frac{x_i - T_n}{S_n}\right) = 0,$$



$$\sum_{i=1}^n \chi\left(\frac{x_i - T_n}{S_n}\right) = 0.$$

Ebból következik, hogy  $T_n = T(F_n)$  és  $S_n = S(F_n)$  kifejezhető a  $T$  és  $S$  funkcionálok segítségével, amelyeket

$$\int \psi\left(\frac{x_i - T(F)}{S(F)}\right) F(dx) = 0$$

és

$$\int \chi\left(\frac{x_i - T(F)}{S(F)}\right) F(dx) = 0$$

definiál.

Ha  $F$ -et  $F_t = (1-t)F + t\delta_x$  -vel helyettesítjük, és vesszük a  $t$  szerinti deriváltat  $t = 0$ -ban, akkor megkapjuk a hatásgörbét. Ha  $F$  szimmetrikus,  $\psi$  páratlan,  $\chi$  páros, akkor egyszerűsítések után

$$IC(x; F, T) = \frac{\psi\left(\frac{x}{S(F)}\right)S(F)}{\int \psi'\left(\frac{x}{S(F)}\right)F(dx)},$$

$$IC(x; F, S) = \frac{\chi\left(\frac{x}{S(F)}\right)S(F)}{\int \chi'\left(\frac{x}{S(F)}\right)\frac{x}{S(F)}F(dx)}.$$

A hely- és skálaparaméter együttes M-becsléséről található részletes leírás a Huber [32] könyv 6. fejezetében. A  $(T, S)$  paraméterek együttes becslését definiáló egyenletrendszerből a becslések értékei iterációs algoritmusokkal meghatározhatóak. A problémára Dutter és Huber [13] többféle numerikus eljárást javasol. A feladat a robusztus regressziós probléma egy speciális esetének is tekinthető, ahol a klasszikus legkisebb négyzetes eltérések összegének minimalizálása helyett a az eltérések egy  $\rho$  függvényét minimalizáljuk, amely a négyzetes függvénynél lassabb ütemben emelkedik. Ezáltal a nagyobb eltérések becslésre gyakorolt súlyát csökkenteni lehet.

A robusztus regresszió problémája a következő: legyen  $n$  megfigyelésünk,  $x_1, x_2, \dots, x_n, f_i(\theta)$  a modell függvény,  $\theta = (\theta_1, \dots, \theta_n)^T$  az ismeretlen paraméter vektor. Tekintsük a következő függvényt:

$$g(\theta, \sigma) = \sum_{i=1}^n \rho\left(\frac{x_i - f_i(\theta)}{\sigma}\right)\sigma + a\sigma, \quad (33)$$

ahol  $\rho$  konvex függvény, kétszeresen differenciálható,  $\rho(t) \geq 0$  minden  $t \in \mathbb{R}$ -re,  $\rho(0) = 0$ ,  $\rho(t)/t$  konvex, ha  $t < 0$  és konkáv, ha  $t > 0$ , valamint  $\sigma > 0$ ,  $\theta \in \mathbb{R}^p$ , és

$$a = (n - p)\frac{\beta}{2},$$

ahol

$$\beta = 2\xi_\Phi(\chi(x)).$$

A  $\xi_\Phi$  a normális eloszlás szerinti várható értéket jelöli,

$$\psi(t) = \rho'(t) = \frac{d\rho(t)}{dt}$$

és

$$\chi(t) = t\psi(t) - \rho(t).$$

Az  $a$  megválasztása azt eredményezi, hogy  $\sigma$  a hibák szórásának konzisztens becslése, ha azok normális eloszlásúak. Az  $f_i$  függvényekről feltételezzük, hogy kétszer differenciálhatóak. A (33) minimalizálása ekvivalens a következő egyenletek megoldásával:

$$\sum_i \psi\left(\frac{x_i - f_i(\hat{\theta})}{\hat{\sigma}}\right) \frac{\partial f_i}{\partial \theta_j} = 0 \quad (j = 1, \dots, p)$$

és

$$\sum_i \chi\left(\frac{x_i - f_i(\hat{\theta})}{\hat{\sigma}}\right) = a.$$

Az egyenletrendszer megoldására Dutter és Huber [13] többféle algoritmust javasol. Az algoritmusoknak egylépéses és többlépéses változata is használható. A H-algoritmus linearizáláson alapszik, és minden lépésben módosítja (winsorizing) a reziduálisokat.

### 4.3. A PIT paraméterbecslési eljárás ismert eloszlástípus esetén

A következő szakaszban egy új robusztus statisztikai technikát mutatok be szimmetrikus stabil eloszlások paraméterbecslésére, az  $1 \leq \alpha \leq 2$  esetre. Az eljárás az M-becslések csoportjába tartozik, és három stabil paraméter együttes, egyidejű becslését teszi lehetővé. A bemutatott becslési eljárás rendelkezik a B-robusztus, V-robusztus és kvalitatív robusztus tulajdonságokkal. A 4.4 szakaszban a becslések kiszámításának algoritmusát is megadom.

Legyenek  $\xi_1, \xi_2, \dots$  az  $S(\alpha, \beta = 0, \gamma, \delta)$  szimmetrikus stabil eloszlásból származó megfigyelések, az  $\alpha, \gamma, \delta$  paraméterek ismeretlenek, melyeket becsülni szeretnénk. A  $\delta$  hely- és  $\gamma$  skálaparaméter becslésére a Huber [32] által definiált M-becslést alkalmazzuk, amelyhez új  $\psi$  és  $\chi$  súlyfüggvényeket definiálunk. Ezen klasszikus M-becslés használata esetén az eloszlás típusa ismert, és csak a hely- és skálaparaméter ismeretlen. Viszont szimmetrikus stabil eloszlások esetében az eloszlástípust az  $\alpha$  alakparaméter határozza meg, amelyet szintén becsülnünk kell, tehát az eloszlás típusa nem ismert.

A klasszikus M-becslés bármely eloszlástípus esetén alkalmazható a hely- és a skálaparaméter becslésére, így a stabil eloszlástípus esetében is. A bemutatott új módszerben az eloszlástípust helyettesítjük a Cauchy és a normális eloszlás ismert eloszlásfüggvényével, és a hely- és skálaparaméter együttes M-becsléséből adódó meghatározott skálaparaméter becsléssorozatot használjuk fel az alakparaméter becslésére.

A következőkben először bemutatom Fegyverneki [19] alapján a hely- és skálaparaméter együttes M-becslését lehetővé tévő súlyfüggvények felírását és egy numerikus megoldási módszert, arra az esetre, ha az alakparaméter ismert. Majd ezt az eljárást felhasználva bemutatom az alakparamétert is becslő, a három ismeretlen stabil paraméter együttes becslését meghatározó új módszert.

Ismert eloszlástípus esetén a klasszikus M-becslés a következő. Legyenek a megfigyeléseink  $x_1, x_2, \dots, x_n$  független, azonos eloszlású valószínűségi változók, melyek az  $F$  eloszlásból származnak. Legyen

$$F_0((x - T)/S) = F(x),$$

azaz  $F$  és  $F_0$  azonos típusú,  $F_0$  az eloszlástípus kitüntetett tagja, és az  $S$  skála és  $T$  helyparamétert  $F_0$ -hoz képest definiáljuk. A hely- és a skálaparaméter  $(T, S)$  együttes M-becslése  $(T_n, S_n)$  a következő egyenletrendszer megoldása:

$$\sum \psi\left(\frac{x_i - T_n}{S_n}\right) = 0, \quad (34)$$

$$\sum \chi\left(\frac{x_i - T_n}{S_n}\right) = 0, \quad (35)$$

ahol a  $T_n$  a helyparaméter,  $S_n$  a skálaparaméter aktuális becslése,  $\psi$  és  $\chi$  alkalmas súlyfüggvények,  $x_i$  jelöli a minta elemeket. A  $\psi$  és  $\chi$  súlyfüggvények megválasztása más-más becslést eredményez.

Alkalmazzuk a Probability Integral Transformation (PIT) technikát és a momentumok módszerét a  $\psi$  és  $\chi$  függvények meghatározásához. Jól ismert, hogy ha egy  $\xi$  valószínűségi változó  $F$  eloszlásfüggvénye invertálható, akkor az  $F(\xi)$  valószínűségi változó egyenletes eloszlású a  $(0, 1]$  intervallumon. Alkalmazzuk a momentumok módszerét a transzformált egyenletes valószínűségi változóra:

$$\int_{-\infty}^{\infty} F dF = \frac{1}{2}, \quad (36)$$

$$\int_{-\infty}^{\infty} \left(F - \frac{1}{2}\right)^2 dF = \frac{1}{12}. \quad (37)$$

Vagy másképpen

$$E_F\left(F_0\left(\frac{\xi - T}{S}\right)\right) = \frac{1}{2},$$

$$D_F^2\left(F_0\left(\frac{\xi - T}{S}\right)\right) = \frac{1}{12}.$$

Rendezzük át az egyenleteket és használjuk a várható érték és szórásnégyzet helyett az átlagot. A  $T$  és  $S$  paraméterek együttes M-becslését definiáló implicit függvények ekkor:

$$\sum \psi\left(\frac{x_i - T_n}{S_n}\right) = 0, \quad (38)$$

$$\sum \psi^2\left(\frac{x_i - T_n}{S_n}\right) = (n - 1)\mathcal{B}, \quad (39)$$

Tehát az M-becslés  $\psi$  és  $\chi$  súlyfüggvényei:

$$\psi(x) = F_0(x) - \frac{1}{2},$$

$$\chi(x) = \left(F_0(x) - \frac{1}{2}\right)^2 - \frac{1}{12},$$

ahol  $\mathcal{B}$  egy konstanst jelöl. Ha a mintaelemek valódi eloszlása éppen  $F_0$ , akkor  $\mathcal{B}$  értéke éppen  $1/12$ . Ha nem  $\xi$ -nek megfelelő eloszlástípust használunk, azaz a mintaelemek eloszlása nem az  $F_0$  típus, akkor

$$D^2\left(F_0\left(\frac{\xi - T}{S}\right)\right) = D^2\left(F_0\left(\frac{\xi - T}{S}\right) - \frac{1}{2} + \frac{1}{2}\right) =$$

$$D^2\left(\psi\left(\frac{\xi - T}{S}\right) + \frac{1}{2}\right) = D^2\left(\psi\left(\frac{\xi - T}{S}\right)\right) =$$

$$D^2\left(\psi\left(\frac{\xi - T}{S}\right)\right) = \frac{1}{n-1} \sum \left(\psi\left(\frac{\xi - T}{S}\right)\right)^2 = \mathcal{B},$$

azaz

$$\mathcal{B} = D_{F_\xi}^2(\psi(\xi)) \quad (40)$$

-ként áll elő.

Az (38) és (39) egyenletrendszer iteratív algoritmussal, az ún. ping-pong módszerrel oldható meg. A módszer numerikus viselkedése, konvergenciája megtalálható Dutter és Huber [13] dolgozatában. A módosított Newton módszer alapján az alábbi két egyenletet felváltva oldjuk meg: az első egyenletből kapott helyparamétert a második egyenletbe helyettesítve új skálaparaméter értékhez jutunk. Ezt a skálaparaméter becslést felhasználva ismét az első egyenletet számítjuk ki. A kívánt pontosság eléréséig ismétljük a lépéseket.

A helyparaméter közelítése:

$$T_n^{(m+1)} = T_n^{(m)} + \frac{1}{n} S_n^{(m)} \sum_{i=1}^n \psi\left(\frac{x_i - T_n^{(m)}}{S_n^{(m)}}\right), \quad (41)$$

A skálaparaméter közelítése:

$$[S_n^{(m+1)}]^2 = \frac{1}{(n-1)\mathcal{B}} \sum_{i=1}^n \psi^2\left(\frac{x_i - T_n^{(m+1)}}{S_n^{(m)}}\right) [S_n^{(m)}]^2. \quad (42)$$

A  $\psi$  súlyfüggvény

$$\psi(x) = F_0(x) - \frac{1}{2},$$

a kezdeti értékek

$$T_n^{(0)} = \text{med}\{x_i\}, \quad (43)$$

$$S_n^{(0)} = C \cdot \text{MAD}, \quad (44)$$

$\text{med}\{x_i\}$  jelöli a mediánt,  $\text{MAD}$  jelöli a medián abszolút eltérést

$$\text{MAD} = \text{med}\{|x_i - \text{med}\{x_i\}|\},$$

valamint  $S_n^{(m)}$  és  $T_n^{(m)}$  az  $S$  skála- és  $T$  helyparaméter aktuális becslései az  $m$ -edik iterációban. A  $C$  konstans értéke

$$C = F_0^{-1}(3/4),$$

amelyet a kezdeti becslés torzítatlansága miatt alkalmazunk ( $F_0$  szimmetrikus eloszlás).

Ismert eloszlástípus esetén, azaz ha  $F_0$  eloszlás ismert, akkor a ping-pong módszer segítségével a hely- és skálaparaméter együttesen becsülhető. Ebben az esetben a becslések együttes eloszlása aszimptotikusan normális, és a kovariancia mátrix megadható, Fegyverneki [19]. Legyen  $\xi = S\eta + T$ , ahol az  $\eta$  valószínűségi változó eloszlásfüggvénye  $G_0(x)$ . Legyen adott az  $\xi_1, \xi_2, \dots$  minta és  $G_0$  eloszlástípus,  $\xi_i$  valószínűségi változó eloszlása  $G_0((x - T)/S)$ .

**8. Tétel (Fegyverneki).** Tegyük fel, hogy  $G_0$  differenciálható, szigorúan monoton növekvő és  $G_0(0) = 0.5$ . Ekkor  $T_n$  és  $S_n$  jól definiáltak, azaz az (38), (39) egyenletrendszernek létezik egyértelmű megoldása, amelyre  $S_n > 0$ .

**9. Tétel.** A  $(T_n, S_n)$  becslések két dimenziós együttes eloszlása a 8. Tétel feltételeinek teljesülése esetén, normális eloszláshoz tart

$$\sqrt{n}((T_n, S_n) - (T, S)) \xrightarrow{D} N(0, \Sigma),$$

ahol a  $\Sigma$  kovariancia mátrix  $\Sigma = C^{-1}D[C^{-1}]^T$ .

A  $C$  és  $D$  mátrixok

$$C = \begin{pmatrix} E\left(\frac{\partial}{\partial T}\psi\left(\frac{\xi-T}{S}\right)\right) & E\left(\frac{\partial}{\partial S}\psi\left(\frac{\xi-T}{S}\right)\right) \\ E\left(\frac{\partial}{\partial T}\chi\left(\frac{\xi-T}{S}\right)\right) & E\left(\frac{\partial}{\partial S}\chi\left(\frac{\xi-T}{S}\right)\right) \end{pmatrix},$$

és

$$D = \begin{pmatrix} E(\psi^2(\eta)) & E(\psi(\eta)\chi(\eta)) \\ E(\psi(\eta)\chi(\eta)) & E(\chi^2(\eta)) \end{pmatrix} = \begin{pmatrix} \frac{1}{12} & 0 \\ 0 & \frac{1}{180} \end{pmatrix},$$

ahol  $\eta \sim G_0$ . A kovariancia mátrix függ az  $F_0$  típustól.

**10. Tétel.** A PIT becslések a 8. Tétel feltételeinek teljesülése esetén  $B$ -robosztusak,  $V$ -robosztusak, kvalitatív robosztusak és a katasztrófpontjaik

$$\varepsilon^*(T_n) = \frac{\delta}{1 + \delta} = 0.5, \quad \text{ahol} \quad \delta = \min \left\{ -\frac{\psi(-\infty)}{\psi(+\infty)}, -\frac{\psi(+\infty)}{\psi(-\infty)} \right\}$$

és

$$\varepsilon^*(S_n) = \frac{-\chi(0)}{\chi(-\infty) - \chi(0)} = \frac{1}{3}.$$

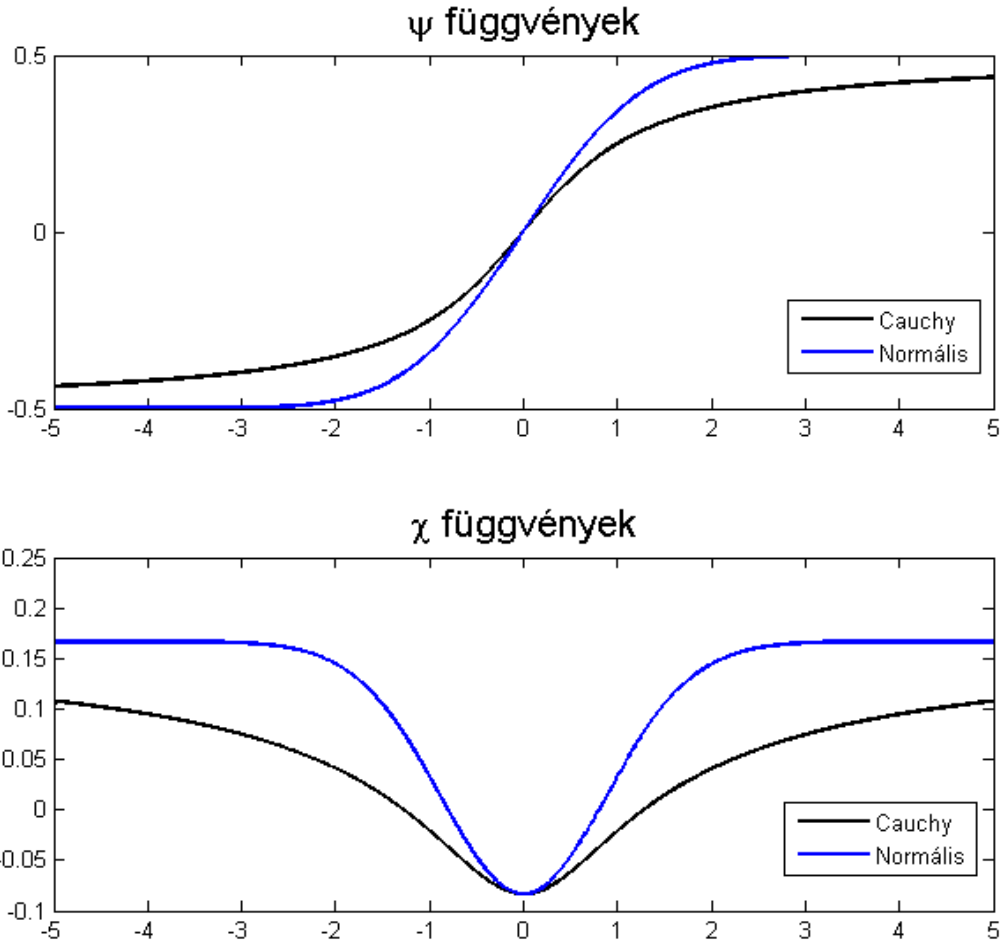
A stabil paraméterek PIT becslésének torzítatlansága következik Hampel *et al.* [27] eredményeiből felhasználva Fegyverneki [19] dolgozatát.

#### 4.4. PIT paraméterbecslés nem ismert eloszlástípus esetén

Az ismert eloszlástípus esetén tehát a hely- és skálaparaméter együttes M-becslésének robusztus jellemzői ismertek, a probléma az ismert algoritmusokkal megoldható. Ha a stabil alakparamétert is a mintából becsüljük, akkor  $F_{0,\alpha}$  eloszlástípus nem ismert. A becslési eljárásban az  $F_{0,\alpha}$  eloszlásra a  $\psi$  függvény, és a  $\mathcal{B}$  érték számításakor van szükség. Mivel az  $F_{0,\alpha}$  nem ismert, a javasolt új módszer szerint használjuk a skálaparaméter meghatározásához a stabil eloszlások családjának két ismert szimmetrikus tagját, a normális eloszlást ( $\alpha = 2$ ), és a Cauchy eloszlást ( $\alpha = 1$ ) az ismeretlen  $F_{0,\alpha}$  helyett a  $\psi$  függvényben.

**26. Definíció.** (Normális eloszlás,  $\alpha = 2$ )

$$f(\gamma, \delta; x) = \frac{1}{\sqrt{4\pi\gamma^2}} \exp\left(-\frac{(x - \delta)^2}{4\gamma^2}\right), \quad \gamma > 0, \quad \delta \in \mathbb{R}$$



3. ábra. A  $\psi$  és  $\chi$  súlyfüggvények,  $\alpha = 1$  és  $\alpha = 2$

**27. Definíció.** (Cauchy eloszlás,  $\alpha = 1$ )

$$f(\gamma, \delta; x) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (x - \delta)^2} \quad \gamma > 0, \quad \delta \in \mathbb{R}.$$

Nemcsak az iteráció súlyfüggvényeinek számításánál, hanem a  $\mathcal{B}$  értékének meghatározásához is szükséges a  $\psi$  függvényben az  $F_{0,\alpha}$  helyettesítése. A  $\mathcal{B}$  értéke továbbá függ a minta  $\alpha$  paraméterétől az integrandus miatt, ezért  $\mathcal{B}(\alpha)$  a továbbiakban az  $\alpha$  paraméter függvénye.

Helyettesítsük be egyenként a két ismert eloszlásfüggvényt,  $\Phi(x)$ -t és

$$F_{Cauchy} = 1/\pi \arctg x + 1/2,$$

a Cauchy eloszlásfüggvényt a  $\psi(x) = F_0(x) - 0.5$  súlyfüggvénybe. A 3 ábrák mutatják a  $\psi$  és  $\chi$  súlyfüggvényeket a normális és a Cauchy eloszlásfüggvény behelyettesítésével. A két  $F_0$

eloszlásfüggvény használatával két különböző  $\mathcal{B}$  függvényhez jutunk: az egyik esetben  $F_0$ -ként a Cauchy eloszlás eloszlásfüggvényét

$$\int_{-\infty}^{\infty} \left(\frac{1}{\pi} \arctg x\right)^2 dF_{\alpha} = \int_{-\infty}^{\infty} \left(\frac{1}{\pi} \arctg x\right)^2 f_{\alpha}(x) dx = \mathcal{B}_1(\alpha), \quad (45)$$

a másik esetben pedig az  $F_0$  eloszlásfüggvényként a normális eloszlás eloszlásfüggvényét használjuk

$$\int_{-\infty}^{\infty} \left(\Phi(x) - \frac{1}{2}\right)^2 dF_{\alpha} = \int_{-\infty}^{\infty} \left(\Phi(x) - \frac{1}{2}\right)^2 f_{\alpha}(x) dx = \mathcal{B}_2(\alpha), \quad (46)$$

ahol  $F_{\alpha}$  és  $f_{\alpha}$  jelöli az  $\alpha$ -stabil minta eloszlás és sűrűségfüggvényét.

Ha a  $\mathcal{B}$  értékét ezek alapján ismerjük, akkor a skálaparaméter becslését mind a két ismert eloszlásfüggvény használatával ki lehet számítani. Az integrál értékének meghatározása helyett a becslés kiszámítása során a  $\mathcal{B}$  függvényeknek egy előzetesen meghatározott racionális törtfüggvény közelítést alkalmazunk. Ezáltal elkerülhető a numerikus integrálás, azaz az algoritmus gyorsítható. A törtfüggvény közelítés meghatározásának lépéseiről szól dolgozatom 4.6. fejezete.

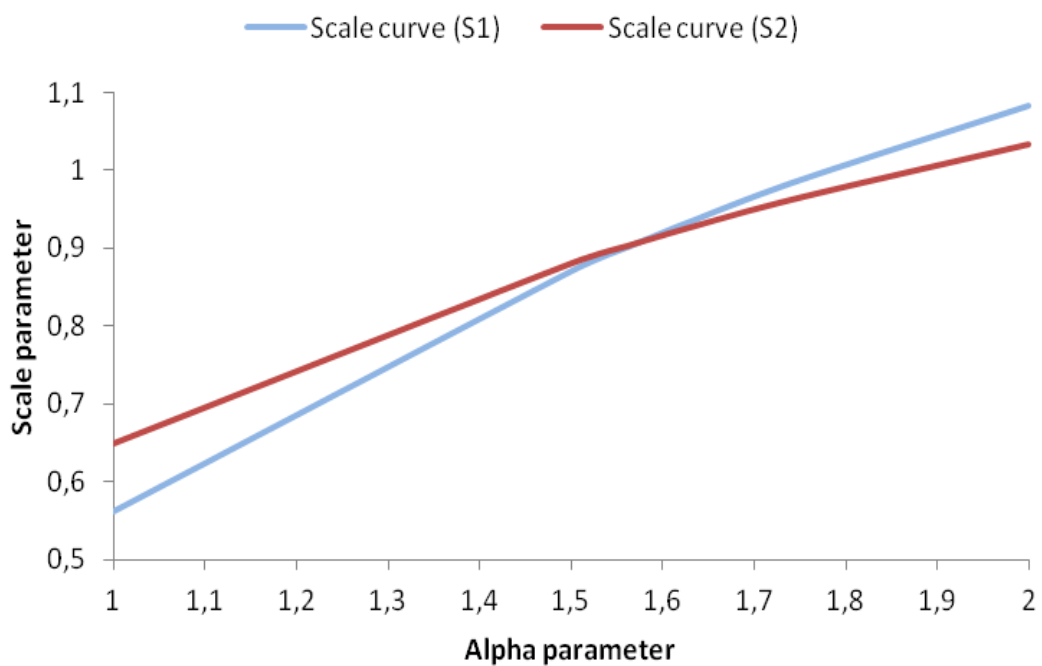
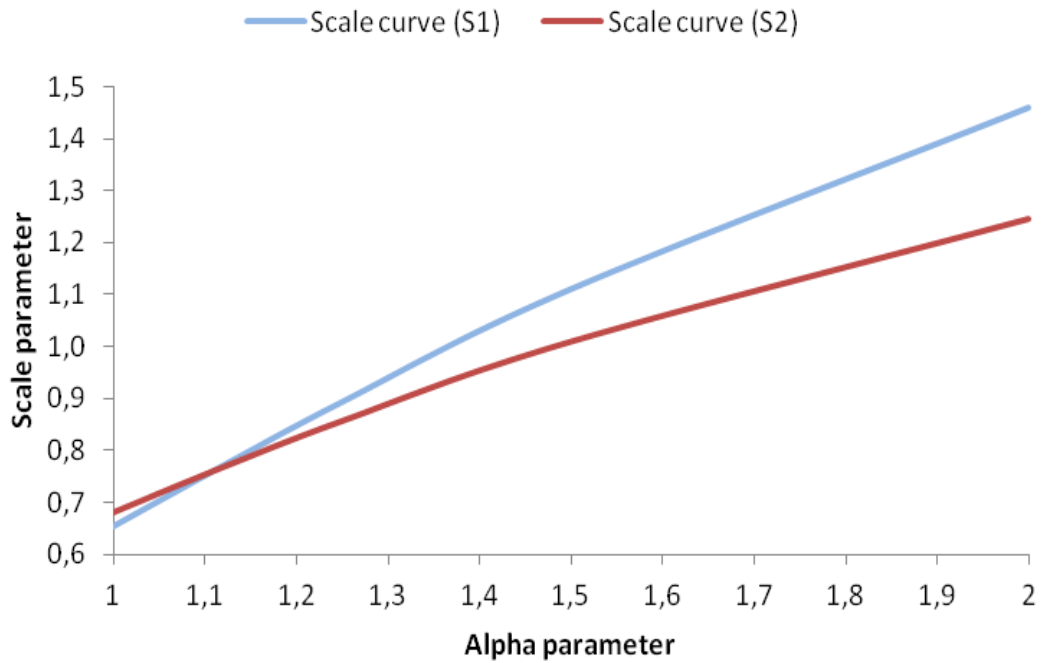
A  $\mathcal{B}$  függvények az integrandusban tartalmazzák az ismeretlen  $f_{\alpha}$   $\alpha$ -stabil sűrűségfüggvényt, tehát a  $\mathcal{B}$  függvények függenek a minta ismeretlen  $\alpha$  alakparaméterétől, amelyet becsülni szeretnénk. Úgy tűnhet, hogy a skála- és a helyparaméter kiszámításához a ping-pong módszer használatához szükségünk van az alakparaméter ismeretére. A javasolt algoritmus ezt a koncepciót fordítja meg. Az  $\alpha$  paraméter valamely feltételezett értékei esetén számítható skálaparaméterek egy becsléssorozata, és ennek a becsléssorozatnak a segítségével becsülhető a minta keresett  $\alpha$  paramétere.

Jelöljük a Cauchy és normális eloszlásfüggvény behelyettesítésével kapott függvényeket  $\mathcal{B}_1(\alpha)$  -val és  $\mathcal{B}_2(\alpha)$ -val. Jelöljük továbbá a  $\mathcal{B}_1(\alpha)$  és  $\mathcal{B}_2(\alpha)$  függvények használatával (42) egyenlet szerint számított skálaparaméter becsléseket  $S_1(\alpha)$ -val és  $S_2(\alpha)$ -val. A skálaparaméter becslések a  $\mathcal{B}$  függvényeken keresztül szintén függenek az alakparamétertől, és a skálaparaméter becslések numerikusan tetszőleges  $\alpha$  esetén kiszámíthatóak az  $[1, 2]$  intervallumon.

Ha egy rögzített minta esetén a mintának megfelelő  $\alpha$  paraméterű  $F_{\alpha} - t$  használjuk a  $\psi$  függvényben, akkor a ping-pong módszer megadja a skálaparaméter torzítatlan becslését. Ha  $F_0$ -ként nem a megfelelő stabil eloszlásfüggvényt használjuk, de az integrandusban szereplő  $f_{\alpha}$  sűrűségfüggvény megfelelő, akkor is megkapjuk a torzítatlan becslését a skálaparaméternek. Tehát, ha nem a megfelelő  $F_{0,\alpha}$  -t használjuk  $\mathcal{B}$ -ben, akkor az  $S_1(a)$  és  $S_2(a)$  skálaparaméter becslések minden  $a \in [1, 2]$ -ra el fognak térni (torzítottak lesznek), kivéve a minta keresett  $\alpha$  paraméterénél, amelyet jelöljünk  $\hat{\alpha}$  -val. A  $\hat{\alpha}$  pontban a skálaparaméter becslések a két eloszlásfüggvény használatával megegyeznek, azaz  $S_1(\hat{\alpha}) = S_2(\hat{\alpha})$ .

Az  $\alpha$  alakparaméter becslését az az  $a \in [1, 2]$  jelenti, amely pontban a két  $S_1(a)$  és  $S_2(a)$  skálaparaméter becslés megegyezik. Ha a skálaparaméter becsléseket  $\alpha$  függvényének tekintjük az  $[1, 2]$  intervallumon, akkor a skálaparaméter függvények két monoton növekvő, konkáv görbét alkotnak, amelyeknek csak egy metszéspontja létezik, a keresett  $\hat{\alpha}$ .

**1. TÉZIS.** *Ismert  $\alpha$  alakparaméterű eloszlástípus esetén az  $M$ -becslés használható a stabil eloszlások hely- és skálaparaméterének becslésére. Ha más alakparaméternek megfelelő  $F_{0,\alpha}$  eloszlásfüggvényt választunk a súlyfüggvényben, a segítségével meghatározott  $\mathcal{B}$  érték felhasználásával a hely- és skálaparaméter torzítatlan becslését adjuk ( $1 \leq \alpha \leq 2$ ).*



4. ábra. A ping-pong módszerrel meghatározott  $S_1(\alpha)$  és  $S_2(\alpha)$  skálaparaméter görbék  $n = 50$  elemből,  $\alpha = 1.2$  (felül) és  $\alpha = 1.8$  (alul) esetén



**2. TÉZIS.** *Ha az alakparaméter sem ismert, akkor a Cauchy és a normális eloszlás alapján a becsült skálaparaméterek összehasonlításával megadható az alakparaméter becslése és így a szimmetrikus stabil eloszlás alak-, hely- és skálaparamétere egyszerre becsülhető.*

#### 4.5. A becslések kiszámításának algoritmus

A skálaparaméter függvények metszéspontjának a meghatározásához egy iteratív intervallumfelezéses algoritmust használunk. A kezdeti értékek legyenek  $\alpha = 1$  és  $\alpha = 2$ , az  $\alpha$  paraméter értékének lehetséges tartományának két végpontja. Kiszámítjuk mindkét végpontban mindkét  $\mathcal{B}$  függvénnyel az  $S_1(\alpha)$  és  $S_2(\alpha)$  skálaparaméter becsléseket, ez összesen négy skála-becslést eredményez. A  $[1, 2]$  intervallum végpontjaiban az összetartozó skálabecslések közötti különbség előjelet vált, ellenkező esetben nincs az intervallumban metszéspont.

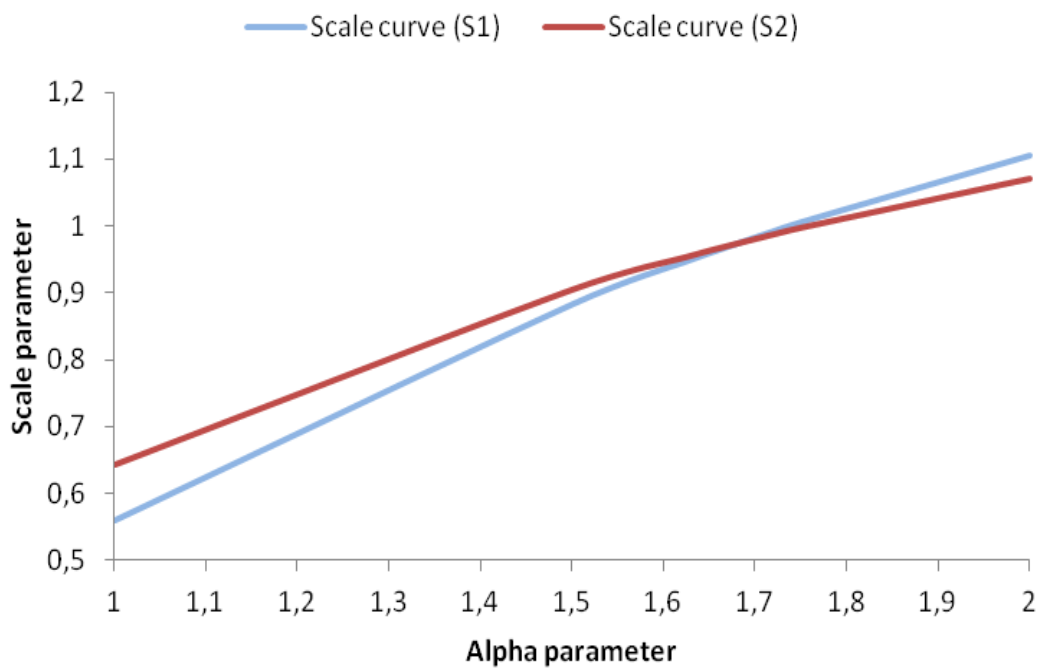
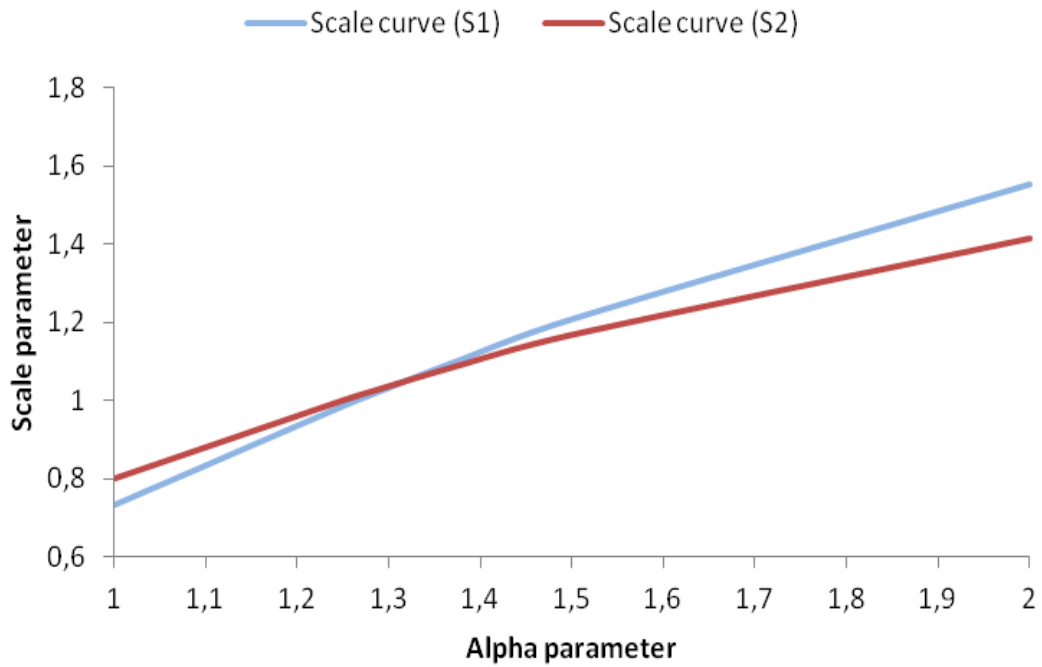
Az inicializálás után kiszámítjuk az intervallum középpontját, és kiszámítjuk az  $S_1$  és  $S_2$  becsléseket ebben a pontban. Meghatározzuk a következő iterációhoz az intervallum végpontjait, azaz úgy változtatjuk meg az intervallumot, hogy a metszéspont még mindig a vizsgált tartományban legyen. A pontosság tetszőleges  $\epsilon$ -ra beállítható, tehát ha az eltérés  $d = |a_{i-1} - a_i|$  két egymást követő iterációban kisebb, mint a meghatározott  $\epsilon$  pontosság érték, akkor megáll az iteráció. A minta  $\hat{\gamma}$  skála- és  $\hat{\delta}$  helyparamétere az utolsó  $\alpha$  iterációban meghatározásra került  $S_1$  és  $S_2$ , valamint  $T_1$  és  $T_2$  átlaga.

#### Algoritmus

1. Az  $\epsilon$  pontosság beállítása.
2. Inicializálás:  $a_0 = a_L = 1$  és  $a_1 = a_U = 2$
3. A  $S_1(a_L)$ ,  $S_2(a_L)$ ,  $S_1(a_U)$ ,  $S_2(a_U)$  kezdeti becslések kiszámítása.
4. Kezdeti feltétel ellenőrzése: ha  $S_1(a_L) < S_2(a_L)$  és  $S_2(a_U) < S_1(a_U)$  akkor van metszéspont, egyébként az algoritmus nem ad becslést  $\alpha$ -ra (kilépés -1).
5. While  $|a_{i-1} - a_i| > \epsilon$ 
  - $a_i := (a_U + a_L)/2$ ,  $S_1(a_i)$  és  $S_2(a_i)$  kiszámítása.
  - Ha  $S_1(a_i) < S_2(a_i)$ , akkor  $a_L := a_i$ , egyébként  $a_U := a_i$ .
6.  $\hat{\alpha} := a_i$
7.  $\hat{\gamma} := (S_1(a_i) + S_2(a_i))/2$  és  $\hat{\delta} := (T_{1,i} + T_{2,i})/2$

Az 4 - 7. ábrákon a PIT algoritmus futtatásával kapott skálaparaméter görbék szerepelnek különböző  $\alpha$  és  $\gamma$  paraméterű, változó mintaelemszámú véletlen minták esetén. A görbék metszéspontja mutatja a vízszintes tengelyen a becsült alakparamétert, a függőleges tengelyen pedig a becsült skálaparamétert. Az ábrák a futások során kapott skálaparaméter becsléssorozatokról készültek, tehát a kiszámított becslés közelében több pont található. A mintaelemszám növelésével –ahogy az várható– egyre pontosabb a paraméterek becslése, mind az alakparaméterre, mind a skálaparaméterre nézve.

Az 4. -6. ábrákon a véletlen minták skálaparaméterének értéke minden esetben  $\gamma = 1$ , és a helyparaméter értéke nulla ( $\delta = 0$ ). A minták elemszáma ( $n$ ) és az alakparaméter ( $\alpha$ ) az 4. ábrán



5. ábra. A ping-pong módszerrel meghatározott  $S_1(\alpha)$  és  $S_2(\alpha)$  skálaparaméter görbék  $n = 500$  elemből,  $\alpha = 1.3$  (felül) és  $\alpha = 1.7$ (alul) esetén

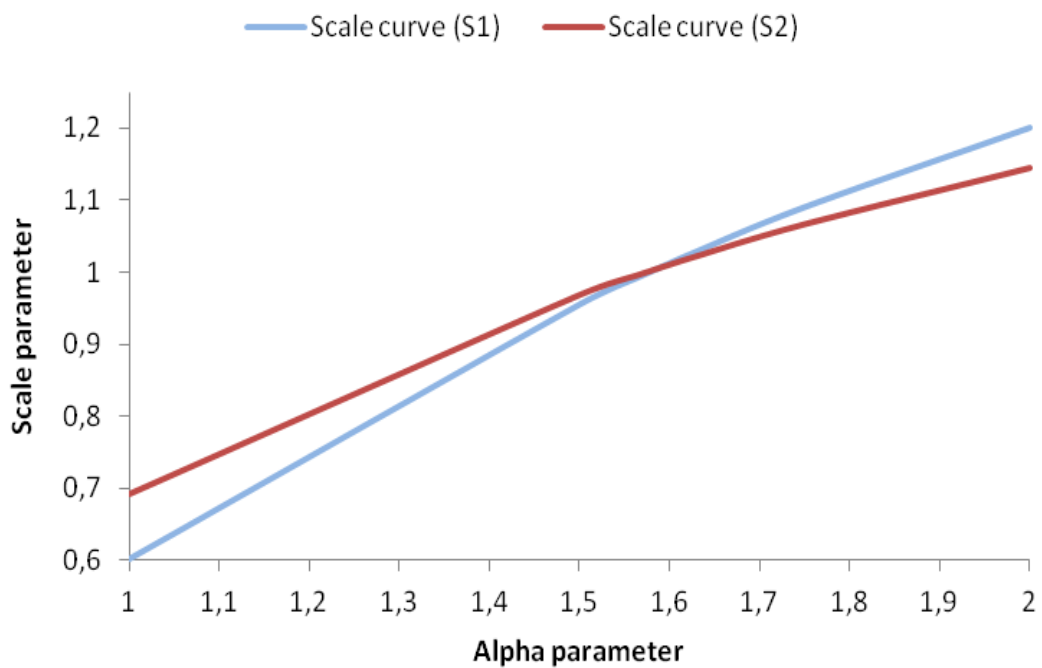
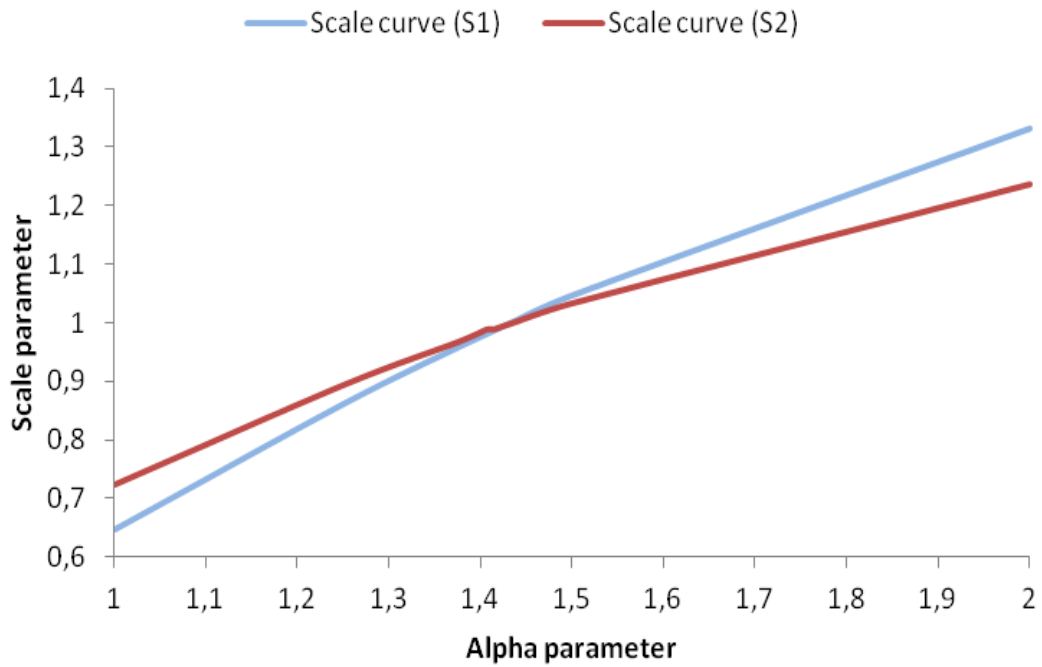
$n = 50$ , fent  $\alpha = 1.2$ , lent  $\alpha = 1.8$ . Az  $n = 50$  elemszám esetén még jelentős a véletlen becslésre gyakorolt hatása, jelentős az eltérés a véletlen minta paramétere és a becsült értékek között. Az 5. ábrán már nagyobb,  $n = 500$  elemű véletlen mintákra futtatva a becslési eljárást, pontosabb becslések adódtak, itt az alakparaméter értéke a fenti ábrán  $\alpha = 1.3$ , a lenti ábrán  $\alpha = 1.7$  volt.

A 6. ábrán a mintaelemszám  $n = 5000$ , és a generált minták alakparamétere a fenti ábrán  $\alpha = 1.4$ , a lenti ábrán  $\alpha = 1.6$  volt. A 7. ábrán nem standardizált mintára látható a kapott becsléssorozat, ekkor a generált minták  $n = 500$  eleműek voltak, az alakparaméter értéke a fenti és a lenti ábrán is egyaránt  $\alpha = 1.5$ , és a skálaparaméter értéke fent  $\gamma = 5$ , lent pedig  $\gamma = 0.5$ .

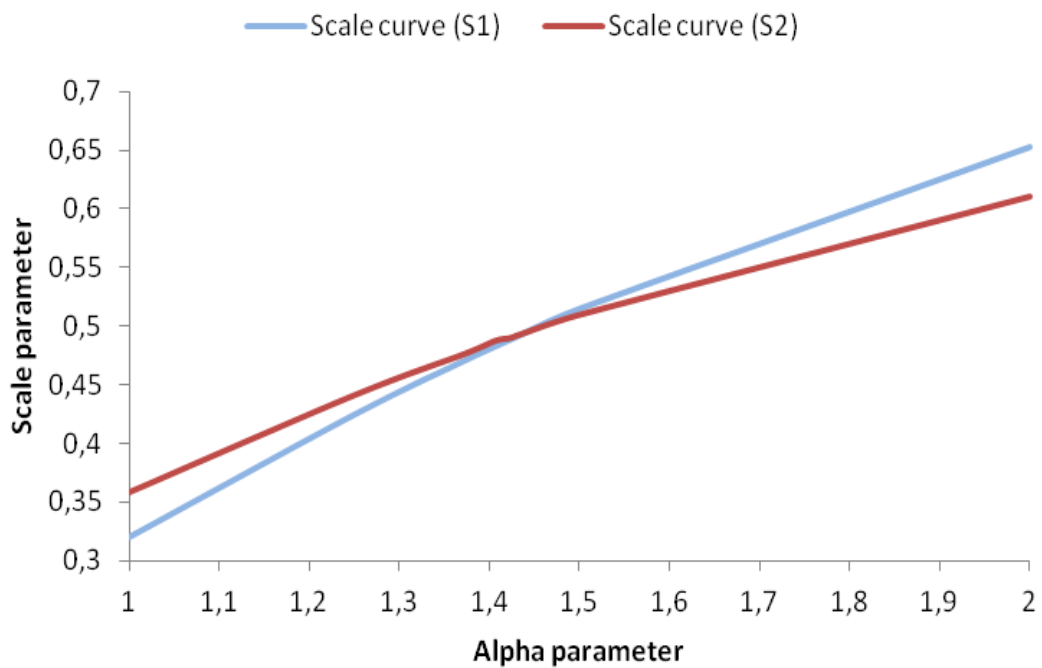
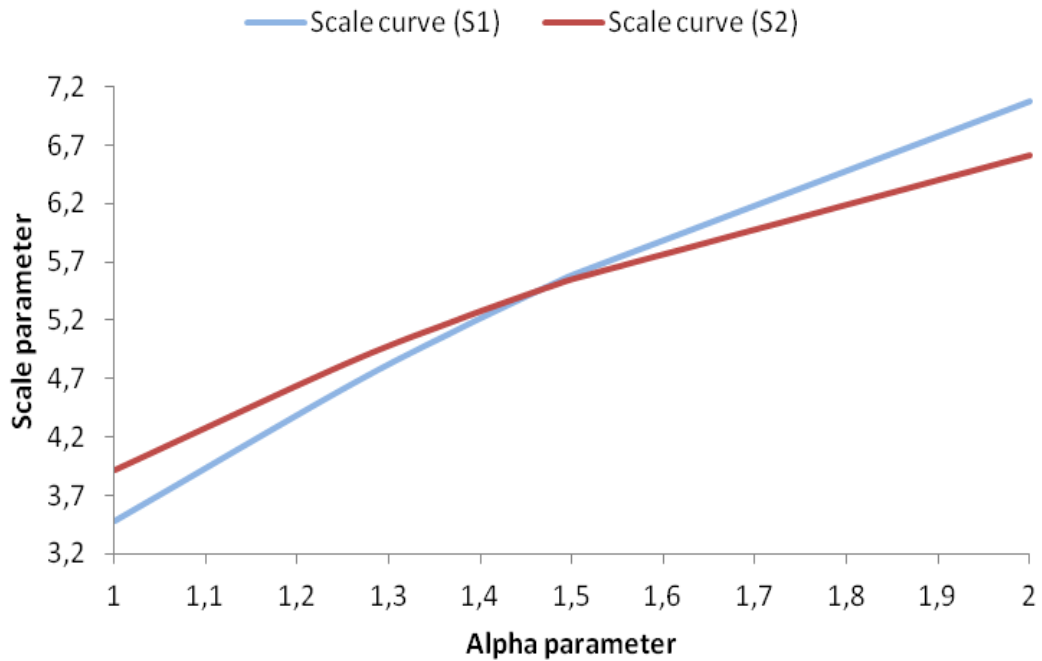
A bemutatott becslési eljárásnak vannak korlátai. A módszer szimmetrikus stabil eloszlások feltételezése mellett használható, mivel az eljárásban használt  $F_0$  típusok (a Cauchy és a normális eloszlás) szimmetrikus eloszlások. A paramétertér, amelyen az  $\alpha$  alakparamétert becsülni tudjuk az  $(1, 2)$  intervallum. Habár a  $\mathcal{B}$  függvények  $\alpha < 1$ -re is meghatározhatóak lennének numerikusan, erre az intervallumra nem számítottam ki a  $\mathcal{B}$  függvények közelítő értékeit. Mivel az  $\alpha \leq 1$  esetben a stabil eloszlás várható értéke nem véges, ez a tartomány a gyakorlat szempontjából kevésbé fontos, mint az  $1 < \alpha \leq 2$  tartomány. A gyakorlati alkalmazásokban, például a portfólió kiválasztási feladatban, ahol a hozamok eloszlásának várható értéke a befektetés várható hozamát adja meg, általában fontos, hogy létezzen a várható érték.

Egy másik probléma, amely főként kis mintaelemszámok esetén jelentkezhet, hogy a skálaparaméter görbék metszéspontja kívül esik az  $[1, 2]$  intervallumon. Ennek esélye nagyobb, hogyha az  $\alpha$  paraméter közel esik a végpontokhoz. Ez a probléma részletesebben is bemutatásra kerül a szimulációkról szóló 5. fejezetben. Ha a metszéspont kívül esik az intervallumon, akkor a módszer sajnos nem szolgáltat érvényes becslést az alakparaméterre.

A 5. fejezetben a módszer implementálásával, és az eredmények megbízhatóságával, pontosságával foglalkozom. A módszer szimulációs vizsgálataim alapján hasonló tulajdonságokkal (variancia, MSE érték) bír, mint a szakirodalomban ismert módszerek (Fama és Roll kvantiliseken alapuló eljárása, McCulloch módszere, a regressziós típusú becslés, és a momentumok módszerének változata). A javasolt eljárás használatának előnye, hogy könnyen implementálható, és az eredmény garantáltan a vizsgált paramétertérbe esik. A módszer futás időben nem használ numerikus integrálást, így gyorsabb lehet, mint a numerikus integrálást használó algoritmusok (pl. ML módszer).



6. ábra. A ping-pong módszerrel meghatározott  $S_1(\alpha)$  és  $S_2(\alpha)$  skálaparaméter görbék  $n = 5000$  elemből,  $\alpha = 1.4$  (felül) és  $\alpha = 1.6$  (alul) esetén



7. ábra. A ping-pong módszerrel meghatározott  $S_1(\alpha)$  és  $S_2(\alpha)$  skálaparaméter görbék  $n = 500$  elemből,  $\alpha = 1,5$ ,  $\gamma = 5$ (felül) és  $\gamma = 0,5$  (alul) esetén

#### 4.6. A $\mathcal{B}$ függvények közelítése

A PIT becslési eljárás használatával a szimmetrikus stabil eloszlású minta alak-, skála- és helyparamétere együttesen becsülhető nagy pontossággal. A becslések kiszámításához az 4.5. szakaszban bemutatott algoritmus alapján szükséges a  $\mathcal{B}$  függvények értékének megfelelő pontosságú ismerete tetszőleges  $1 \leq \alpha \leq 2$  pontban. A  $\mathcal{B}$  függvények egy-egy szórásnégyzetet jelenítenek meg,

$$\mathcal{B} = D_{F_\xi}^2(\psi(\xi)), \quad (47)$$

ahol

$$\psi(x) = F_0(x) - \frac{1}{2},$$

és  $F_\xi$  a mintának megfelelő stabil eloszlásfüggvény,  $F_0$  és  $F_\xi$  azonos típusú, és  $F_0$  az eloszlástípus kitüntetett tagja. A  $\mathcal{B}$  értéke  $1/12$ , ha a  $\xi$  minta eloszlásfüggvénye éppen  $F_0$ , egyébként pedig a függvények pontos értékét a (45) és (46) egyenletek alapján a

$$\mathcal{B}_1(\alpha) = \frac{1}{\pi} \int_{-\infty}^{\infty} \int_0^{\infty} \left(\frac{1}{\pi} \arctg x\right)^2 e^{-k^\alpha} \cos(kx) dk dx, \quad (48)$$

$$\mathcal{B}_2(\alpha) = \frac{1}{\pi} \int_{-\infty}^{\infty} \int_0^{\infty} \left(\Phi(x) - \frac{1}{2}\right)^2 e^{-k^\alpha} \cos(kx) dk dx, \quad (49)$$

formulák adják meg, ahol  $f_\alpha$  helyére a (6) szerint definiált stabil szimmetrikus sűrűségfüggvényt helyettesítettem. A (48) formula esetén a  $\psi$  függvényben szereplő  $F_0$  eloszlásfüggvény a standard Cauchy eloszlásfüggvény, míg a (49) formula esetén az  $F_0$  függvény a standard normális eloszlásfüggvény.

A (7) formula alapján, amely az általános szimmetrikus stabil sűrűségfüggvényének sorfejtéses reprezentációját tartalmazza a  $\mathcal{B}$  függvények alakja:

$$\mathcal{B}_1(\alpha) = \frac{1}{\pi\alpha} \int_{-\infty}^{\infty} \left(\frac{1}{\pi} \arctg x\right)^2 \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m)!} \Gamma\left(\frac{2m+1}{\alpha}\right) x^{2m} dx,$$

$$\mathcal{B}_2(\alpha) = \frac{1}{\pi\alpha} \int_{-\infty}^{\infty} \left(\Phi(x) - \frac{1}{2}\right)^2 \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m)!} \Gamma\left(\frac{2m+1}{\alpha}\right) x^{2m} dx.$$

A  $\mathcal{B}$  függvények tehát tartalmazzák az integrál alakban vagy sorfejtéses alakban ismert  $\alpha$ -stabil sűrűségfüggvényeket. Az 4.5. szakaszban bemutatott algoritmusbeli iterációk aktuális  $\alpha$  közelítéseinek megfelelő függvény értékeket a (48) és (49) képletek alapján numerikus integrálással kellene meghatározni, de ez jelentősen növelné a számítási igényt. A futási időben történő numerikus integrálás elkerülése, és ezáltal az algoritmus gyorsítása érdekében a  $\mathcal{B}_1$  és  $\mathcal{B}_2$  függvényeket racionális törtfüggvényekkel közelítettem.

A racionális törtfüggvény közelítés a legjobb egyenletes közelítések közül lényegesen kisebb hibával rendelkezik, mint például a polinom approximáció. A függvények legjobb egyenletes közelítésének meghatározáshoz legyen  $f \in C[a, b]$  a közelítendő függvény, amelyet egy  $F(x) = F_A(x) \in C[a, b]$  paraméteres függvényvel közelítünk, ahol  $A = [a_1, \dots, a_n]^T \in \Omega$ ,  $\Omega \subset \mathbb{R}^n$  adott paraméterhalmaz. A függvényközelítés jóságát az  $e(x) = f(x) - F_A(x)$  hibafüggvény normájával mérjük.

A legjobb függvényközelítés (approximáció) azt jelenti, hogy adott norma esetén keressük azt az  $A^*$  paramétervektort ( $F_{A^*}(x)$  közelítő függvényt), amelyre fennáll, hogy

$$\|f - F_{A^*}\| \leq \|f - F_A\|, \quad \forall A \in \Omega.$$

Az  $F_{A^*}(x)$  megoldást a legjobb approximációnak nevezzük. Ha a norma azonos a

$$\|g\|_C = \max_{x \in [a,b]} |g(x)|$$

Csebisev normával, akkor legjobb egyenletes közelítésről, vagy Csebisev-féle approximációról beszélünk.

Legyen  $R(m, n)$  mindazon  $r(x) = p(x)/q(x)$  alakú racionális törtfüggvények halmaza, ahol  $p(x) \in P_m$ ,  $q(x) \in P_n$ ,  $P_n$  jelöli a legfeljebb  $n$ -edfokú polinomok halmazát, és a  $p(x)$  és  $q(x)$  polinomoknak nincs közös zérushelyük. Az  $f \in C[a, b]$  függvények

$$r(x) = \frac{a_0 + a_1x + \dots + a_mx^m}{b_0 + b_1x + \dots + b_nx^n}$$

alakú legjobb approximációját keressük Csebisev normában.

**11. Tétel.** *Ha  $f \in C[a, b]$ , akkor létezik legjobb egyenletes approximáció, azaz olyan  $r^*(x) \in R(m, n)$  racionális törtfüggvény, hogy*

$$\|f - r^*\|_C \leq \|f - r\|_C \quad r(x) \in R(m, n)$$

**28. Definíció.** *Az  $a \leq x_1 < x_2 < \dots < x_N \leq b$  pontokat az  $f(x) - r(x)$  hibafüggvényre nézve alternáló pontoknak nevezzük, ha*

$$|f(x_j) - r(x_j)| = \|f - r\|_C \quad (j = 1, \dots, N)$$

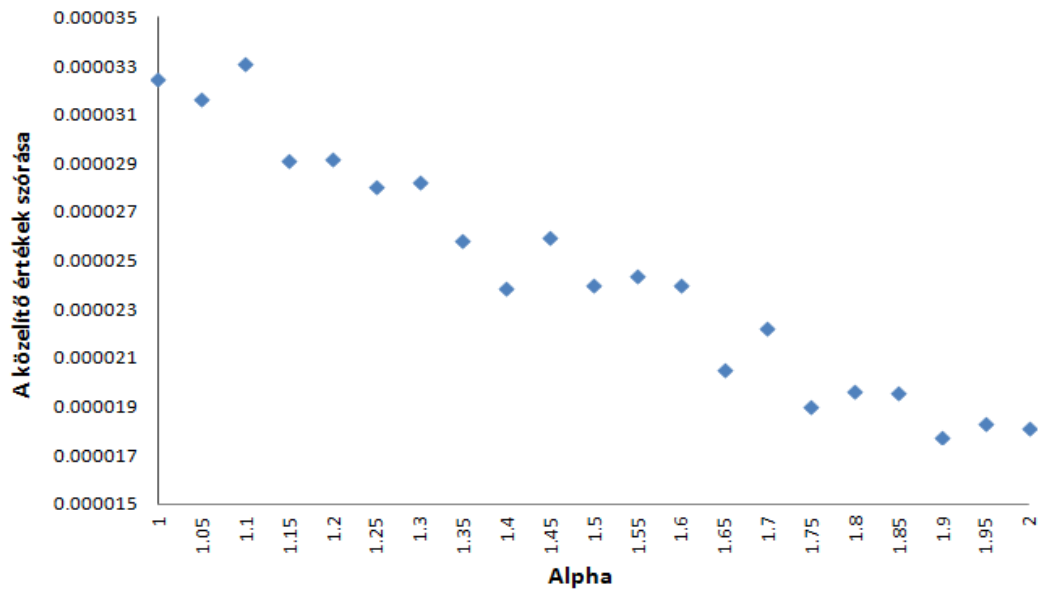
$$f(x_j) - r(x_j) = -[f(x_{j+1}) - r(x_{j+1})] \quad (j = 1, \dots, N - 1)$$

**12. Tétel.** *Legyen adott  $f(x) \in C[a, b]$ . Az  $r(x) = p(x)/q(x) \in R(m, n)$  racionális törtfüggvény akkor és csak akkor az  $f$  függvény legjobb egyenletes approximációja, ha az  $f(x) - r(x)$  függvénynek létezik egy  $N = 2 + \max\{n + \partial p, m + \partial q\}$  pontból álló alternáló ponthalmaza ( $\partial p$  a  $p(x)$  fokszáma,  $\partial q$  a  $q(x)$  fokszáma. )*

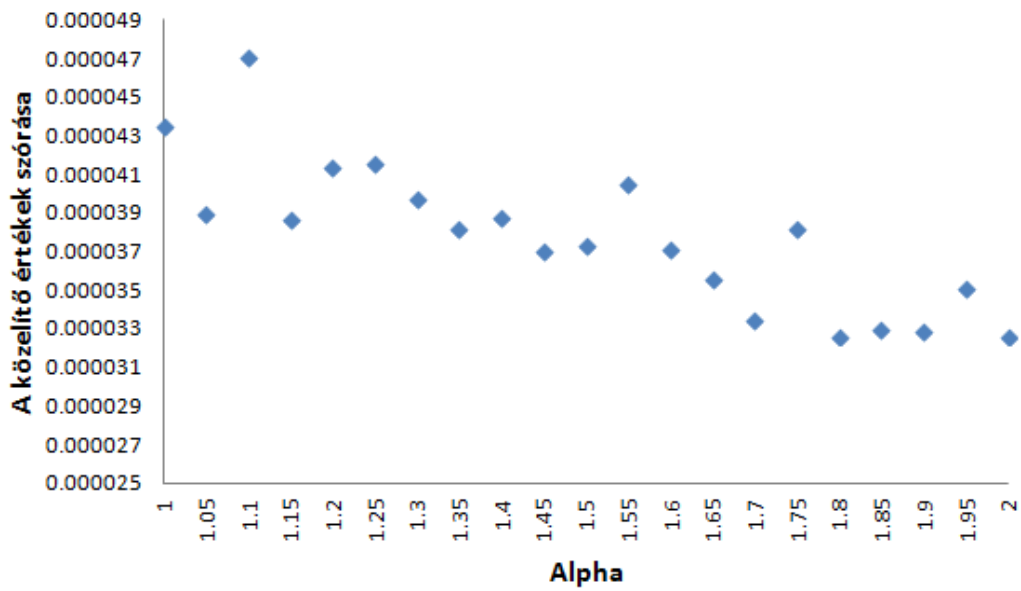
**13. Tétel.** *A legjobb egyenletes racionális approximáció egyértelmű.*

A függvények Csebisev-féle legjobb közelítésének elmélete, az itt megadott tételek bizonyításai megtalálhatók például Kincaid és Cheney [37] könyvében.

A következő szakaszban a  $\mathcal{B}$  függvények legmegfelelőbb racionális törtfüggvény közelítései meghatározásának lépéseit ismertetem. A közelítést két lépésben határoztam meg. Először a függvények kiválasztott alappontokban felvett értékeit közelítettem véletlen minták segítségével, majd a függvényértékekre támaszkodva meghatároztam a racionális törtfüggvényeket. Az alappontokat és a törtfüggvény közelítés fokszámát ( $m$  és  $n$ ) magam választottam ki, és a függvényértékeket is csak közelíteni tudjuk, a kapott racionális törtfüggvények nem az egyenletesen legjobb racionális approximációt jelentik, csak egy azt megközelítő approximációt.



8. ábra. A közelítő függvényértékek szórása a  $\mathcal{B}_1$  függvény esetén



9. ábra. A közelítő függvényértékek szórása a  $\mathcal{B}_2$  függvény esetén



A közelítendő függvények választott alappontokban felvett függvényértékeinek kiszámítását a nagy számok törvénye teszi lehetővé, azaz a függvényértékek közelíthetők az átlaggal. Az  $F_\alpha$  mérték szerinti integrálás adott  $\alpha$ -stabil eloszlású véletlen minta generálásával, és a generált mintaelemek helyettesítésével határozható meg:

$$\mathcal{B}_1(\alpha) \approx \frac{1}{n} \sum_{i=1}^n (\Phi(x_i) - 0.5)^2, \quad (50)$$

$$\mathcal{B}_2(\alpha) \approx \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{\pi} \arctg x_i \right)^2. \quad (51)$$

A racionális törtfüggvényes közelítés meghatározásához az  $\alpha = 1, 1.05, 1.1, \dots, 1.95, 2$  értékeket választottam, azaz összesen 21 alappontot vettem fel. A választott alappontokban a (50) és (51) formulák alapján 5 millió elemből álló  $\alpha$ -stabil véletlen minták segítségével közelítettem a  $\mathcal{B}_1$  és  $\mathcal{B}_2$  függvényeket. Az 5 millió elemből álló mintákból készített közelítést 200-szor ismételttem meg, és az ismétlésekből kapott értékeket átlagoltam. Tehát ezzel a módszerrel összesen 1 milliárd véletlenszámból határoztam meg a függvények alappontokban felvett értékeit. A véletlenszámok generálását a (60) formula alapján végeztem.

A közelítés meghatározásához MATLAB függvényeket készítettem. Az .m fájlokat részletebben a dolgozatom 7. fejezetében ismertetem. A függvényértékek kiszámításánál a normális eloszlás eloszlásfüggvényének értékét a MATLAB-ban rendelkezésre álló (beépített) függvénnyel kaptam meg.

Az alappontokban érvényes közelítő függvényértékek pontosságáról információval szolgálhat az  $\alpha = 1$  alappontban kapott  $\mathcal{B}_1(1)$ , valamint a  $\alpha = 2$  alappontban kapott  $\mathcal{B}_2(2)$  érték, ugyanis ezekben a pontokban a függvény értékei pontosan  $1/12$ -del egyenlőek. Az intervallum másik két végpontjában is meghatározható a függvényérték nagy pontossággal (a MAPLE 15 szoftverrendszer segítségével), amelyek  $\mathcal{B}_1(2) = 0.04556423$ , valamint  $\mathcal{B}_2(1) = 0.12680134$ . A szimulációval meghatározott közelítő értékek a nagy pontosságú értékekkel összehasonlítva 5 tizedesjegy ( $10^{-5}$ ) pontosságúak.

A 200 ismétlésből kapott közelítő értékek szórása a  $\mathcal{B}_1$  függvény esetében  $0.00003$ -ról ( $\alpha = 1$ )  $0.00001$ -re ( $\alpha = 2$ ), míg a  $\mathcal{B}_2$  függvény esetében  $0.00004$ -ről  $0.00003$ -ra csökkent. A csökkenés mind a két függvénynél közel lineáris jellegű, a  $\mathcal{B}_2$  függvénynél egy kicsit talán lassabb ütemű. A szórások alakulását az  $\alpha$  paraméter függvényében mutatja a 8. és a 9. ábra. Az 1. táblázat tartalmazza a két függvény közelített értékeit az alappontokban.

A 10. és 11. ábrákon a közelített függvények láthatók. A közelített függvények jellege miatt a törtfüggvényes közelítés fokszámának megválasztásakor érdemes a számlálót eggyel magasabb fokú polinomnak választani, mint a nevezőt. Ezzel egy olyan kifejezést kapunk, amelynek képe hasonlóan a közelítendő függvényekhez, lineárishoz hasonló. A legmegfelelőbb racionális törtfüggvényt különböző fokszámú polinom párok használatával kerestem.

1. táblázat. A  $\mathcal{B}_1$  és  $\mathcal{B}_2$  függvények értékei az  $\alpha = 1, 1.05, \dots, 2$  alappontokban

$\alpha$	$\mathcal{B}_1(\alpha)$	$\mathcal{B}_2(\alpha)$
1.00	0.08333268	0.12680106
1.05	0.07941931	0.12271293
1.10	0.07588209	0.11895612
1.15	0.07267831	0.11548698
1.20	0.06976327	0.11227325
1.25	0.06711412	0.10931583
1.30	0.06469494	0.10657474
1.35	0.06248928	0.10403059
1.40	0.06046750	0.10167604
1.45	0.05860996	0.09948424
1.50	0.05690556	0.09745172
1.55	0.05533790	0.09556008
1.60	0.05389250	0.09379953
1.65	0.05255457	0.09215979
1.70	0.05132104	0.09062897
1.75	0.05017641	0.08919795
1.80	0.04911322	0.08786609
1.85	0.04812502	0.08661367
1.90	0.04721128	0.08544831
1.95	0.04636014	0.08436018
2.00	0.04556340	0.08333157

A törtfüggvények, amelyeket vizsgáltam az  $R(5, 4)$ ,  $R(4, 3)$ ,  $R(3, 2)$ ,  $R(2, 1)$  voltak, azaz

$$\mathcal{B}_{i,A}(x) = \frac{a_5x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0}{x^4 + b_3x^3 + b_2x^2 + b_1x + b_0}, \quad (52)$$

$$\mathcal{B}_{i,B}(x) = \frac{a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0}{x^3 + b_2x^2 + b_1x + b_0}, \quad (53)$$

$$\mathcal{B}_{i,C}(x) = \frac{a_3x^3 + a_2x^2 + a_1x + a_0}{x^2 + b_1x + b_0}, \quad (54)$$

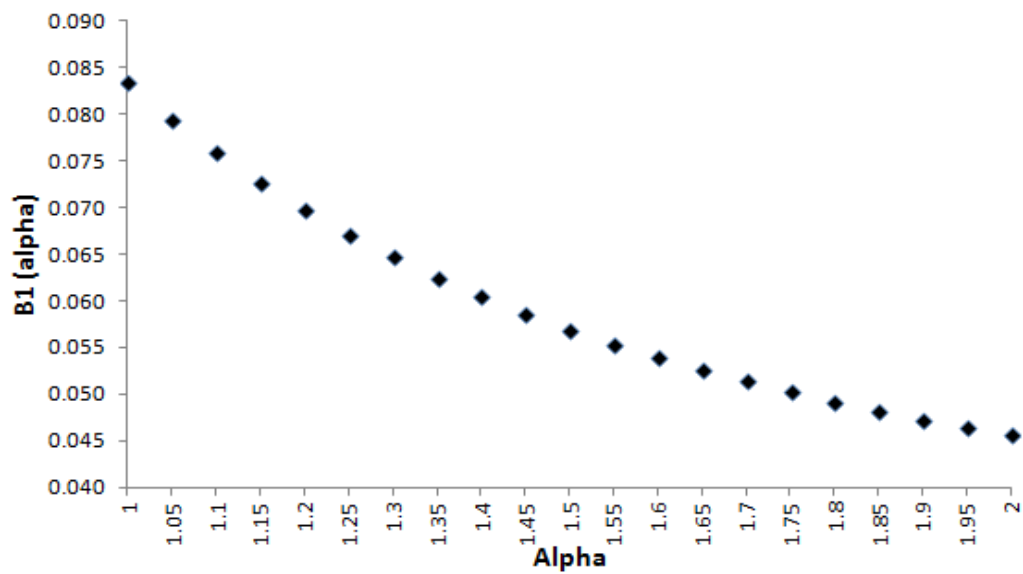
$$\mathcal{B}_{i,D}(x) = \frac{a_2x^2 + a_1x + a_0}{x + b_0}, \quad (55)$$

$$(56)$$

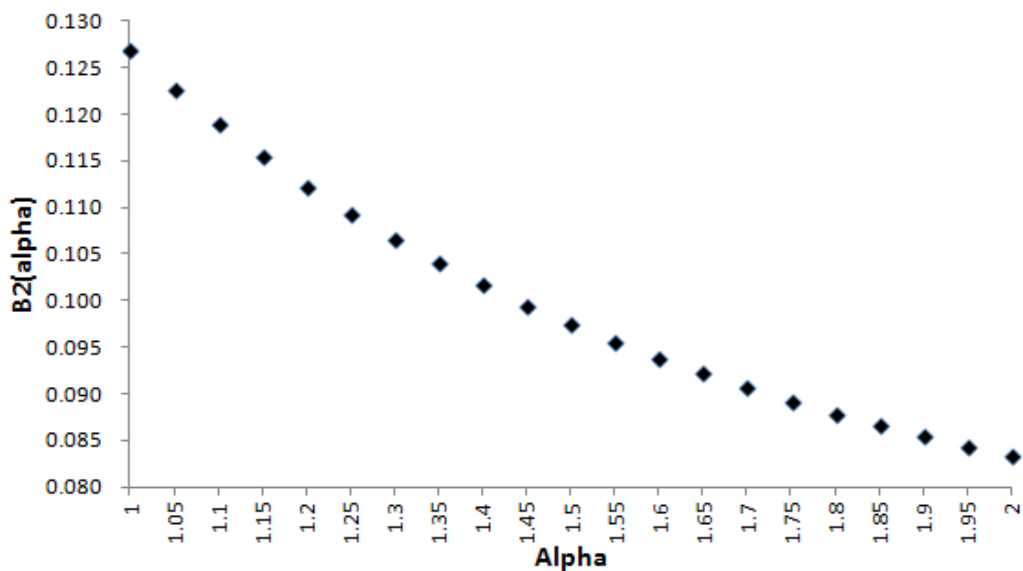
ahol  $i = 1, 2$ , és  $A, B, C$  és  $D$  jelöli a vizsgált törtfüggvény közelítéseket.

A törtfüggvény közelítés alakja általánosan:

$$\mathcal{B}_i(x) = \frac{\sum_{j=0}^m a_j x^j}{\sum_{k=0}^{n-1} b_k x^k + x^n} \quad (57)$$



10. ábra. A  $\mathcal{B}_1(\alpha)$  függvény közelített értékei



11. ábra. A  $\mathcal{B}_2(\alpha)$  függvény közelített értékei

Az alappontok függvényértékeinek ismeretében az ismeretlen  $a_5, \dots, a_0, b_3, \dots, b_0$  együtthatók meghatározhatók. Jelölje  $\alpha_i$  a kiválasztott alappontokat,  $i \in \mathbb{Z}$ . A (57) formula alapján

$$\mathcal{B}(\alpha_i) \left( \sum_{k=0}^{n-1} b_k(\alpha_i)^k + (\alpha_i)^n \right) = \left( \sum_{j=0}^m a_j(\alpha_i)^j \right), \quad (58)$$

$$\left( \sum_{j=0}^m a_j(\alpha_i)^j \right) - \mathcal{B}(\alpha_i) \left( \sum_{k=0}^{n-1} b_k(\alpha_i)^k \right) = \alpha_i^n \mathcal{B}(\alpha_i). \quad (59)$$

Tehát összesen 21 lineáris egyenletet készíthetünk, amelyekben az  $a_j$  és  $b_k$  együtthatók az ismeretlenek. Az egyenletrendszer egyenleteit tehát úgy kapjuk, ha valamilyen módon kiválasztunk bizonyosakat a 21 alappont közül. Az alappontok kiválasztásánál fontos, hogy a kiválasztott ponthalmaz tartalmazza az intervallum végpontjait, és a közbeeső pontok nagyjából egymástól azonos távolságra helyezkedjenek el. A választott pontok számától függően a megoldandó lineáris egyenletrendszer túlhatározott (ha több alappontot választunk, mint ahány ismeretlen együtthatót szeretnénk kiszámolni), vagy olyan egyenletrendszer, amely négyzetes együtthatómátrixszal rendelkezik (az alappontok és ismeretlenek száma megegyezik).

2. táblázat. A legmegfelelőbb törtfüggvény közelítés keresése során vizsgált fokszámok és alappontok

sorszám	$m$	$n$	i. sz.	a.sz.	alappontok
1	5	4	10	10	1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2
2	5	4	10	11	1, 1.05, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2
3	5	4	10	12	1, 1.05, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 1.95, 2
4	4	3	8	8	1, 1.1, 1.2, 1.4, 1.6, 1.8, 1.9, 2
5	4	3	8	9	1, 1.1, 1.2, 1.35, 1.5, 1.65, 1.8, 1.9, 2
6	4	3	8	10	1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2
7	3	2	6	6	1, 1.2, 1.4, 1.6, 1.8, 2
8	3	2	6	7	1, 1.1, 1.3, 1.5, 1.7, 1.9, 2
9	3	2	6	8	1, 1.1, 1.2, 1.4, 1.6, 1.8, 1.9, 2
10	2	1	4	4	1, 1.35, 1.65, 2
11	2	1	4	5	1, 1.25, 1.5, 1.75, 2
12	2	1	4	6	1, 1.2, 1.4, 1.6, 1.8, 2

Meghatároztam a fenti (52), (53), (54), (55) törtfüggvény közelítéseket a MATLAB szoftvercsomag beépített lineáris egyenletrendszer megoldó *linsolve* programjával három különböző alappont halmaz kiválasztásával. A három kipróbált eset, amikor az ismeretlen együtthatók számával pontosan megegyező, eggyel több, és kettővel több alappontot választottam. A kiválasztott alappontokat, és a törtfüggvény együtthatók számát foglalja össze a 2. táblázat. Az i.sz. rövidítés az ismeretlenek számát, a.sz. az alappontok számát jelenti,  $m$  és  $n$  a törtfüggvény számlálójának és nevezőjének a fokszáma.

A törtfüggvény közelítéseket tehát mind a 12 esetben meghatároztam a *linsolve* függvénnyel. Túlhatározott egyenletrendszer esetén a *linsolve* MATLAB függvény a legkisebb négyzetes értelemben legjobban illeszkedő megoldást választja. Ezután megkerestem a kipróbált esetek közül az alappontbeli értékekhez legjobban illeszkedő közelítést, a közelítés hibája alapján. A közelítés hibáját a szimulált alappontbeli függvényértékek, és a közelített függvényértékek közötti legnagyobb eltérés abszolút értékével mértem. A közelítések hibáit a 3. táblázat tartalmazza.

3. táblázat. A közelítések legnagyobb abszolút eltérései a szimulált függvényértékektől

sorszám	$\mathcal{B}_1$	$\mathcal{B}_2$
1	1.065136e-005	9.320114e-006
2	7.191584e-006	3.105706e-005
3	1.507770e-005	2.549295e-005
4	3.062506e-006	1.215120e-005
5	5.371650e-006	8.040216e-005
6	3.767749e-006	1.911747e-005
7	3.223983e-006	1.125222e-005
8	3.523025e-006	7.966478e-006
9	2.921388e-006	1.491678e-005
10	3.532933e-005	3.033719e-005
11	2.894798e-005	2.190907e-005
12	2.513574e-005	2.337243e-005

Megállapítottam, hogy az egyes közelítések maximális eltérései nagyon közel vannak egymáshoz, azt is mondhatjuk, hogy nincs közöttük számottevő különbség. Mivel az alappontokban felvett függvényértékek közelítésénél 5 tizedesjegyig sikerült pontosan a szimuláció, ezért az is lehetséges, hogy mindegyik közelítés megfelelő pontosságú.

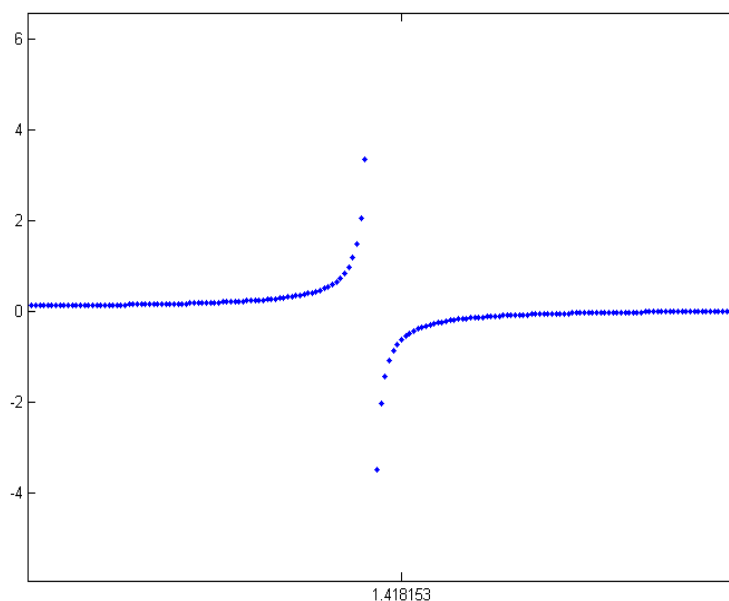
A kipróbált törtfüggvény közelítések pontosságának összehasonlítása során azt állapítottam meg, hogy azoknál a közelítéseknél, ahol az egyenletrendszernek pontos megoldása van, ott a felhasznált alappontokban  $10^{-17}$ , a közbeeső pontoknál viszont  $10^{-5}$  pontosságú a közelítés. A *linsolve* függvénnyel legkisebb négyzetes értelemben kiszámolt túlhatározott megoldásoknál viszont egyenletesebbek az eltérések, tehát nem annyira pontos a felhasznált alappontokban, viszont a közbeeső pontoknál a pontos megoldással rendelkező esetenél kisebb az eltérés, ekkor  $10^{-7} - 10^{-9}$  pontosságokat kaptam. Az eltéréseket minden közelítésnél meghatároztam, de terjedelmi okokból ezeket a táblázatokat nem részletezem.

A legmegfelelőbb törtfüggvény kiválasztásánál arra is ügyelnem kellett, hogy se a számláló, se a nevezőnek ne legyen zérushelye az  $[1, 2]$  intervallumon. (A PIT módszer pontosságának vizsgálatára készített nagy ismétlésszámú szimulációknál jelentkezett az a probléma, hogy pontatlan, akár negatív közelítést eredményezett a törtfüggvényes approximáció.) A problémát illusztrálja a 12. ábra, amely a  $\mathcal{B}_1$  függvény első ( $m = 5, n = 4$ ) közelítésével kiszámolt függvényértékeit mutatja. Ennél a törtfüggvény közelítésnél a számlálóra kapott polinomnak  $x = 1.4181$ -nél zérushelye van, ami miatt ez a közelítés nem használható, hiszen a zérushely környezetében nem ad pontos értékeket.

A közelítések polinomjainak zérushelyeit az összes törtfüggvényre a MAPLE program segítségével vizsgáltam, és olyan törtfüggvényt választottam, amelynél ez a numerikus hiba nem jelentkezik, a közelítő függvénynek nincsen zérushelye az  $[1, 2]$  intervallumban. A közelítések közül az  $R(3, 2)$  és  $R(2, 1)$  fokszámú (7-12. számú) esetek megfelelőek ebből a szempontból.

A törtfüggvény közelítések vizsgálatának eredménye alapján a 4. táblázatban közölt törtfüggvény közelítést választottam, azaz a 8. sorszámú approximáció adódott a legmegfelelőbbnek.

**A  $\mathcal{B}$  függvények törtfüggvényes közelítésének meghatározása jelentős számítási munkával járt. Az egyes részfeladatok elvégzése, például az alappontokban érvényes függ-**



12. ábra. A  $\mathcal{B}_1(\alpha)$  hibás közelítő függvénye, ( $m = 5, n = 4$ ) eset

4. táblázat. A  $\mathcal{B}_1$  és  $\mathcal{B}_2$  legmegfelelőbb racionális törtfüggvény közelítésének együtthatói

Együtthatók	$\mathcal{B}_1$	$\mathcal{B}_2$
$a_3$	0.00343013	0.00631315
$a_2$	0.00605670	0.01943904
$a_1$	0.04709978	0.09332481
$a_0$	0.00972618	0.01619877
$b_1$	-0.38087590	-0.09345095
$b_0$	0.17663917	0.16029569

vényértékek meghatározása a milliárdos nagyságrendű véletlenszám generálás miatt alapponként önmagában több órát vett igénybe. Ugyanakkor, a közelítések meghatározása révén az algoritmus implementálhatósága, használhatósága jelentősen egyszerűsödött. A törtfüggvény közelítés használata egyszerű felhasználó számára is lehetővé teszi a becslési eljárás alkalmazását.

A törtfüggvény közelítés használatával a PIT becslési eljárás gyorsítható és egyszerűsíthető, mivel nem szükséges az egyébként zárt alakban ismeretlen stabil sűrűségfüggvény és eloszlásfüggvény közvetlen, numerikus integrálással történő számítása futási időben.

## 5. Statisztikai vizsgálatok

### 5.1. Stabil eloszlású véletlen számok generálása

A stabil eloszlások általában csak a karakterisztikus függvény segítségével írhatók le, ezért az alkalmazásokhoz, a paramétereik becsléséhez számos esetben közelítő eljárásokra, módszerekre van szükség, amelyekhez a Monte-Carlo módszereket alkalmazzuk. Ezért a különféle statisztikai vizsgálatokhoz elengedhetetlen, hogy tetszőleges paraméterű stabil eloszlású változót generálhassunk. Általában erre a feladatra az eloszlásfüggvény inverzén alapuló klasszikus technikát használjuk, amelynek lényege, hogy egy egyenletes eloszlású változót helyettesítünk a generálni kívánt változó eloszlásfüggvényének inverzébe, amely így a kívánt eloszlású változót fogja eredményezni. Ha adott egy  $\xi$  folytonos valószínűségi változó  $F_\xi$  invertálható eloszlásfüggvénnyel, akkor az  $Y = F(\xi)$  valószínűségi változó egyenletes eloszlású a  $(0, 1]$  intervallumon. Az inverz transzformáció (inverse probability integral transformation) alapján pedig az  $F_\xi^{-1}(Y)$  valószínűségi változó eloszlásfüggvénye éppen  $F_\xi$ .

Stabil eloszlások esetén nem használható a klasszikus inverz függvény módszeren alapuló véletlenszám generálási technika, mivel nem ismert az általános eloszlásfüggvény, és így az eloszlásfüggvény inverze sem zárt alakban. A problémára általános esetben (tetszőleges  $\alpha, \beta, \gamma, \delta$  paraméter esetén) Chambers et al. [8] adtak 1976-ban először formulát. Szintén az általános esetre használható képletet adott meg Zolotarev [84]. Weron [82] tételeket adott a különböző parametrizációjú karakterisztikus függvényekkel megadott változók generálásához használható formulákról.

Weron [82] dolgozatának 3.1-es tétele alapján az  $\alpha = 1$ -nél folytonos parametrizáció esetén az általános módszer a következő<sup>7</sup>.

Legyen

$$K(\alpha) = \alpha - 1 + \text{sign}(1 - \alpha) = \begin{cases} \alpha, & \alpha < 1, \\ \alpha - 2, & \alpha > 1, \end{cases}$$

és

$$\gamma_0 = -\frac{\pi}{2}\beta_2 \frac{K(\alpha)}{\alpha}.$$

Legyen  $\xi$  egyenletes eloszlású változó a  $(-\frac{\pi}{2}, \frac{\pi}{2})$  intervallumon, és legyen  $\eta$  független, exponenciális valószínűségi változó 1 várható értékkel. Ekkor

- ha  $\alpha \neq 1$ , akkor  $X \sim S(\alpha, \beta_2, 1, 0)$

$$X = \frac{\sin(\alpha(\xi - \gamma_0))}{(\cos\xi)^{\frac{1}{\alpha}}} \left( \frac{\cos(\xi - \alpha(\xi - \gamma_0))}{\eta} \right)^{\frac{1-\alpha}{\alpha}},$$

- ha  $\alpha = 1$ , akkor  $X \sim S(1, \beta_2, 1, 0)$

$$X = \left( \frac{\pi}{2} + \beta_2 \xi \right) \tan \xi - \beta_2 \log \left( \frac{\eta \cos(\xi)}{\frac{\pi}{2} + \beta_2 \xi} \right)$$

<sup>7</sup>A  $\beta_2$  jelölés arra vonatkozik, hogy ebben a parametrizációban más értéket vesz fel a  $\beta$  aszimmetria paraméter is. Az  $\alpha = 1$  paraméterértéknél folytonos és nem folytonos parametrizáció szimmetria paramétereinek összefüggése megtalálható Weron [82] cikkben.

A statisztikai vizsgálatokhoz Zolotarev [84] formulája alapján generáltam stabil eloszlású véletlen mintákat. Eszerint a formula szerint egy általános  $\alpha$  - stabil, szimmetrikus  $Z \sim S(\alpha, 0, 1, 0)$  változót megkaphatunk a következőként:

$$Z = \frac{\sin(\alpha\xi)}{(\cos\xi)^{\frac{1}{\alpha}}} \left( \frac{\cos((1-\alpha)\xi)}{\eta} \right)^{\frac{1-\alpha}{\alpha}}, \quad (60)$$

ahol  $\eta$  standard exponenciális,  $\xi$  egyenletes valószínűségi változó a  $(-\pi/2, \pi/2)$  intervallumon. Standard stabil változónak azt a valószínűségi változót tekintjük, amely a fenti változóból

$$Z/\alpha^{\frac{1}{\alpha}}$$

képlettel generálható. Az így kapott valószínűségi változó esetében a skálaparaméter értéke  $\gamma = 1$  lesz.

A nem standardizált  $S(\alpha, 0, \gamma, \delta)$  eloszlású változót a  $\delta$  hely- és  $\gamma$  skálaparaméter segítségével  $\gamma Z + \delta$ -ként kaphatjuk meg, hasonlóan a  $(\sigma, \mu)$  paraméterű normális eloszlás generálásához, mivel a stabil eloszlások esetén is használhatjuk az azonos típusú eloszlásfüggvénnyel rendelkező eloszlásokra vonatkozó tulajdonságot.

Az egyenletes eloszlású valószínűségi változók generálását minden statisztikai és magasabb szintű általános célú programnyelv lehetővé teszi. Ezek a programcsomagok általában beépített függvényként tartalmazzák a szükséges rutint. Az exponenciális változó generálásához használható a már említett inverz függvényes előállítás. Ha  $\eta$  exponenciális eloszlású változó, azaz

$$F_{\eta}(x) = 1 - e^{-x},$$

és  $U$  egyenletes eloszlású változó a  $[0, 1]$  intervallumon, akkor  $\eta$  generálható

$$\eta = -\ln U$$

alapján.

A normális eloszlású változók generálásának széles körű szakirodalma van, és a statisztikai programcsomagok tartalmazzák az egyenletes eloszlást generáló rutin mellett általában normális eloszlást generáló beépített rutint is. Az

$$f_{Cauchy} = \frac{1}{\pi} \frac{1}{1+x^2}$$

sűrűségfüggvényű Cauchy eloszlás generálható

$$Y = \operatorname{tg}\left(\left(U - \frac{1}{2}\right)\pi\right)$$

formula alapján, ahol  $U \in U(0, 1)$ .

Többváltozós stabil eloszlású minták generálása az egyváltozós esettől sokkal összetettebb probléma. Ebben az esetben a változók közötti függési struktúrát a spektrálmérték definiálja. A spektrálmérték becslése nagyon nehéz feladat, és általános megoldás még nem is született a problémára. Véletlen számok generálása is csak bizonyos esetekben, stabil eloszlások egy csoportjára lehetséges.

Viszonylag egyszerűen generálhatunk többváltozós véletlen vektort akkor, ha a spektrálmérték diszkrét mérték. A generálás számítási igénye függ az  $\alpha$ -tól, a pontok, a tömegek és a pont



tömegek szóródásától. Nolan két dimenziós sűrűségfüggvény kiszámításához készített szotvert (MVSTABLE), amely letölthető a [61] weboldalról. Magasabb dimenzióban a szimuláció számítási igénye alapvetően a pont tömegek lineáris függvénye, Nolan [57].

A szimulációs módszer Modarres és Nolan [54] eredményén alapul. Ha  $X$  karakterisztikus függvénye a (13) formula szerinti, akkor

$$X \stackrel{D}{=} \begin{cases} \sum_{j=1}^n \gamma_j^{1/\alpha} Z_j s_j, & \alpha \neq 1, \\ \sum_{j=1}^n \gamma_j (Z_j + \frac{2}{\pi} \ln \gamma_j) s_j, & \alpha = 1, \end{cases}$$

ahol  $Z_1, \dots, Z_n$  független azonos eloszlásúak, teljesen ferde standardizált egyváltozós  $\alpha$  stabil valószínűségi változók, vagyis  $Z_i \sim S_\alpha(1, 1, 0)$ .

Nolan [58] bemutat stabil eloszlások egy csoportjára vonatkozó eredményeket is, amelyekkel többváltozós eloszlások szimulálhatók. Ezek a radiál szimmetrikus (sugarasan szimmetrikus) vagy izotróp eloszlások, valamint az elliptikus eloszlások.

**29. Definíció.** Ha  $X$   $\alpha$ -stabil és radiál szimmetrikus vagy izotróp, akkor  $X$  karakterisztikus függvénye

$$E \exp(i\langle u, X \rangle) = \exp(-\gamma_0^\alpha |u|^\alpha + i\langle u, \delta \rangle) \quad (61)$$

és a projekciós paraméterek  $\gamma(u) = \gamma_0 |u|^\alpha$ ,  $\beta(u) = 0$ ,  $\delta(u) = \langle u, \delta \rangle$ . A spektrálmérték ebben az esetben egyenletes eloszlás  $\mathbb{S}$ -en.

**30. Definíció.** Ha  $X$   $\alpha$ -stabil és elliptikus (elliptically contoured), akkor az együttes karakterisztikus függvény

$$E \exp(i\langle u, X \rangle) = \exp(-(u^T \Sigma u)^{\alpha/2} + i\langle u, \delta \rangle) \quad (62)$$

és a projekciós paraméterek  $\gamma(u) = (u^T \Sigma u)^{\alpha/2}$ ,  $\beta(u) = 0$ ,  $\delta(u) = \langle u, \delta \rangle$ .  $\Sigma$  valamilyen pozitív definit mátrix,  $\delta \in \mathbb{R}^d$  eltolás vektor.

Feltételezzük, hogy  $X$  nonsinguláris, vagyis  $\Sigma$  szigorúan pozitív definit:  $\forall u \neq 0 : u^T \Sigma u > 0$ . Az elliptikus stabil eloszlások többváltozós normális eloszlások skálázási keverékei (Samorodnitsky és Taqqu [71]).

Legyen

$$A \sim S(\alpha/2, 1, \gamma, 0)$$

független egyváltozós pozitív  $\alpha/2$ -stabil véletlen vektor,  $0 < \alpha < 2$  és  $G \sim N(0, \Sigma)$   $d$ -dimenziós normális eloszlás.

Ekkor  $A^{1/2}G$   $\alpha$ -stabil elliptikus, és a karakterisztikus függvénye

$$\exp(-(\gamma/2)^{\alpha/2} (\sec \pi\alpha/4) (u^T \Sigma u)^{\alpha/2}).$$

Az elliptikus stabil eloszlást sub-Gauss stabil eloszlásnak is nevezik. Ez a felírás lehetővé teszi a szimulációt. Legyen  $0 < \alpha < 2$ ,

$$A \sim S(\alpha/2, 1, 2\gamma_0^2 (\cos \pi\alpha/4)^{2/\alpha}, 0), \quad (63)$$

és  $G \sim N(0, \Sigma)$ , ekkor  $X = A^{1/2}G + \delta$  karakterisztikus függvénye a (62) formula szerint alakul.

A radiál szimmetrikus eset akkor áll elő, ha  $A$  ugyanaz mint fent, és  $G \sim N(0, I)$ , akkor  $X = A^{1/2}G + \delta$  karakterisztikus függvénye a (61) formula szerinti.

Nolan [59] az  $X$  többváltozós stabil eloszlás hosszának (normájának) alapján vezet le az elliptikus és az izotróp stabil eloszlások sűrűségfüggvényére kifejezéseket. Ehhez legyen  $X$  centrált  $d$ -dimenziós izotróp stabil véletlen vektor, amelynek karakterisztikus függvénye

$$\exp(-\gamma_0^\alpha |u|^\alpha).$$

$X$  normája (hossza) ekkor

$$R = |X| = \sqrt{X_1^2 + \dots + X_d^2}.$$

Tehát  $R$  eloszlása használható a sűrűségfüggvény megadására. Tegyük fel, hogy  $d \geq 2$ . Ha  $0 < \alpha < 2$ , és  $X$  eloszlását tekintve egyenlő  $A^{1/2}Z$ -vel, ahol  $A$  mint a (63) formulában, pozitív stabil eloszlás, valamint  $Z \sim N(0, I)$ ,  $A$  és  $Z$  függetlenek egymástól, akkor

$$R^2 \stackrel{D}{=} A(Z_1^2 + \dots + Z_d^2) = AT, \quad (64)$$

ahol  $T \chi^2$ -eloszlású  $d$  szabadsági fokkal és független  $A$ -tól. A (64) képlet segítségével szimulálhatjuk a norma eloszlását közvetlenül, anélkül hogy ismernénk az általános  $X$  véletlen vektort. A képlet ad egy másik lehetőséget is radiál szimmetrikus eloszlás szimulálására  $d$ -dimenzióban. Legyen  $A$  mint (63) formulában,  $T \sim \chi^2$   $d$ -szabadsági fokkal és  $s$  egyenletes eloszlású  $\mathbb{S}$ -en, ekkor

$$X \stackrel{D}{=} \sqrt{AT} s$$

radiál szimmetrikus  $\alpha$ -stabil eloszlás  $\gamma_0$  skálaparaméterrel. Két dimenzióban  $T$  exponenciális és  $-2 \log U_1$  -ként generálható,  $S = (\cos(2\pi U_2), \sin(2\pi U_2))$  ahol  $U_1$  és  $U_2 \sim U(0, 1)$  függetlenek.

A három dimenziós gömbszimmetrikus (izotróp) stabil eloszlású véletlen vektorok szimulálásának egy eltérő módját adja meg Uchaikin és Zolotarev [79], (8.4 fejezet, 221. oldal). Az algoritmus két lépésből áll. Meghatározandó egy izotróp vektor

$$\Omega = (\Omega_1, \Omega_2, \Omega_3),$$

azaz egyenletes eloszlás szerint választunk egy pontot az  $\mathbb{S}$  gömb felszínén. A koordináták két egyenletes eloszlású véletlen szám segítségével készíthetők:

$$\begin{aligned} \Omega_1 &= \sqrt{1 - \mu^2} \sin \phi, \\ \Omega_2 &= \sqrt{1 - \mu^2} \cos \phi, \\ \Omega_3 &= \mu = \cos \phi, \\ \phi &= 2\pi U_1 \text{ és } \mu = 2U_2 - 1. \end{aligned}$$

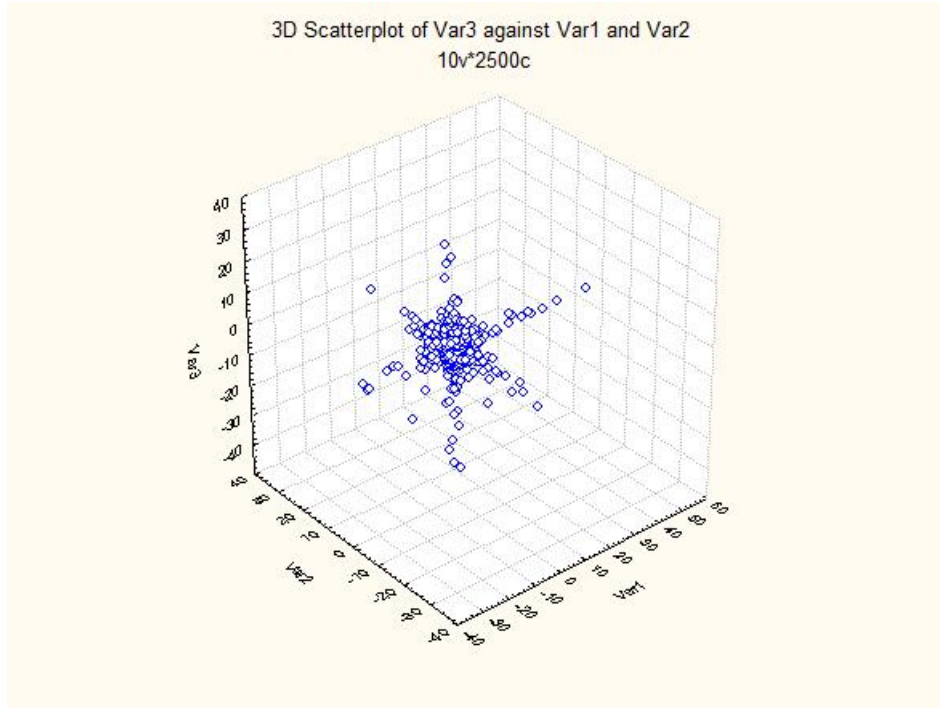
Az  $R$  eloszlásának megfelelő véletlen szám készítése bonyolultabb. A módszer  $F(r)$  numerikus invertálásán alapul.

$$r = F^{-1}(u) \equiv r(u), \quad 0 < u < 1$$

A transzformáció elvégzéséhez szükség van  $F(r)$  sorfejtésére:

$$1 - F(r) = \frac{2}{\pi\alpha} \sum_{n=1}^{\infty} (-1)^{n+1} \Gamma(n\alpha + 2) \sin(n\alpha\pi/2) n^{-1} r^{-n\alpha}, \quad (65)$$

13. ábra.  $\alpha = 1.5$  paraméterű stabil eloszlású három dimenziós generált minta.

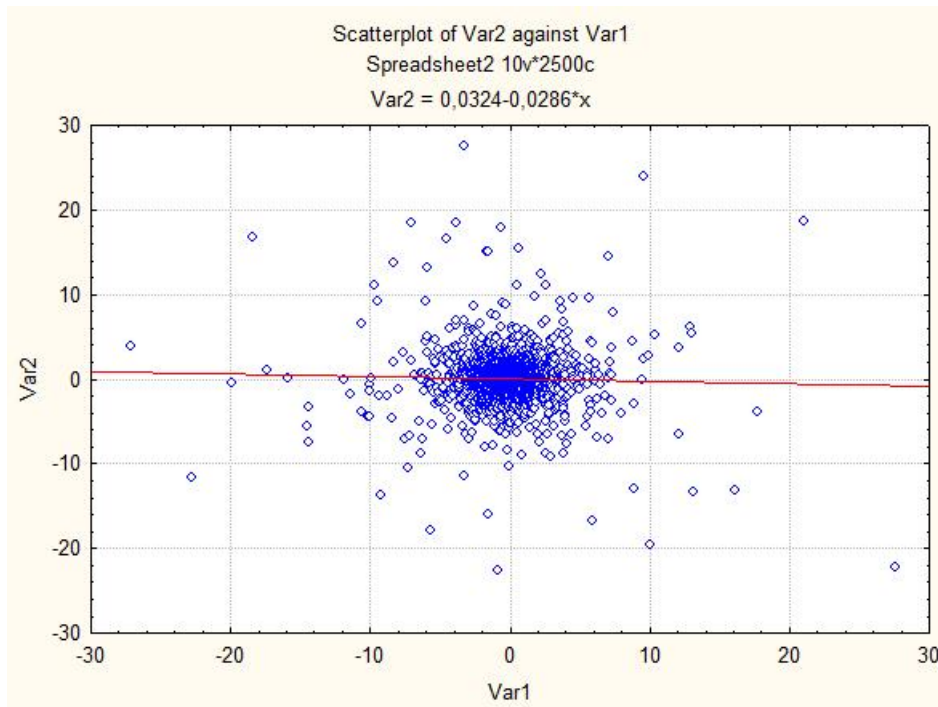


$$F(r) = \frac{4}{\pi\alpha} \sum_{n=1}^{\infty} (-1)^{n+1} \Gamma\left(\frac{2n+1}{\alpha}\right) \frac{n}{(2n+1)!} r^{2n+1}. \quad (66)$$

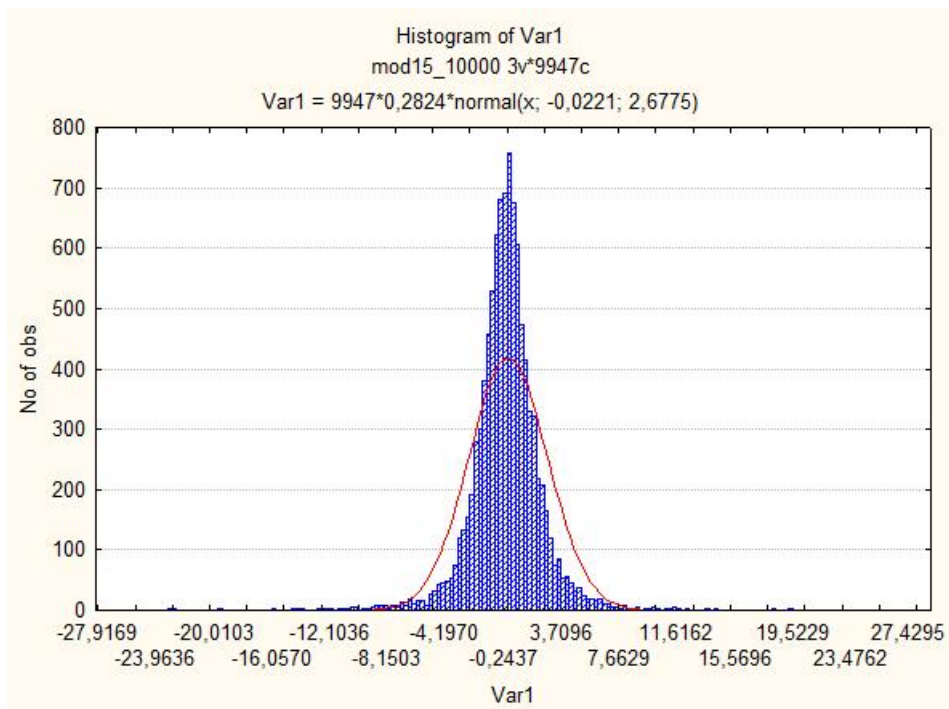
A sorfejtés alapján Uchaikin és Zolotarev egy összetett, több részből álló közelítést mutat be, a módszer implementálásához szükséges közelítések megtalálhatók a függelékben.

Végül az egydimenziós eloszlás generálásra ismert (60) képlet alkalmazásával készíthető független mintákat vektorba foglalva tetszőleges dimenziójú többváltozós valószínűségi változót szimulálhatunk. Ez megfelel a diszkrét spektrálmértékű esetnek az  $(1, 0, 0)$ ,  $(0, 1, 0)$ ,  $(0, 0, 1)$  pontok választásával, ekkor a választott pontok irányában a sűrűségfüggvény nagyobb értéket vesz fel. Három dimenziós esetben ( $\alpha=1.5$ , 2500 pont) egy jellemző képet mutat a véletlen pontok szóródásáról az 13. ábra. A gömbszimmetrikus generálással kapott két változós eloszlást ( $\alpha=1.5$ , 2500 pont) jellemzi a 14. ábra.

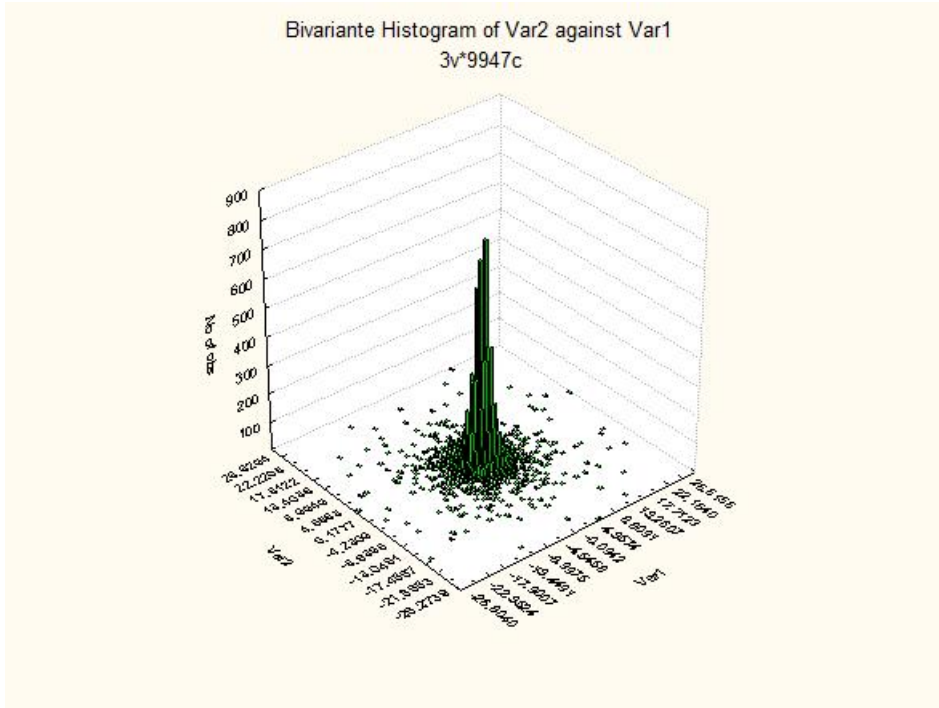
14. ábra.  $\alpha = 1.5$  paraméterű két dimenziós gömbszimmetrikus generált minta.



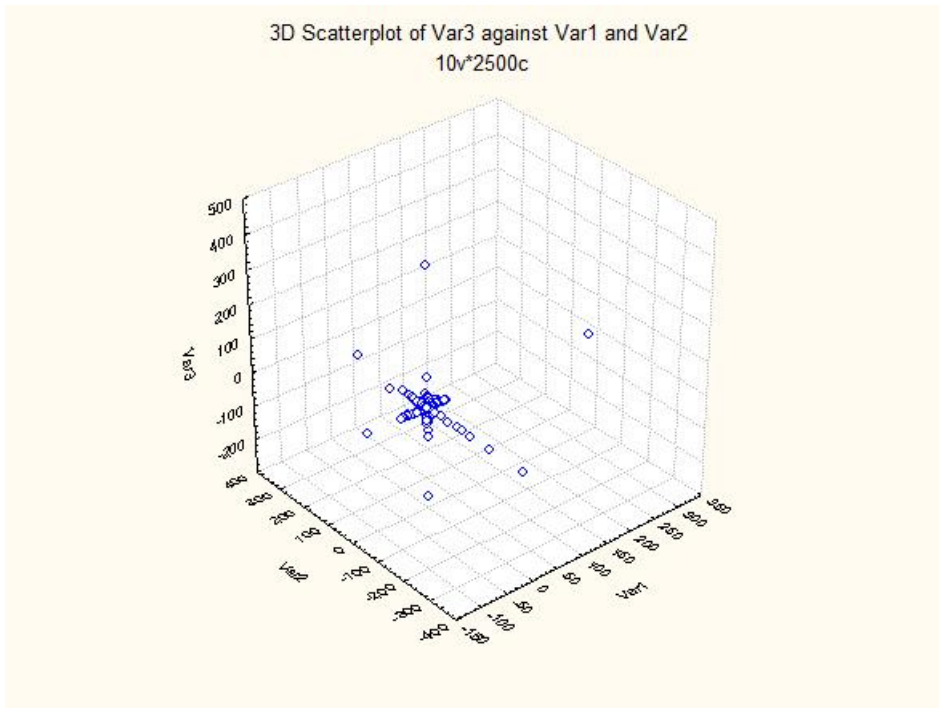
15. ábra.  $\alpha = 1.5$  paraméterű egy dimenziós generált minta.



16. ábra.  $\alpha = 1.5$  paraméterű két dimenziós generált minta.



17. ábra.  $\alpha = 1.3$  paraméterű három dimenziós generált minta.



## 5.2. A PIT paraméterbecslési módszer statisztikai vizsgálata

A PIT becslési eljárás statisztikai jellemzőinek vizsgálatára több Monte-Carlo szimulációból álló szimuláció sorozatot végeztem. A szimulációs vizsgálattal meghatároztam a paraméterek becslésének pontosságát, megbízhatóságát, az együttes becslésből adódó korrelációt, illetve kovarianciát a három paraméter között. A szimulációkban rögzített  $\alpha$  paraméterű, standardizált véletlen mintákat generáltam, és becsültem a PIT módszerrel a paramétereket, majd kiszámoltam a becsült paraméterek ismert, tényleges paraméterektől való átlagos eltérését és a Monte-Carlo szimulációkban kapott becslések szórását, amely információval szolgál a becslési eljárás pontosságára nézve, a generált minták elemszámai és a Monte-Carlo ismétlések függvényében.

Egy becslési eljárás jóságára vonatkozóan akkor tudunk megállapítást tenni, hogyha ismert a becslés viselkedése kis mintákra, azaz véges elemszám esetén, valamint ismert a becslés, mint valószínűségi változó határeloszlása. A leggyakrabban használt becslések aszimptotikusan normális eloszlást követnek (pl. ML módszer), és így a becsült paraméterekre vonatkozóan konfidenciaintervallumokat tudunk konstruálni. Ha adhatóak konfidenciaintervallumok a becslésre, akkor a becslés hibája is becsülhetővé válik, vagy, ha a becslés hibája becsülhető, akkor adhatunk konfidenciaintervallumokat.

A hely- és skálaparaméter becslésére ismert elméleti eredmény szerint a PIT becslés torzítatlan és konzisztens, és rögzített (ismert)  $\alpha$  alakparaméter esetén aszimptotikusan normális. Abban az esetben, amikor mindhárom paramétert becsüljük, még nem született elméleti eredmény a becslések eloszlására vonatkozóan. Célom a statisztikai vizsgálatokkal a becslések viselkedésének megismerése volt kis minták esetén és aszimptotikusan is. A statisztikai vizsgálat fő motivációja az volt, hogy három becsült paraméter esetén is kimutathassuk a becslések együttes normalitását.

A szimulációk során vizsgáltam a becslések egyváltozós normalitását, valamint az együttes normalitást is különböző többváltozós normalitás tesztekkel. Az egyváltozós normalitás tesztelése az elméleti eredményeknek megfelelően igazolta a helyparaméter becslésének normalitását. A skálaparaméter becslésének normalitását kevés kivételtől eltekintve szintén elfogadhatjuk. Az alakparaméter becslésének egyváltozós normalitását is sikerült igazolni a legtöbb szimulációban, bár voltak olyan esetek, amikor a normalitást adott szignifikancia szinten elvettem.

A PIT becslés statisztikai jellemzőinek megismerésére a következő vizsgálatokat végeztem el a Monte-Carlo ismétlésekkel kapott becslés mintákra

- leíró statisztika: átlag, szórás, minimum, maximum
- korrelációs együtthatók, kovariancia mátrixok
- MSE értékek (mean squared error)
- érvényes becslések számának meghatározása
- egyváltozós normalitás tesztek paraméterenként -  $\chi^2$  próba, Kolmogorov-Smirnov próba, Sarkadi próba
- konfidencia intervallumok meghatározása paraméterenként
- a többváltozós normalitás tesztelésére
  - Mardia-féle többváltozós ferdeség és lapultság mutató [46] kiszámítása

– többváltozós omnibus teszt Doornik és Hansen [12] alapján.

A szimulációs vizsgálat második szakaszában a szakirodalomban ismert paraméterbecslési módszerek performanciájával hasonlítottam össze a PIT módszert. Az ismert eljárások teljesítményét Weron [81] foglalta össze szimulációs tanulmányban. A Weron tanulmányában szereplő paraméter értékek, mintaelemszámok és Monte-Carlo ismétlésszámok alkalmazásával készített szimuláció sorozat segítségével igazoltam, hogy a PIT módszer hasonló statisztikai jellemzőkkel bír, mint a Weron által vizsgált módszerek.

A Monte-Carlo szimulációs technika nagy mennyiségű véletlen kísérlet egymás utáni végrehajtását és a kapott eredmények átlagolását jelenti. Ez a fajta szimulációs eszköz azokban az esetekben hasznos, amikor több paraméter hatását szeretnénk vizsgálni, illetve az egyes kísérleteket nagy mértékben befolyásolja a véletlen, azaz a vizsgált jelenség vagy rendszer jellemzője nem számítható ki determinisztikus algoritmussal.

A szimuláció sorozatban a Monte - Carlo szimulációk változtatott paraméter értékei

- $n$ , a generált minta elemszáma;  $n = 50, 100, 400, 2500$
- $r$ , a Monte - Carlo ismétlések száma, azaz hányszor készítjük el az  $n$  elemű mintát  
 $r = 100, 400, 2500$
- az  $\alpha$  karakterisztikus kitevő értéke;  $\alpha = 1.3, 1.5, 1.7$

voltak. A szimuláció sorozat tehát 36 különálló szimuláció esetet foglal magába.

A szimulációk paramétereinek megadásánál az  $\alpha = 1.3, 1.5, 1.7$  értékek választásának magyarázata az, hogy a PIT becslés az  $[1, 2]$  intervallum végpontjaihoz közel eső  $\alpha$  paraméter esetén nem biztos, hogy minden esetben megoldást szolgáltat, mivel a véletlen hatása is szerepet játszik ekkor a becslés alakulásában. Azért, hogy minimális legyen az érvénytelen becslések száma, de még a teljes intervallumon vizsgálhassam a módszer jellemzőit, a végpontoktól kicsit messzebb, az  $[1, 2]$  intervallum közepe felé eső értékeket választottam. Az 1.5 -ös értéknek pedig közgazdaságtani jelentősége is van azon túl, hogy pontosan a tekintett intervallum középpontja.

A szimulációkban a stabil eloszlású véletlen minták generálása a Zolotarev-féle (60) formula alkalmazásával történt. A vizsgálatban standardizált, azaz  $\gamma = 1$  és  $\delta = 0$  skála- és helyparaméterrel rendelkező mintákat generáltam, az alakparaméter a fentieknek megfelelően változott a szimuláció sorozatban. A  $\beta$  ferdeségi paraméter értéke minden esetben nulla, azaz csak a szimmetrikus esetet vizsgáltam. A generált minták mindegyikére alkalmazva a PIT becslést, összetartozó  $\hat{\alpha}, \hat{\sigma}, \hat{\mu}$  becslés hármassokat kaptam.

5. táblázat. A PIT becslés performanciájának jellemzői,  $r = 2500$ 

$\alpha$	$n$	$\bar{\hat{\alpha}}$	$\bar{\hat{\gamma}}$	$\bar{\hat{\delta}}$	$\sigma(\hat{\alpha})$	$\sigma(\hat{\gamma})$	$\sigma(\hat{\delta})$	$r_{\alpha,\gamma}$	$r_{\alpha,\delta}$	$r_{\gamma,\delta}$
1.3	50	1.362996	1.054485	0.000862	0.206463	0.218358	0.194215	0.574019	0.022507	0.003332
1.3	100	1.322489	1.022797	0.004336	0.147954	0.152204	0.137281	0.607487	-0.040014	-0.026304
1.3	400	1.306025	1.006280	0.000084	0.073299	0.077120	0.069101	0.658297	-0.015864	-0.002690
1.3	2500	1.299937	1.000347	-0.000792	0.028803	0.029952	0.026553	0.642502	0.022661	0.033307
1.5	50	1.534916	1.021137	0.003957	0.207883	0.178588	0.175584	0.534839	0.003675	-0.000334
1.5	100	1.520216	1.010721	0.002316	0.158329	0.130271	0.124165	0.595818	0.013889	-0.004671
1.5	400	1.505993	1.001907	0.001505	0.078686	0.064437	0.061768	0.581140	-0.011002	0.028515
1.5	2500	1.500685	1.000939	-0.000185	0.031431	0.025221	0.024555	0.589993	0.010055	0.010404
1.7	50	1.688076	0.996272	-0.001882	0.187891	0.157124	0.159264	0.529540	-0.011794	0.010194
1.7	100	1.709380	1.006829	0.001146	0.140511	0.107856	0.109804	0.482513	-0.001954	-0.001560
1.7	400	1.704027	1.000218	-0.000536	0.073661	0.053958	0.055382	0.513515	0.002858	0.020375
1.7	2500	1.701075	1.000620	-0.000805	0.030251	0.022020	0.022022	0.518613	0.020298	-0.008963



A becslési eljárás pontosságára, megbízhatóságára vonatkozó információkat tartalmaz a 5. táblázat, amely az  $r = 2500$ -as ismétlések esetében a becslés aszimptotikus viselkedésének megfelelő szimulációkat összegzi. A táblázatban a paraméterenkénti átlagok, szórások, és páronkénti korrelációs együtthatók szerepelnek. A korrelációs együtthatók azt mutatják, hogy a  $\hat{\alpha}$  és  $\hat{\gamma}$  között erős kapcsolat áll fenn. Ez az eredmény abból következik, hogy a skálaparaméter segítségével becsüli az alakparamétert az eljárás. A korrelációs együttható  $0.5 - 0.6$  közötti értékű. A táblázat alapján úgy tűnik, hogy a korrelációs együttható értéke az alak- és skálaparaméter becslése között az  $\alpha$  paraméter növekedésével csökken. A skála- és helyparaméter, valamint a hely- és alakparaméter között egyáltalán nem jelentős a korreláció.

6. táblázat. MSE értékek  $r=2500$

$\alpha$	$n$	$MSE(\hat{\alpha})$	$MSE(\hat{\gamma})$	$MSE(\hat{\delta})$
$\alpha = 1.3$	50	0.046577	0.050629	0.037704
$\alpha = 1.3$	100	0.022387	0.023677	0.018857
$\alpha = 1.3$	400	0.005407	0.005985	0.004773
$\alpha = 1.3$	2500	0.000829	0.000897	0.000705
$\alpha = 1.5$	50	0.044417	0.032327	0.030833
$\alpha = 1.5$	100	0.025467	0.017079	0.015416
$\alpha = 1.5$	400	0.006225	0.004154	0.003816
$\alpha = 1.5$	2500	0.000988	0.000637	0.000603
$\alpha = 1.7$	50	0.035430	0.024691	0.025357
$\alpha = 1.7$	100	0.019823	0.011675	0.012053
$\alpha = 1.7$	400	0.005440	0.002910	0.003066
$\alpha = 1.7$	2500	0.000916	0.000485	0.000485

A leíró statisztikai jellemzők (átlagok és szórások) mellett a becslések pontosságát jellemzi az MSE (Mean Squared Error) érték, amelyeket szimulációnként a 6. táblázat tartalmaz, az aszimptotikus esetben ( $r = 2500$ ), mindhárom paraméterre. A táblázat jól mutatja az  $n$  mintaelemszám hatását a becslés pontosságára; azaz, hogy az  $n$  mintaelemszám növelésével hogyan csökken a becslések eltérése az ismert paraméterértékektől. A legalacsonyabb MSE értékeket az  $\alpha = 1.7$  alakparaméterű szimulációnál kaptam, míg az  $\alpha = 1.3$  paraméterű esetekhez szinte mindegyik szimuláció esetén a magasabb MSE értékek tartoznak.

A szimuláció sorozat minden szimulációs esetében meghatároztam a paraméter becslések közötti kapcsolatot jellemző kovariancia, és korrelációs mátrixokat. Az  $r = 2500$ -as aszimptotikus viselkedés jellemzőit mutató szimulációk korrelációs mátrixai részben már a performancia jellemzőket bemutató 5. táblázatban szerepelnek. A kovariancia mátrixok tanulmányozása érdekes lehet a változók együttes eloszlásának vizsgálatakor. Az elméleti eredmény alapján a hely- és skálaparaméter együttes eloszlásának kovariancia mátrixa kiszámítható. A három ismeretlen becsült paraméter esetén az empirikus eredmény eléréséhez a kovariancia mátrixok elemzése lehet szükséges. Az  $\alpha = 1.5$ ,  $r = 2500$  esethez tartozó kovariancia mátrixokat tartalmazzák a 7-10. táblázatok, amelyeknél a mintaelemszámok rendre  $n = 50, 100, 400, 2500$ .

7. táblázat. Kovariancia mátrix a három becsült paraméterre -  $n = 50$

	$\alpha$	$\gamma$	$\delta$
$\alpha$	0.043215	0.019856	0.000134
$\gamma$	0.019856	0.031894	-0.000010
$\delta$	0.000134	-0.000010	0.030830

8. táblázat. Kovariancia mátrix a három becsült paraméterre -  $n = 100$

	$\alpha$	$\gamma$	$\delta$
$\alpha$	0.025068	0.012289	0.000273
$\gamma$	0.012289	0.016970	-0.000076
$\delta$	0.000273	-0.000076	0.015417

9. táblázat. Kovariancia mátrix a három becsült paraméterre -  $n = 400$

	$\alpha$	$\gamma$	$\delta$
$\alpha$	0.006192	0.002947	-0.000053
$\gamma$	0.002947	0.004152	0.000113
$\delta$	-0.000053	0.000113	0.003815

10. táblázat. Kovariancia mátrix a három becsült paraméterre -  $n = 2500$

	$\alpha$	$\gamma$	$\delta$
$\alpha$	0.000988	0.000468	0.000008
$\gamma$	0.000468	0.000636	0.000006
$\delta$	0.000008	0.000006	0.000603

11. táblázat. Az érvényes becslések száma a generált minták elemszáma ( $n$ ) és az  $\alpha$  paraméter alapján

$\alpha$		1.3			1.5			1.7		
$r$		100	400	2500	100	400	2500	100	400	2500
$n$	50	94	380	2383	94	389	2419	91	360	2248
	100	99	398	2481	100	399	2493	98	389	2444
	400	100	400	2500	100	400	2500	100	400	2500
	2500	100	400	2500	100	400	2500	100	400	2500

A szimulációk alapján ha az  $\alpha$  paraméter értéke közel esik az  $[1, 2]$  intervallum valamelyik végpontjához, és a mintaelemszám alacsony, akkor előfordulhat, hogy a két skálaparaméter görbének nincsen metszéspontja az intervallumon, azaz a becslés nem szolgáltat érvényes eredményt. Ennek oka a véletlen hatása, illetve a számítások pontatlansága. Nagyobb mintaelemszámok esetén ennek a jelenség előfordulási gyakorisága csökken: kisebb minta elemszámoknál ( $n = 50, n = 100$ ) fordult leginkább elő, hogy a módszer nem adott megoldást. Az  $n = 50$  esetben ez a futtatások 5-10%-át érintette, az  $n = 100$ -as esetben 2-3%-ban jelentkezett a probléma. A 11. táblázat tartalmazza az érvényes becslések számát szimulációként. A táblázatban  $r$  a Monte-Carlo ismétlések számát,  $n$  generált minta elemszámát jelöli.

### 5.3. A becslések normalitásának vizsgálata

A  $\hat{\gamma}$ ,  $\hat{\delta}$ ,  $\hat{\alpha}$  becslések aszimptotikus normalitását a  $\chi^2$ , a Kolmogorov-Smirnov (KS) és a Sarkadi illeszkedésvizsgálat (goodness-of-fit) tesztekkel vizsgáltam. A  $\chi^2$  és a KS teszt gyakran használt statisztikai próbák, a Sarkadi próba kevésbé ismert. A  $\chi^2$  teszt előnye, hogy jól ismertek a statisztikai tulajdonságai, hátránya, hogy eredménye függ a teszt során alkalmazott intervallumok számának megválasztásától. A KS teszt normális nullhipotézis esetén szintén jól ismert tulajdonságokkal rendelkezik, és kevésbé szigorú, mint a  $\chi^2$  teszt. A Sarkadi próba egy véletlenített eljárás, amelyben a tesztstatisztika számításánál egy véletlenül kiválasztott elemmel csökkentjük a mintát, és ezt a mintaelemet felhasználjuk a tesztstatisztika kiszámolásához.

#### A $\chi^2$ próba eredményei

A Pearson-féle  $\chi^2$  próba illeszkedésvizsgálatra történő alkalmazásához a megfigyeléseket kategóriákba (intervallumokba) osztjuk, és az intervallumokban megfigyelt gyakoriságokat hasonlítjuk a nullhipotézisben szereplő elméleti eloszlásból számított gyakoriságokhoz. Legyen a megfigyelések száma  $N$ , és jelölje  $p_i$  az intervallumokba esés elméleti eloszlásból származó valószínűségeit, valamint  $k_i$  jelöli a megfigyelt gyakoriságokat. A  $\chi^2$  próba tesztstatisztikájának számítása

$$\chi^2 = \sum_{i=1}^n \frac{(k_i - Np_i)^2}{Np_i}.$$

A nullhipotézis fennállása esetén a  $\chi^2$  tesztstatisztika eloszlása  $\chi^2$  eloszlású  $n - 1$  szabadsági fokkal. A hipotézisvizsgálat ismert módszere szerint  $1 - a$  szignifikancia szinten, ahol az  $a$  az elsőfajú hiba elkövetésének valószínűségét jelenti, ha a tesztstatisztika értéke kisebb, mint a  $\chi^2$  eloszlásból kapott kritikus érték, akkor elfogadható a nullhipotézis.

A PIT eljárás becsléseinek vizsgálatakor a  $\chi^2$  próba számítását a MATLAB beépített *chi2gof* függvényével végeztem. Ez a program a  $\chi^2$  próbát a mintából becsült átlagú és szórású normális eloszláshoz illeszti. A tesztstatisztika számításánál az intervallumok száma kezdetben 9, 11, 13 és 15 volt, valamint ha valamelyik intervallumba túl kevés elem esik, akkor a tesztet számító *chi2gof* függvény két egymás melletti intervallumot összevon. Az elsőfajú hiba valószínűségét 0.05-re választottam.

A skála- és helyparaméter becslésekre kapott eredményeket a 12. és 13. táblázatokban közlöm. A táblázatokban a tesztek p-értéke, a p-értékek alatt a próba szabadsági foka található. A tesztek 9, 11, 13 és 15 intervallumra osztással is elkészítettem, és a táblázatokban a kapott legjobb illeszkedéseket (legmagasabb p-érték) tüntettem fel. A helyparaméter becslésének

12. táblázat. A meghatározott p-értékek és a szabadsági fokok a  $\gamma$  paraméter egyváltozós normalitásának tesztelése során  $\chi^2$  próbával

		$n=50$	$n=100$	$n=400$	$n=2500$
$\alpha = 1.3$	$r= 100$	0.2056 (8)	0.4858 (5)	0.9417 (8)	0.7841 (4)
	$r= 400$	0.0098 (8)	0.6162 (7)	0.8662 (8)	0.9900 (6)
	$r=2500$	0.0000 (4)	0.0087 (9)	0.8075 (10)	0.1320 (8)
$\alpha = 1.5$	$r= 100$	0.2746 (5)	0.2627 (7)	0.9283 (4)	0.7503 (8)
	$r= 400$	0.2494 (6)	0.1383 (9)	0.7935 (4)	0.8591 (7)
	$r=2500$	0.2780 (8)	0.1892 (9)	0.8564 (8)	0.7954 (7)
$\alpha = 1.7$	$r= 100$	0.8198 (4)	0.2510 (4)	0.3999 (5)	0.9975 (5)
	$r= 400$	0.3458 (8)	0.9148 (7)	0.8608 (4)	0.8946 (7)
	$r=2500$	0.5492 (8)	0.2512 (9)	0.9306 (6)	0.6685 (7)

normalitását az elméleti eredményeknek megfelelően mindegyik szimulációban elfogadhatjuk.

A skálaparaméter becslésének normalitását néhány szimulációs esetet kivéve szintén elfogadhatjuk, tehát a várakozásnak megfelelő eredményt kaptam. Azok a szimulációk, amelyek esetén alacsony p-értékek adódtak, és a normalitást el kell utasítani, az  $\alpha = 1.3$ -mas alakparaméter értékhez és a kis mintás, magas ismétlésszámú esetekre adódtak (1:  $n = 50, r = 400$ ; 2:  $n = 50, r = 2500$ ; 3:  $n = 100, r = 2500$ ). Ennek a jelenségnek az oka még további vizsgálatokat igényel.

Az alakparaméter normalitásának igazolására még nem született elméleti eredmény, ezért csak a szimulációval kapott eredmények állnak rendelkezésre. Ennek a paraméternek a normalitásvizsgálata ezért kiemelt szerepet kap, mert ez a három paraméter együttes normalitásvizsgálatának előfeltétele is egyben.

Az alakparaméter normalitásvizsgálatának eredményeit tartalmazza a 14. táblázat. Az  $\alpha$  paraméter  $\chi^2$  tesztjeinek eredményei alapján az aszimptotikus normalitás a legtöbb szimulációban elfogadható, viszont általában az  $r = 2500$  nagy ismétlésszámú, kis mintás ( $n = 50$  és  $n = 100$ ) esetekben az alakparaméter értékek egyikénél sem teljesült. Az alakparaméter becslése függ a skálaparaméter becslésétől, ezért elképzelhető, hogy az alakparaméter normalitása azért sérül, mert a skálaparaméter becslés sem normális eloszlású kis mintás esetben ( $\alpha = 1.3$  paraméterérték).

Egy másik lehetséges oka a normalitás sérülésének az  $\alpha = 1.3$  alakparaméterű szimulációkban tapasztalt skálaparaméter normalitásának sérülésén kívül, hogy mivel kis mintás szimulációról van szó, a véletlen hatása miatt több esetben kerül a skálaparaméter görbék metszéspontja

13. táblázat. A meghatározott p-értékek és a szabadsági fokok a  $\delta$  paraméter egyváltozós normalitásának tesztelése során  $\chi^2$  próbával

		$n=50$	$n=100$	$n=400$	$n=2500$
$\alpha = 1.3$	$r= 100$	0.9374 (7)	0.9273 (7)	0.3552 (6)	0.5093 (4)
	$r= 400$	0.5003 (5)	0.4898 (6)	0.8759 (11)	0.7158 (8)
	$r=2500$	0.7164 (10)	0.4003 (6)	0.7846 (8)	0.9336 (9)
$\alpha = 1.5$	$r= 100$	0.9348 (8)	0.3313 (7)	0.4424 (6)	0.8370 (7)
	$r= 400$	0.0910 (6)	0.6099 (7)	0.1604 (10)	0.6405 (7)
	$r=2500$	0.6960 (12)	0.7007 (8)	0.8856 (7)	0.9688 (7)
$\alpha = 1.7$	$r= 100$	0.7385 (8)	0.9688 (5)	0.5362 (5)	0.8654 (7)
	$r= 400$	0.5786 (8)	0.6390 (9)	0.9801 (6)	0.6876 (9)
	$r=2500$	0.6250 (7)	0.9126 (6)	0.6851 (7)	0.4936 (5)

az intervallumon kívülre, és így a becslések eloszlása aszimmetrikussá válik, azaz az eloszlások farka levágásra kerül amiatt, mert csak az (1, 2) intervallumban lehet az  $\alpha$  paraméter értéke.

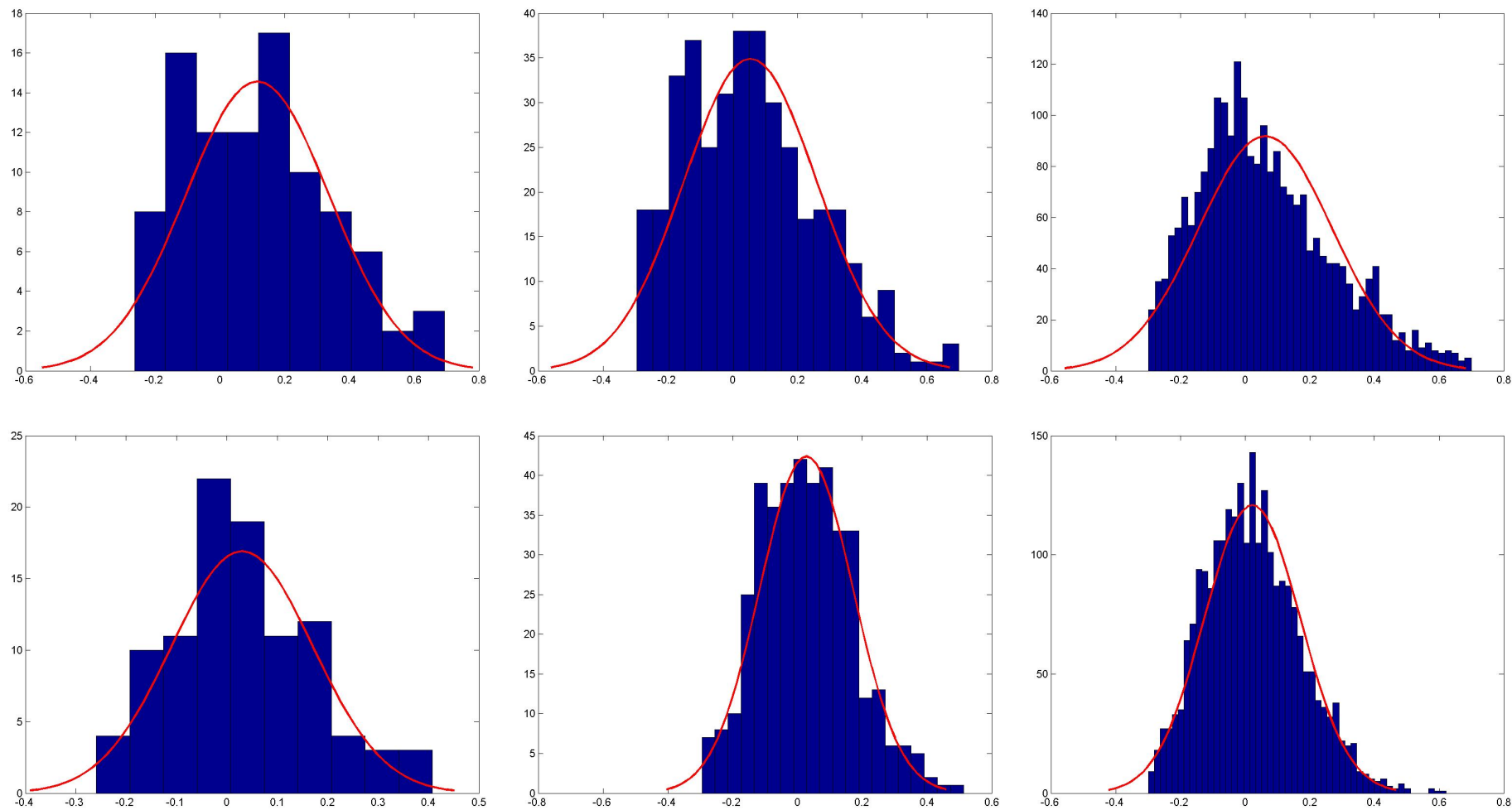
Az alakparaméter becslésekből képzett mintákat mutatja a 18. és 19. ábra, abban a szimuláció sorozatban, amikor az alakparaméter értéke  $\alpha = 1.3$ . Az egy sorban lévő ábrák az azonos mintaelemszámokhoz tartozó szimuláció esetek, és a Monte-Carlo ismétlések száma azonos az oszlopokban. Azaz, az első sor első ábrája tartozik az  $n = 50, r = 100$  szimulációhoz, az első sor második ábrája az  $n = 50, r = 400$ , az utolsó pedig az  $n = 50, r = 2500$  szimulációhoz. Az ábrákon feltüntettem a becslésekhez illesztett, a mintából becsült paraméterű normális eloszlás görbét is. Ezek az ábrák a MATLAB *histfit* függvényével készültek, és az ábrák intervallumai nem egyeznek meg a  $\chi^2$  próba felosztásában használt intervallumokkal.

Az ábrákon látható a már említett aszimmetria probléma, ami a becslési eljárás jellemzőjéből ered, azaz, hogy alacsony mintaelemszámnál a becslés nem határozható meg az [1, 2] intervallumban. A magasabb mintaelemszámú esetekben viszont látható, hogy a módszer egyre pontosabbá válik, és már nem játszanak torzító szerepet az intervallumból kieső becslések. Azoknál a szimulációknál, ahol a  $\chi^2$  próba a 9, 11, 13 és 15 intervallumos felosztással is elutasította a normalitást az alakparaméter esetén, levágott (trimmed) normális eloszláshoz való illeszkedésvizsgálattal tovább vizsgálható.

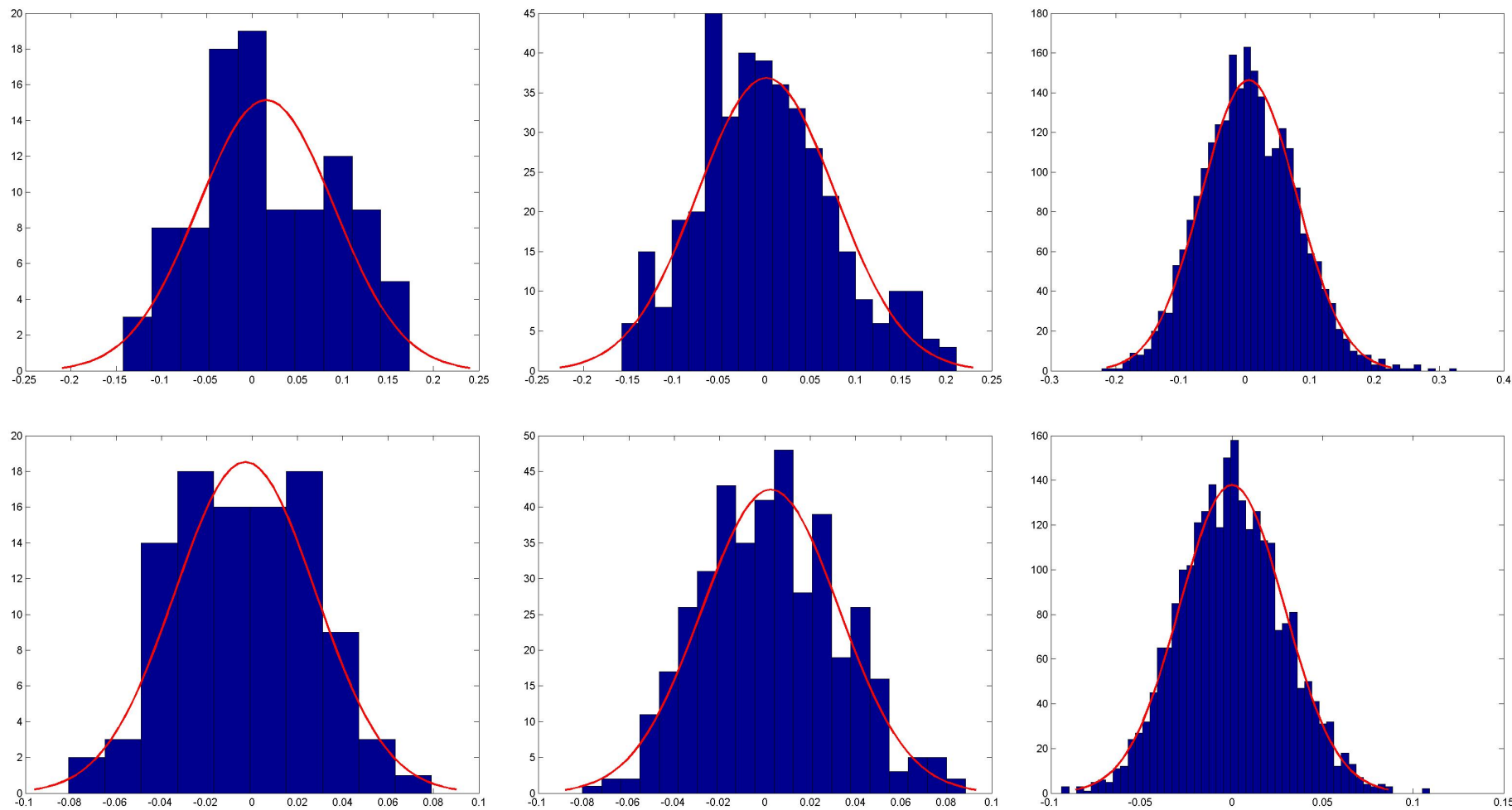
14. táblázat. A meghatározott p-értékek és a szabadsági fokok az  $\alpha$  paraméter egyváltozós normalitásának tesztelése során  $\chi^2$  próbával

		$n=50$	$n=100$	$n=400$	$n=2500$
$\alpha = 1.3$	$r= 100$	0.2053 (7)	0.7346 (8)	0.3214 (8)	0.2946 (7)
	$r= 400$	0.0282 (5)	0.5215 (8)	0.0522 (6)	0.5043 (7)
	$r=2500$	0.0000 (6)	0.0000 (6)	0.1544 (9)	0.8829 (11)
$\alpha = 1.5$	$r= 100$	0.8489 (8)	0.0767 (8)	0.9939 (8)	0.8856 (7)
	$r= 400$	0.1805 (12)	0.3239 (11)	0.2755 (10)	0.8731 (9)
	$r=2500$	0.0368 (8)	0.0010 (10)	0.9351 (5)	0.6571 (9)
$\alpha = 1.7$	$r= 100$	0.9844 (6)	0.5725 (5)	0.7544 (5)	0.5716 (4)
	$r= 400$	0.2756 (9)	0.3259 (11)	0.2926 (8)	0.3751 (9)
	$r=2500$	0.0000 (5)	0.1889 (8)	0.7703 (10)	0.7197 (5)

18. ábra. Normalitásvizsgálat az  $\alpha$  paraméterre (felső sor:  $n = 50$ , alsó sor:  $n = 100$ )



19. ábra. Normalitásvizsgálat az  $\alpha$  paraméterre (felső sor:  $n = 400$ , alsó sor  $n = 2500$ )





## A Kolmogorov - Smirnov próba eredményei

Az egymintás Kolmogorov-Smirnov statisztika az empirikus eloszlásfüggvény és az elméleti nullhipotézisben szereplő eloszlásfüggvény maximális eltérésén alapszik. A teszt többféle képpen elvégezhető, egy lehetséges alakja a következő:

$$D = \max(|F(x) - F_n(x)|),$$

ahol  $F_n(x)$  az empirikus eloszlásfüggvényt,  $F(x)$  a nullhipotézis szerinti eloszlást, ebben az esetben a  $\Phi(x)$  standard normális eloszlásfüggvényt jelöli.

A Kolmogorov - Smirnov próba számítását a MATLAB *kstest* függvényével végeztem. A program képes egyoldali és kétoldali próbát számítani, a kétoldali próba esetén azt vizsgáljuk, hogy a minta empirikus eloszlásfüggvénye megegyezik-e a nullhipotézisben szereplő eloszlásfüggvénnyel, az egyoldali próbánál lehetőség van a nullhipotézisbeli eloszlásfüggvénytől "nagyobb", és "kisebb" eloszlásfüggvények tesztelésére. A KS próba kritikus értékei kritikus érték táblázatban ismertek, amelyből a MATLAB közelítő formulával, vagy interpolációval határozza meg a minta esetén érvényes pontos kritikus értéket<sup>8</sup>.

A becslésekből képzett mintát standardizáltam a minta empirikus szórásával és –az átlag helyett– az ismert paraméterértékkel (a Monte-Carlo szimulációs paraméterek,  $\alpha = 1.3$ ,  $\alpha = 1.5$ ,  $\alpha = 1.7$ ;  $\gamma = 1$ ,  $\delta = 0$ ), majd az így kapott mintát a standard normális eloszláshoz képest teszteltem.

A 15. táblázatban a kapott tesztstatisztika értékekhez tartozó  $p$ -értékeket tüntettem fel. Az elsőfajú hibát most is 5%-nak választottam. A KS-próba a  $\chi^2$ -próbához hasonló eredményt adott, azaz a kis mintás esetekben ( $n = 50$ ,  $n = 100$ ,  $n = 400$ ) az  $r = 2500$ -as ismétlésű szimulációkból kapott becslések normalitását minden  $\alpha$  paraméterértékre el kellett utasítanom.

15. táblázat. A meghatározott  $p$ -értékek az  $\alpha$  paraméter normalitásának vizsgálatára Kolmogorov - Smirnov próbával

		$n=50$	$n=100$	$n=400$	$n=2500$
$\alpha = 1.3$	$r = 100$	0.000775	0.321002	0.110517	0.240236
	$r = 400$	0.005550	0.014744	0.748535	0.230706
	$r = 2500$	0.000000	0.000309	0.013647	0.423348
$\alpha = 1.5$	$r = 100$	0.055635	0.043293	0.816827	0.599934
	$r = 400$	0.080290	0.136900	0.238664	0.879341
	$r = 2500$	0.000000	0.000120	0.026103	0.271226
$\alpha = 1.7$	$r = 100$	0.193768	0.260638	0.620750	0.988977
	$r = 400$	0.227625	0.010404	0.494786	0.278681
	$r = 2500$	0.000002	0.000022	0.019143	0.420449

<sup>8</sup>A Kolmogorov-Smirnov próbát számító *kstest* függvényről bővebben: <http://www.mathworks.com/help/stats/kstest.html>

## A Sarkadi próba eredményei

A Sarkadi próba (Sarkadi-Störmer próba) egy véletlenített normalitás teszt, melynek számítása során a mintaelemek közül véletlenül választott elemeket felhasználunk a minta transzformálására és így csökkentjük a minta elemszámát, Sarkadi és Tusnady [72]. A módszer a Durbin transzformáció módosítása. A próba alapötlete az, hogy az eredeti mintaelemek transzformálásával a nullhipotézis fennállása esetén független, azonos eloszlású valószínűségi változók egy halmazát kapjuk. A transzformációhoz valamelyik mintaelemet használjuk fel, ezért a teszt kiemetele függ az elemek sorrendjétől, és így szerepet játszik a véletlen a teszt eredményében. A próba elvégzéséhez saját MATLAB függvényt készítettem.

Legyenek  $x_1, \dots, x_n$  a véletlen minta elemei, amelyek ismeretlen szórásnégyzetű és várható értékű normális eloszlásból származnak. A transzformáció a következő:

$$y_i = \frac{x_i - \bar{x}''}{S} \psi_{n-2} \left( \frac{|x_{n-1} - x_n| \sqrt{n-2}}{S \sqrt{2}} \right) \quad (i = 1, 2, \dots, n-2)$$

$$\bar{x}'' = \frac{\sum_{i=1}^n x_i + \sqrt{\frac{n}{2}}(x_{n-1} + x_n)}{n + \sqrt{2n}}$$

$$S = \sqrt{\sum_{i=1}^n x_i^2 - \frac{1}{n} \left( \sum_{i=1}^n x_i \right)^2 - \frac{1}{2} (x_{n-1} - x_n)^2}$$

ahol a  $\psi_\nu(t)$  függvényt a következő összefüggés definiálja:

$$Q([\psi_\nu(t)]^2 | \nu) = 2P(t | \nu) - 1$$

( $\psi_\nu(t) \geq 0$ )

$$P(t | \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) \sqrt{\nu\pi}} \int_{-\infty}^t \left(1 + \frac{u^2}{\nu}\right)^{-\frac{\nu+1}{2}} du$$

$$Q(t | \nu) = \frac{1}{2^{n/2} \Gamma(\frac{\nu}{2})} \int_t^\infty e^{-\frac{u}{2}} u^{\frac{\nu}{2}-1} du$$

A  $P(t | \nu)$  és  $1 - Q(t | \nu)$  függvények a Student eloszlás és a  $\chi^2$  eloszlás eloszlásfüggvényei, mindkettő szabadsági foka  $\nu$ . A transzformált minta ezután a standard normális eloszláshoz illeszthető tetszőleges, ismert nemparaméteres eljárással.

A PIT becslés vizsgálatára a Sarkadi próbával kapott eredmények találhatóak a 16. táblázatban, amelyek alapján megállapítható, hogy ez a teszt is elutasította az  $n = 50$  és  $n = 100$  elemű mintás szimulációval kapott becslések normalitását  $r = 2500$  ismétlés esetén. Az  $\alpha$  paraméter szerint is változnak az értékek: egy kicsivel magasabb p-értékek adódtak az  $\alpha = 1.7$ -es esetekben, mint az  $\alpha = 1.3$ -mas szimulációkban, azaz a módszer aszimptotikus normalitása függ a minta alakparaméterétől.

16. táblázat. A meghatározott p-értékek az  $\alpha$  paraméter normalitásának vizsgálatára Sarkadi próbával

		$n=50$	$n=100$	$n=400$	$n=2500$
$\alpha = 1.3$	$r = 100$	0.160355	0.594312	0.094813	0.491981
	$r = 400$	0.001478	0.243142	0.004439	0.195978
	$r = 2500$	0.000000	0.000000	0.145810	0.491990
$\alpha = 1.5$	$r = 100$	0.100746	0.070665	0.914536	0.685101
	$r = 400$	0.029355	0.119937	0.136593	0.418395
	$r = 2500$	0.002463	0.000065	0.639667	0.815317
$\alpha = 1.7$	$r = 100$	0.894649	0.761276	0.095115	0.799296
	$r = 400$	0.174968	0.093929	0.126697	0.060433
	$r = 2500$	0.000000	0.011084	0.825292	0.575386

### A többváltozós normalitásvizsgálatra alkalmazott módszerek és eredmények

A következő szakaszban a szimulációk során kapott becslés-hármasok többváltozós normalitásának tesztelését mutatom be. A többváltozós normalitás tesztelés módszereiről jó áttekintést nyújt Thode [78] könyve, valamint Henze [30] kritikai összehasonlító tanulmánya. A korábbi munkáim közül a PIT paraméterbecslés többváltozós normalitás tesztelésével saját publikációim között megadott [S8], [S9], [S14] hivatkozásokkal jelölt dolgozatokban foglalkoztam.

A paraméterbecslések többváltozós normalitásának tesztelésére az egyik legismertebb módszert használtam, amely a Mardia-féle ferdeség és lapultság mutatókat használja, Mardia [47]. Ezen mutatók viszonylag egyszerűen kiszámíthatóak, és a teszt elvégzéséhez rendelkezésre állnak kritikus érték táblázatok. A másik többváltozós teszt, amelyet alkalmaztam, a Doornik és Hansen [12] féle omnibusz teszt, amely a ferdeség és lapultság mutatók ötvözését, egyetlen statisztikába való tömörítését alkalmazza.

#### *A Mardia-féle ferdeség és lapultság mutatók*

Legyen  $X^T = (X_1, \dots, X_p)$  egy  $p \times n$ -es mátrix, amely tartalmazza az  $n$  megfigyelést  $p$ -dimenziós oszlopvektorokban. A minta átlagát jelölje

$$\bar{X} = n^{-1}(X_1 + \dots + X_n),$$

a kovariancia mátrixot jelölje

$$S = n^{-1}\tilde{X}^T\tilde{X}, \text{ ahol } \tilde{X}^T = (X_1 - \bar{X}, \dots, X_n - \bar{X}).$$

Határozzuk meg a

$$D = (d_{ij}) = \tilde{X}S^{-1}\tilde{X}^T \quad (67)$$

mátrixot, az ún. Mahalanobis távolságokat. Legyenek a ferdeségi és lapultsági mutatók

$$b_{1p} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d_{ij}^3, \quad (68)$$

$$b_{2p} = \frac{1}{n} \sum_{i=1}^n d_{ii}^2. \quad (69)$$

A hipotézisvizsgálat elvégezhető az alapján, hogy aszimptotikusan

$$\frac{nb_{1p}}{6} \sim \chi_{p(p+1)(p+2)/6}^2$$

és

$$b_{2p} \sim \mathcal{N}(p(p+2), 8p(p+2)/n),$$

ahol  $\mathcal{N}(\mu, \sigma^2)$  a normális eloszlást jelöli  $\mu$  várható értékkel és  $\sigma^2$  szórásnégyzettel. A teszt részletes leírása megtalálható [30] tanulmányban.

A többváltozós normalitásteszteléssel kapcsolatos fontos tény, hogy a nullhipotézis elvetése nem mutatja meg, hogy a minta hogyan, milyen módon tér el a normális eloszlástól (Henze [30] 3. fejezet). Továbbá a szignifikancia pontokat a normalitás hipotézisének teljesülése mellett határozták meg, ezért a teszt semmilyen információt nem szolgáltat a minta tényleges eloszlásáról.

A PIT eljárás szimulációs vizsgálatának adataira meghatároztam a ferdeség és lapultság mutatókat, melyeket a 17. táblázat tartalmaz. A szimuláció paramétereit után a táblázat oszlopai sorrendben a ferdeség mutató hipotézisvizsgálatára a döntés (D), az  $nb_{1p}/6$  tesztstatisztika értéke, a hozzá tartozó  $p$ -érték, valamint a lapultság mutató hipotézisvizsgálatára a döntés (D), a standardizált  $b_{2p}$  értéke, és a statisztikához tartozó  $p$ -érték. A döntés értéke a nullhipotézis elfogadása esetén 0, egyébként 1.

A többváltozós normalitás hipotézisét nem döntenek el egyértelműen az eredmények. A lapultság mutató alapján sokkal több szimulációs esetben elfogadhatjuk a normalitást, mint a ferdeség mutató alapján. A legmegbízhatóbban akkor fogadhatjuk el egy szimulációs esetben a normalitást, ha az mindkét mutató alapján elfogadható. Ilyen eset inkább a nagyobb elemszámú generált mintákra  $n = 2500$  fordult elő, azaz azt mondhatjuk, hogy a PIT becslések csak nagyobb mintákra együttes normális eloszlásúak. Kisebb mintaelemszámú szimulációkat inkább alacsonyabb ismétlésszámmal ( $r = 100$ ) fogadott el a teszt többváltozós normálisnak.

Ismert, hogy a többváltozós normalitás feltétele az egyváltozós peremeloszlások normalitása. Tehát azokban az esetekben, ahol már az egyváltozós normalitás sem teljesült, a többváltozós normalitás elvetését vártam. A többváltozós tesztek eredményei valóban ennek a várakozásnak megfelelően alakultak.

17. táblázat. A Mardia-féle többváltozós ferdeség és lapultság teszt döntései(D), a tesztstatisztika értéke és a p-érték

			D	$nb_{1p}/6$	p-érték	D	$b_{2p}$ standard	p-érték
$\alpha = 1.3$	$n = 50$	$r = 100$	1	22.398081	0.013200	0	0.265623	0.395265
	$n = 50$	$r = 400$	1	41.800912	0.000008	0	0.067850	0.472953
	$n = 50$	$r = 2500$	1	384.587157	0.000000	1	7.142642	0.000000
	$n = 100$	$r = 100$	0	14.091107	0.168876	0	1.032508	0.150917
	$n = 100$	$r = 400$	1	28.037849	0.001780	1	2.575136	0.005010
	$n = 100$	$r = 2500$	1	151.268178	0.000000	0	1.050224	0.146808
	$n = 400$	$r = 100$	0	5.753187	0.835556	1	3.910089	0.000046
	$n = 400$	$r = 400$	1	19.789087	0.031312	0	0.131562	0.447665
	$n = 400$	$r = 2500$	1	50.174041	0.000000	0	0.001991	0.499206
	$n = 2500$	$r = 100$	0	6.523418	0.769540	0	1.141685	0.126792
	$n = 2500$	$r = 400$	0	15.656387	0.109900	0	0.025705	0.489746
	$n = 2500$	$r = 2500$	1	25.240587	0.004908	0	0.010173	0.495942
$\alpha = 1.5$	$n = 50$	$r = 100$	0	4.221211	0.936816	0	1.508475	0.065716
	$n = 50$	$r = 400$	1	33.604329	0.000215	0	0.631936	0.263714
	$n = 50$	$r = 2500$	1	104.800776	0.000000	0	0.023828	0.490495
	$n = 100$	$r = 100$	0	11.436966	0.324502	0	0.298019	0.382844
	$n = 100$	$r = 400$	1	24.426016	0.006546	0	0.019138	0.492365
	$n = 100$	$r = 2500$	1	106.558322	0.000000	0	0.044863	0.482108
	$n = 400$	$r = 100$	0	12.963902	0.225697	0	0.027898	0.488872
	$n = 400$	$r = 400$	1	21.231148	0.019538	0	0.533860	0.296719
	$n = 400$	$r = 2500$	1	46.562315	0.000001	0	0.072398	0.471143
	$n = 2500$	$r = 100$	0	9.086179	0.523945	0	0.361614	0.358820
	$n = 2500$	$r = 400$	0	8.003159	0.628528	0	0.246190	0.402768
	$n = 2500$	$r = 2500$	1	18.372345	0.048998	0	0.281065	0.389330
$\alpha = 1.7$	$n = 50$	$r = 100$	1	23.053209	0.010552	0	0.000459	0.499817
	$n = 50$	$r = 400$	1	24.119597	0.007289	0	1.137085	0.127751
	$n = 50$	$r = 2500$	1	124.086225	0.000000	0	0.325204	0.372513
	$n = 100$	$r = 100$	0	7.687082	0.659372	1	2.679831	0.003683
	$n = 100$	$r = 400$	1	30.293531	0.000767	1	4.800182	0.000001
	$n = 100$	$r = 2500$	1	62.872934	0.000000	0	1.425348	0.077028
	$n = 400$	$r = 100$	0	9.724771	0.464963	0	0.074475	0.470316
	$n = 400$	$r = 400$	0	8.15017	0.614171	0	0.017879	0.492868
	$n = 400$	$r = 2500$	1	25.66026	0.004225	0	0.337840	0.367742
	$n = 2500$	$r = 100$	0	2.206083	0.994503	0	0.296225	0.383529
	$n = 2500$	$r = 400$	0	10.830614	0.370866	0	0.968045	0.166511
	$n = 2500$	$r = 2500$	0	9.092656	0.523333	0	0.257538	0.398382

### *Doornik és Hansen többváltozós normalitás tesztje*

A többváltozós normalitás tesztelésére alkalmazott második módszer a ferdeség és lapultság mutatókat együtt vizsgálja, amely a megfigyelések transzformációjának felhasználásával számítható, Doornik és Hansen [12]. Legyen  $X, \bar{X}, \tilde{X}, S$  mátrixok a megfigyelések, a megfigyelések átlag vektora, a standardizált minta és a kovariancia mátrix. Legyen

$$V = \text{diag}(S_{11}^{-1/2}, \dots, S_{pp}^{-1/2})$$

diagonális mátrix, amelynek főátlójában a szórások szerepelnek. A korrelációs mátrix  $C = VS V$ . Az eredeti minta transzformálásával készítünk egy mátrixot a következők szerint:

$$R' = H\Lambda^{-1/2}H^T V\tilde{X}^T, \quad (70)$$

ahol  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$  tartalmazza  $C$  sajátértékeit a diagonálisban és  $H$  oszlopai tartalmazzák a megfelelő sajátvektorokat. Ez a transzformáció standard normális változókat eredményez. Doornik és Hansen (DH) tesztje alapján a többváltozós statisztika

$$E_p = Z_1^T Z_1 + Z_2^T Z_2 \approx \chi_{(2p)}^2,$$

ahol  $Z_1^T = (z_{11}, \dots, z_{1p})$  és  $Z_2^T = (z_{21}, \dots, z_{2p})$  vektorok a többváltozós ferdeség és lapultság mutatókból kiszámíthatók, lásd Doornik és Hansen [12] függelék. A  $\chi^2$  eloszlás kritikus értékei 6-os szabadsági foknál 12.59 (95%), 16.81 (99%), 22.46 (99,9%).

A PIT becsléshármasok többváltozós normalitásának tesztelésére elvégzett DH teszt eredményeit a 18. táblázatban foglaltam össze. Ez a teszt összességében szigorúbbnak bizonyult a Mardia-féle ferdeség és lapultság tesztnél. A többváltozós normalitást ez a teszt is egyértelműen elutasította a kis mintás, nagy ismétlésszámú szimulációknál. Az  $\alpha = 1.7$  alakparaméterű szimulációknál több esetet fogadott el, mint az  $\alpha = 1.5$  és  $\alpha = 1.3$  alakparaméterű szimulációknál.

**A többváltozós normalitásvizsgálatok eredményeit összegezve azt mondhatjuk, hogy a három becsült paraméter együttes normalitása csak nagy mintákra ( $n = 2500$ ) állítható, és kis mintás esetben ( $n = 50, n = 100$ ), aszimptotikusan ( $r = 2500$ ) nem teljesül, viszont a PIT becslés robusztus.**

18. táblázat. Doornik és Hansen féle többváltozós teszt eredményei (D: döntés)

			D	DH statisztika	p-érték
$\alpha = 1.3$	$n = 50$	$r = 100$	1	25.260327	0.000306
	$n = 50$	$r = 400$	1	55.244072	0.000000
	$n = 50$	$r = 2500$	1	506.760696	0.000000
	$n = 100$	$r = 100$	1	13.427618	0.036726
	$n = 100$	$r = 400$	1	37.682759	0.000001
	$n = 100$	$r = 2500$	1	185.511807	0.000000
	$n = 400$	$r = 100$	0	6.503782	0.369180
	$n = 400$	$r = 400$	1	15.657785	0.015714
	$n = 400$	$r = 2500$	1	45.268559	0.000000
	$n = 2500$	$r = 100$	0	3.243214	0.777760
	$n = 2500$	$r = 400$	0	2.772454	0.836816
	$n = 2500$	$r = 2500$	1	27.527921	0.000115
$\alpha = 1.5$	$n = 50$	$r = 100$	0	7.890457	0.246239
	$n = 50$	$r = 400$	1	47.252034	0.000000
	$n = 50$	$r = 2500$	1	106.325875	0.000000
	$n = 100$	$r = 100$	1	23.606429	0.000617
	$n = 100$	$r = 400$	1	23.269263	0.000711
	$n = 100$	$r = 2500$	1	127.901426	0.000000
	$n = 400$	$r = 100$	0	2.556077	0.862140
	$n = 400$	$r = 400$	1	18.173520	0.005813
	$n = 400$	$r = 2500$	1	33.890413	0.000007
	$n = 2500$	$r = 100$	0	6.144038	0.407250
	$n = 2500$	$r = 400$	0	4.260192	0.641511
	$n = 2500$	$r = 2500$	1	14.995016	0.020296
$\alpha = 1.7$	$n = 50$	$r = 100$	0	5.042797	0.538336
	$n = 50$	$r = 400$	0	8.082860	0.232095
	$n = 50$	$r = 2500$	1	81.772383	0.000000
	$n = 100$	$r = 100$	0	1.208186	0.976478
	$n = 100$	$r = 400$	1	23.886761	0.000548
	$n = 100$	$r = 2500$	1	63.155934	0.000000
	$n = 400$	$r = 100$	0	6.724202	0.347108
	$n = 400$	$r = 400$	0	4.761545	0.574740
	$n = 400$	$r = 2500$	0	9.127458	0.166536
	$n = 2500$	$r = 100$	0	0.859518	0.990382
	$n = 2500$	$r = 400$	0	5.830887	0.442399
	$n = 2500$	$r = 2500$	0	4.212093	0.647999

## 5.4. Becslési eljárások összehasonlítása

A statisztikai vizsgálat alapján a PIT becslési eljárás performanciája megegyezik a Weron [81] dolgozatában vizsgált módszerek performancia jellemzőivel. A vizsgált becslési eljárások Fama és Roll kvantilis módszere (FR), McCulloch (CULL) módszere, a regressziós típusú becslés (REG), és a momentumok módszere (MOM) voltak. A [81] dolgozat 4.5 és 4.6 Táblázatai különböző  $\alpha$  és  $\beta$  stabil paraméterekkel,  $n = 100$ ,  $n = 500$ ,  $n = 2000$  elemű mintákkal és  $r = 25$  valamint  $r = 50$  ismétléssel végrehajtott Monte-Carlo szimulációk eredményeit tartalmazzák. Ezekből a táblázatokból kigyűjtöttem a szimmetrikus esetre vonatkozó eredményeket, és megismételtem ezekkel azonos beállításokkal a PIT becslési eljárás szimulációs vizsgálatát. A szakasz eredményeit a [S2] dolgozatban publikáltam.

A 19 - 22. táblázatok a becslések átlagát, minimumát, maximumát, az  $\hat{\alpha}$  és  $\hat{\gamma}$  becslések MSE értékét tartalmazzák a különböző Monte-Carlo beállítások esetén. A 19. táblázatban a generált minták skálaparamétere  $\gamma = 0.1, 1, 10$ , a további paraméterek:  $\alpha = 1.5, \delta = 0, n = 500, r = 25$ , tehát ebben a szimulációban a skálaparaméter becslésre gyakorolt hatását vizsgáljuk. A 20-22. táblázatok eredményei standard mintákhoz tartoznak, a táblázatok egy-egy  $n$  mintaelemszámhoz tartozó eredményeket mutatnak. Az egyes táblázatokban közölt eredményeket különböző alakparaméterek ( $\alpha = 1.2, 1.5, 1.8$ ) esetén határozták meg. Ezekben a szimulációkban főleg a mintaelemszám becslésre gyakorolt hatása mutatkozik meg (a három táblázatot együttesen vizsgálva).

**Összességében azt mondhatjuk, hogy a PIT becslési eljárás pontossága, megbízhatósága a performancia jellemzők alapján nem tér el szignifikánsan a szakirodalomban ismert többi, Weron által vizsgált módszertől.**

A Monte-Carlo szimulációk ismétlésszáma  $r = 25$ , és a kisebb mintás ( $n = 100$ ) esetben  $r = 50$  volt, ami a jelenlegi számítási kapacitások lehetőségeit figyelembe véve kicsit alacsonynak tűnhet. A beállításokon viszont az összehasonlíthatóság kedvéért nem változtattam, a PIT módszert is ezekkel az ismétlésszámokkal futtattam.

19. táblázat. Becslési eljárások összehasonlítása. ( $\alpha = 1.5, \delta = 0, n = 500, r = 25$ ) \*

$\gamma$	Módszer	$\bar{\alpha}$	$\alpha_{min}$	$\alpha_{max}$	$MSE_{\alpha}$	$\bar{\gamma}$	$\gamma_{min}$	$\gamma_{max}$	$MSE_{\gamma}$
$\gamma = 10$	FR	1.502	1.298	1.745	0.01128	10.230	8.989	11.291	0.35687
	CULL	1.502	1.288	1.771	0.01491	10.100	9.073	11.095	0.33746
	REG	1.486	1.245	1.628	0.00972	10.067	8.999	10.937	0.28761
	PIT	1.512	1.382	1.682	0.00467	9.937	8.829	11.592	0.40294
$\gamma = 1$	FR	1.518	1.351	1.677	0.00558	1.020	0.908	1.205	0.00447
	CULL	1.518	1.415	1.692	0.00553	1.017	0.903	1.183	0.00394
	REG	1.515	1.332	1.643	0.00729	1.011	0.914	1.148	0.00271
	PIT	1.495	1.319	1.663	0.00630	0.989	0.850	1.080	0.00296
$\gamma = 0.1$	FR	1.512	1.374	1.643	0.00595	0.099	0.088	0.110	0.00003
	CULL	1.498	1.370	1.603	0.00437	0.098	0.087	0.108	0.00003
	REG	1.495	1.361	1.620	0.00549	0.098	0.086	0.108	0.00004
	PIT	1.512	1.360	1.594	0.00395	0.100	0.088	0.107	0.00001

\* Az FR, CULL és REG adatok forrása [81], 4.5. Táblázat.



20. táblázat. Becslési eljárások összehasonlítása. ( $\gamma = 1, \delta = 0, n = 2000, r = 25$ ) \*

$\alpha$	Módszer	$\bar{\alpha}$	$\alpha_{min}$	$\alpha_{max}$	$MSE_{\alpha}$	$\bar{\gamma}$	$\gamma_{min}$	$\gamma_{max}$	$MSE_{\gamma}$
$\alpha = 1.8$	FR	1.812	1.726	1.950	0.00306	0.993	0.952	1.031	0.00055
	CULL	1.817	1.714	2.000	0.00475	0.994	0.938	1.032	0.00054
	MOM	1.800	1.745	1.856	0.00092	0.992	0.940	1.029	0.00038
	REG	1.801	1.735	1.848	0.00101	0.992	0.948	1.023	0.00033
	PIT	1.795	1.751	1.846	0.0008	0.999	0.956	1.049	0.0007
$\alpha = 1.5$	FR	1.496	1.410	1.588	0.00260	1.003	0.946	1.047	0.00073
	CULL	1.498	1.367	1.595	0.00288	1.001	0.931	1.051	0.00104
	MOM	1.487	1.409	1.561	0.00208	0.999	0.953	1.048	0.00061
	REG	1.492	1.371	1.630	0.00351	1.002	0.948	1.053	0.00082
	PIT	1.504	1.454	1.574	0.0012	1.004	0.952	1.060	0.0007
$\alpha = 1.2$	FR	1.207	1.159	1.264	0.00099	0.994	0.935	1.060	0.00082
	CULL	1.211	1.135	1.275	0.00143	0.985	0.945	1.038	0.00096
	MOM	1.204	1.111	1.276	0.00148	0.993	0.909	1.044	0.00095
	REG	1.201	1.132	1.272	0.00122	0.992	0.927	1.071	0.00099
	PIT	1.195	1.116	1.264	0.0017	0.997	0.915	1.068	0.0017

\* Az FR, CULL, MOM és REG adatok forrása [81], 4.6. Táblázat.

21. táblázat. Becslési eljárások összehasonlítása. ( $\alpha = 1.5, \gamma = 1, \delta = 0, n = 500, r = 25$ ) \*

$\alpha$	Módszer	$\bar{\alpha}$	$\alpha_{min}$	$\alpha_{max}$	$MSE_{\alpha}$	$\bar{\gamma}$	$\gamma_{min}$	$\gamma_{max}$	$MSE_{\gamma}$
$\alpha = 1.8$	FR	1.794	1.597	1.989	0.00978	1.007	0.912	1.124	0.00316
	CULL	1.806	1.597	2.000	0.01605	1.011	0.891	1.131	0.00378
	MOM	1.776	1.560	1.909	0.00630	1.003	0.897	1.110	0.00243
	REG	1.774	1.628	1.887	0.00528	1.003	0.895	1.108	0.00250
	PIT	1.818	1.716	1.911	0.0027	1.011	0.894	1.084	0.0019
$\alpha = 1.5$	FR	1.497	1.329	1.733	0.00833	0.992	0.900	1.109	0.00285
	CULL	1.487	1.274	1.690	0.01077	0.983	0.883	1.091	0.00320
	MOM	1.501	1.277	1.698	0.00677	0.996	0.847	1.110	0.00389
	REG	1.504	1.338	1.663	0.00664	0.991	0.894	1.147	0.00361
	PIT	1.498	1.346	1.610	0.0046	1.010	0.908	1.156	0.0030
$\alpha = 1.2$	FR	1.174	1.063	1.353	0.00379	0.995	0.867	1.100	0.00442
	CULL	1.169	1.041	1.329	0.00470	0.982	0.883	1.078	0.00420
	MOM	1.155	1.017	1.299	0.00692	0.970	0.857	1.115	0.00459
	REG	1.195	1.059	1.347	0.00614	1.000	0.870	1.143	0.00486
	PIT	1.218	1.092	1.332	0.0043	0.996	0.865	1.063	0.0027

\* Az FR, CULL, MOM és REG adatok forrása [81], 4.6. Táblázat.

22. táblázat. Becslési eljárások összehasonlítása. ( $\alpha = 1.5, \gamma = 1, \delta = 0, n = 100, r = 50$ ) \*

$\alpha$	Módszer	$\bar{\alpha}$	$\alpha_{min}$	$\alpha_{max}$	$MSE_{\alpha}$	$\bar{\gamma}$	$\gamma_{min}$	$\gamma_{max}$	$MSE_{\gamma}$
$\alpha = 1.8$	FR	1.779	1.217	2.000	0.03902	0.987	0.803	1.233	0.01037
	CULL	1.788	1.284	2.000	0.04584	0.988	0.742	1.168	0.00957
	MOM	1.828	1.427	2.000	0.02498	1.001	0.840	1.185	0.00756
	REG	1.812	1.448	2.000	0.02528	0.995	0.824	1.182	0.00852
	PIT	1.779	1.488	1.999	0.0191	0.996	0.788	1.135	0.0077
$\alpha = 1.5$	FR	1.497	1.140	2.000	0.04066	1.004	0.679	1.374	0.02051
	CULL	1.504	1.113	2.000	0.04872	0.992	0.676	1.378	0.01980
	MOM	1.527	1.116	1.958	0.04003	1.016	0.646	1.320	0.01970
	REG	1.495	0.955	1.909	0.04424	0.998	0.654	1.313	0.01939
	PIT	1.516	1.244	1.875	0.0169	1.034	0.715	1.229	0.0140
$\alpha = 1.2$	FR	1.165	0.887	1.453	0.02202	0.995	0.800	1.264	0.01196
	CULL	1.150	0.870	1.460	0.02429	0.949	0.693	1.161	0.01371
	MOM	1.198	0.848	1.564	0.02624	0.992	0.721	1.273	0.01654
	REG	1.208	0.865	1.543	0.01851	1.004	0.754	1.289	0.01255
	PIT	1.237	1.023	1.620	0.0204	1.011	0.758	1.368	0.0191

\* Az FR, CULL, MOM és REG adatok forrása [81], 4.6. Táblázat.

## 6. A BÉT részvényeinek modellezése a PIT módszerrel

### 6.1. A hozambecslés modelljei

Ebben a fejezetben bemutatom a PIT becslési eljárás egy lehetséges alkalmazását. A Budapesti Értéktőzsde vezető papírjainak árfolyam ingadozásait vizsgáltam a naponkénti logaritmikus hozamok eloszlásának alapján. Vizsgálatomban a napi hozamokat függetlennek és szimmetrikus stabil eloszlásúnak feltételeztem. A szimmetrikus stabil modellből kiindulva becsültem a hozamok eloszlásának paramétereit és hipotézis vizsgálatot ellenőriztem az illeszkedést. A becsült paraméterek segítségével konfidencia intervallumokat adtam a hozamok alakulására, amely korlátozott előrejelzési lehetőséget is biztosít.

Az árfolyamváltozások vizsgálatára több modell használható: a relatív (százalékban kifejezett) hozam

$$R_t = \frac{P_t}{P_{t-1}} - 1, \quad (71)$$

és a folytonosan számított (continuously-compounded rate) logaritmikus hozam

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln(R_t + 1), \quad (72)$$

ahol  $P_t$  és  $P_{t-1}$  jelenti a részvény árát az  $t$ -edik és  $t - 1$ -edik időpillanatban. Ezt a modellt egy napos (one-day, single-period) modellnek nevezik. Ha  $R_t$  értéke kicsi, akkor  $\ln(1 + R_t) \approx R_t$  a logaritmus függvény sorfejtése miatt, ezért a két modell nagyon hasonló eredményt ad.

A logaritmikus hozam modell használatának előnye, hogy az eltelt időre nézve additív. A portfólió hozama a logaritmikus modell szerint:

$$r_{pt} = \ln\left(\sum_{i=1}^N w_i e^{r_{it}}\right),$$

ahol  $w_i$  jelöli az  $i$ -edik részvény súlyát a portfólióban ( $\sum_{i=1}^N w_i = 1, \forall i : w_i > 0$ ).

Gyakorlatban használatos a következő közelítés is:

$$r_{pt} \approx \sum_{i=1}^N w_i r_{it}.$$

Elemzésemben a Budapesti Értéktőzsde vezető részvényeinek  $r_i$  logaritmikus hozam eloszlásának paramétereit számítottam ki a PIT becslés és a STABLE program<sup>9</sup> segítségével. A négy stabil paraméter árfolyamok esetén szemléletes jelentéssel rendelkezik.

- Az  $\alpha$  alakparaméter, amely talán a négy paraméter közül a legfontosabb, jellemzi a farkak vastagságát (heaviness) és a csúcsosságot (peakedness). Az árfolyamok esetében minél alacsonyabb az  $\alpha$  paraméter értéke, annál nagyobb valószínűséggel következik be extrém áringadozás az adott részvény árfolyamában, ezért annál kockázatosabbnak tekinthető.

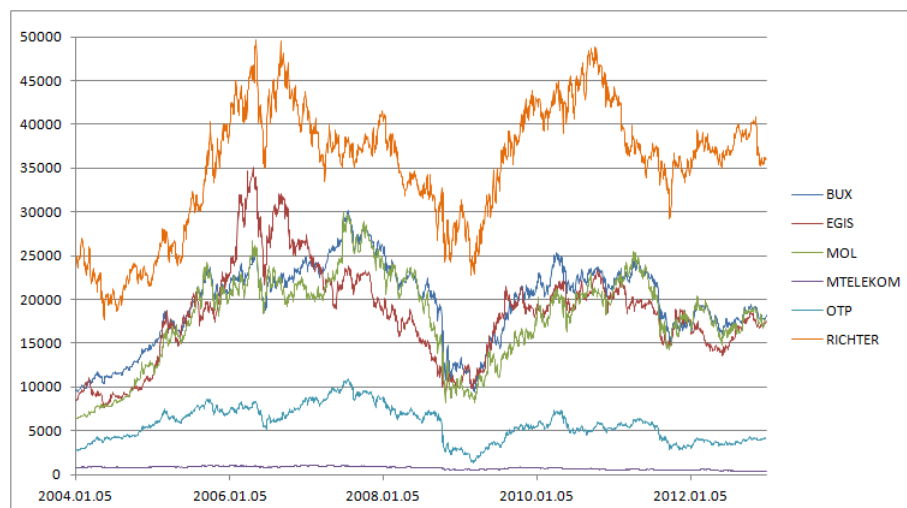
---

J. P. Nolan stabil eloszlások ML módszerrel való paraméterbecslésére, a sűrűség és eloszlásfüggvény közelítésére, stb. kidolgozott, szabadon hozzáférhető szoftvere, mely letölthető a [61] weboldáról.

- A  $\beta$  aszimmetria paraméter írja le az eloszlás ferdeségét: ha értéke negatív, akkor az eloszlás balra ferde, és nagyobb valószínűséggel csökken az árfolyam, míg ha pozitív, akkor jobbra ferde, és emelkedik az árfolyam.
- A  $\gamma$  skálaparaméter és a  $\delta$  helyparaméter a normális eloszlás szórásához és várható értékéhez hasonló szerepet tölt be.
- A  $\delta$  helyparaméter az eloszlás középpontját adja meg, tehát a várható hozamot jelenti. A skálaparaméter a megfigyelések szóródását jellemzi, azaz a változékonyságot (volatilitást), a kockázatosságot írja le.

A dolgozatomban vizsgált adatsorok részvények napi záróár adatait tartalmazzák 2004.01.01. és 2012. 12. 31. között. Az adatok forrása a [www.portfolio.hu](http://www.portfolio.hu) weboldal. A vizsgált részvények: OTP, Richter, Egis, Magyar Telekom, MOL, valamint a BUX Budapesti Értéktőzsde hivatalos indexe. Abban az esetben, amikor az adatsorból hiányzott az aznapi záróár, az árfolyamváltozást az előző napi meglévő adatból számítottam. Az 20. ábra a napi záróár adatokat mutatja.

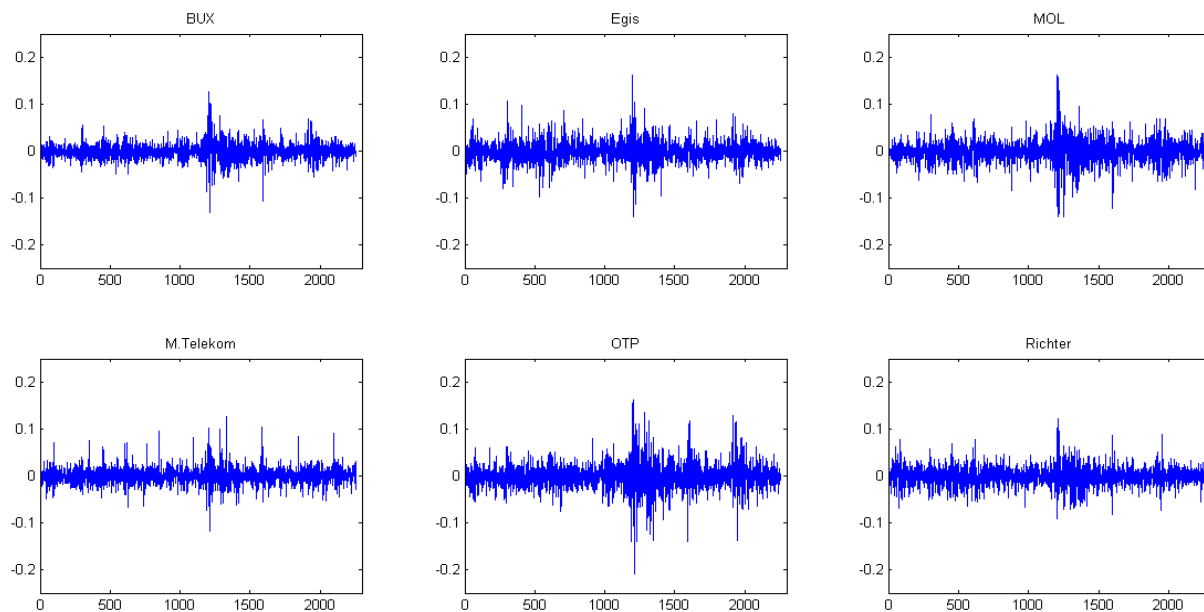
20. ábra. A vizsgált részvények napi záróárainak alakulása 2004.01.01. és 2012. 12. 31. között



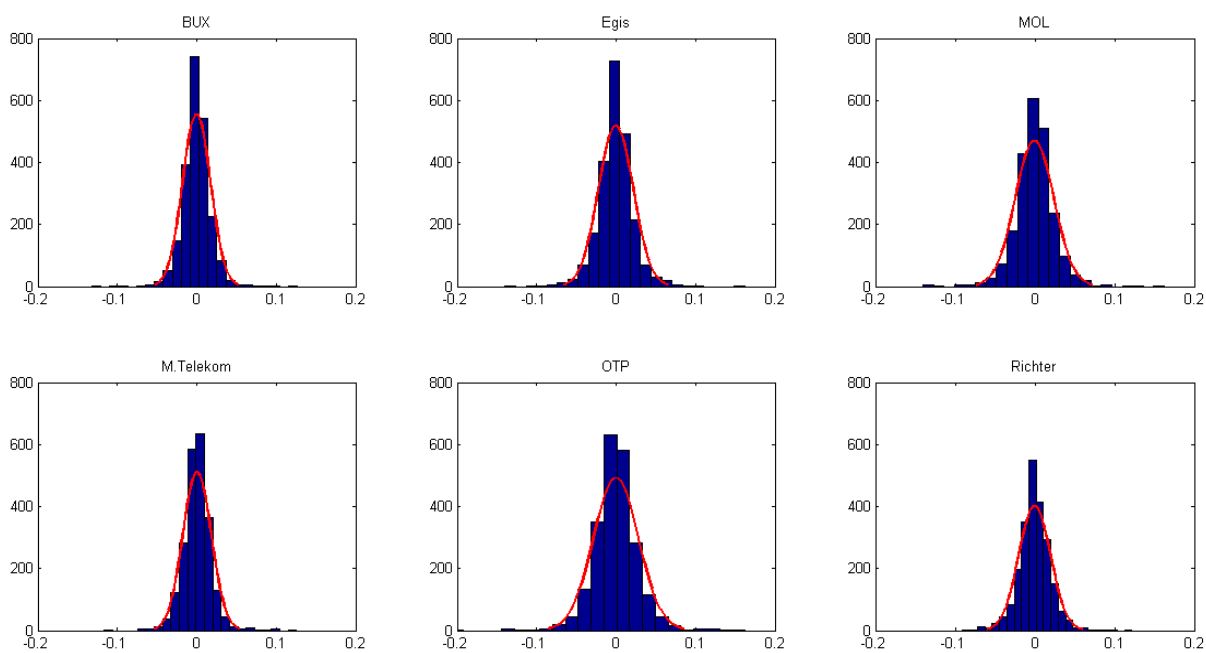
Feltételeztem, hogy a napi árfolyamadatokról képzett minta független, azonos eloszlásból származik. A részvényenkénti adatsorokat két változóra bontottam páros-páratlan sorszám alapján, és  $\chi^2$  teszttel vizsgáltam a két változó függetlenségét. A teszt nem utasította el a függetlenséget 95%-os szignifikancia szinten abban az esetben, amikor a felosztás intervallumainak száma 2-3 volt, de több intervallumra osztás esetén már igen.

A részvényenkénti adatsorokból a (72) formula alapján meghatároztam a napi logaritmikus hozamokat, melyeket a 21. ábra mutat. A hozamok ábráján megfigyelhető a pénzügyi válság hatására az ún. *volatility clustering* jelenség, azaz látszik, hogy a válság kirobbanását követő időben sokkal volatilisabbak lettek a részvények, mindegyik részvény ára drasztikus ingadozásnak volt kitéve. (A válság az ábrák középső részénél, kb. az 1200. megfigyeléstől kezdődően látszik.)

21. ábra. A vizsgált részvények logaritmusos hozamai



22. ábra. A vizsgált részvények logaritmusos hozamainak gyakorisági hisztogramjai



23. táblázat. A PIT módszerrel a logaritmikus hozamokból számított  $\hat{\alpha}$  alak-,  $\hat{\delta}$  hely- és  $\hat{\gamma}$  skálaparaméter becslések

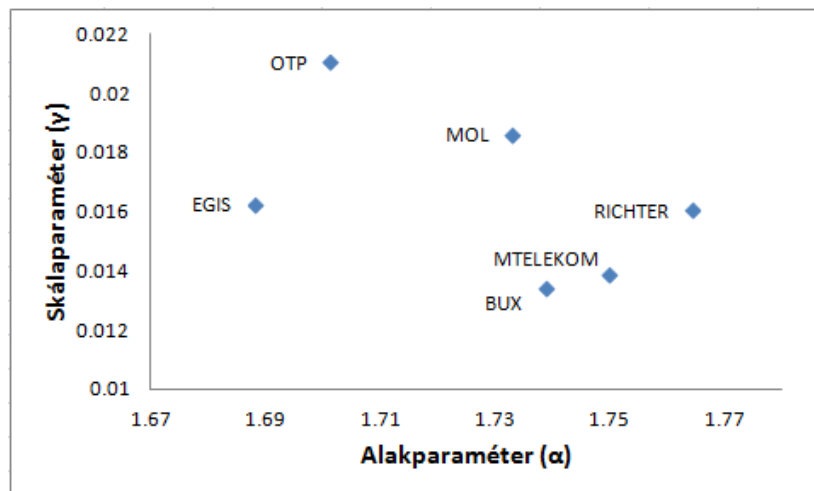
Részvény	$\hat{\alpha}$	$\hat{\gamma}$	$\hat{\delta}$
BUX	1.738710	0.013422	0.000371
EGIS	1.688192	0.016257	0.000356
MOL	1.732882	0.018634	0.000324
MTELEKOM	1.749846	0.013878	-0.000163
OTP	1.701002	0.021068	0.000409
RICHTER	1.764326	0.016072	0.000178

A logaritmikus hozamok eloszlásának vizsgálatához gyakorisági hisztogramon (22. ábra) ábrázoltam az adatokat. A hisztogramokon láthatóak a hozam adatokra illesztett normális eloszlások görbéi is. A hisztogramokból kitűnik, hogy a normális modell nem illeszthető az adatokra, azok eloszlása inkább stabil eloszlást követ.

## 6.2. A BÉT hozamainak PIT becsléssel számított paraméterei

Meghatároztam a PIT becsléssel a minták  $\hat{\alpha}$  alak-,  $\hat{\delta}$  hely- és  $\hat{\gamma}$  skálaparaméter becsléseit a szimmetria feltételezése mellett. A számított paraméterértékeket mutatja a 23. táblázat. Az eredmények szerint az OTP és az Egis valamivel kockázatosabbnak bizonyult, a Richter, Mol és M. Telekom papírjai egy kicsivel stabilabbak, kevésbé kockázatosak. A  $\gamma$  skálaparaméter értéke az OTP esetében a legmagasabb,  $\gamma = 0.0211$ , ami arra utal, hogy a vizsgált papírok közül az OTP-re volt legnagyobb hatása a pénzügyi válságnak. A becsült paraméterek közötti kapcsolat még jobban megfigyelhető, ha a részvények alakparaméterét és skálaparaméterét együtt ábrázoljuk (23. ábra).

23. ábra. A becsült alak- és skálaparaméterek közötti kapcsolat



Mivel a volatilisabb időszakokban az árfolyam változások átlaga eltávolodhat nullától, a logaritmusos hozam és a (71) formula alapján definiált százalékos hozam ebben az időszakban eltérhet. A stabil paramétereket a százalékos hozamok esetén is meghatároztam, azonban az eredmények alapján a teljes vizsgált időszakban megfigyelt eloszlás paramétereiben ez a különbség nem jelent számottevő változást. A dolgozat eredményeit a logaritmusos modellben ismertetem. A (71) formula alapján számított százalékos hozamok eloszlásának becsült paramétereit mutatja a 24. táblázat.

24. táblázat. A PIT módszerrel a százalékos hozamokból számított  $\hat{\alpha}$  alak-,  $\hat{\delta}$  hely- és  $\hat{\gamma}$  skálaparaméter becslések

Részvény	$\hat{\alpha}$	$\hat{\gamma}$	$\hat{\delta}$
BUX	1.738707	0.013425	0.000418
EGIS	1.688284	0.016265	0.000426
MOL	1.731482	0.018630	0.000414
MTELEKOM	1.749941	0.013877	-0.000114
OTP	1.702944	0.021094	0.000526
RICHTER	1.762995	0.016065	0.000244

A STABLE programmal számított paraméter becslések eredményeit tartalmazza a 25. táblázat. Az egyes részvények alakparaméterére a két módszerrel nagyon hasonló eredményt kaptam. A STABLE programmal számított  $\alpha$  paraméter értékek minden esetben egy kicsivel (kb. 1-2 századdal) alacsonyabbak, de kockázatosság szempontjából ugyanaz a sorrend alakul ki. A  $\beta$  ferdeségi paraméter értéke minden részvény esetén nullához közeli, a leginkább szimmetrikus árfolyamú papír az Egis ( $\beta = -0.0037$ ) és az M. Telekom ( $\beta = -0.0070$ ). A skála és helyparaméter becslések is hasonlóan alakultak mindkét módszerrel.

A hely és skálaparaméternek kiszámíthatók a robusztus becslései úgy, mint a medián és a medián abszolút eltérés (Median Absolute Deviation, MAD), illetve a normális modellből kiindulva megbecsülhető az átlag és szórás is. Ezeket a becsléseket tartalmazza a 26. táblázat. Az OTP és Egis részvényekre a magas skálaparaméter érték mellett magas szórást és MAD érté-

25. táblázat. A STABLE programmal a logaritmusos hozamokból számított alak-  $\hat{\alpha}$ , ferdeségi-  $\hat{\beta}$ , hely-  $\hat{\delta}$  és skála-  $\hat{\gamma}$  paraméter becslések

Részvény	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\delta}$
BUX	1.7246	-0.0322	0.00991	0.00043
EGIS	1.6623	-0.0037	0.01212	0.00042
MOL	1.7211	0.0323	0.01376	0.00018
M.TELEKOM	1.7359	-0.0070	0.01016	-0.00005
OTP	1.6904	-0.0852	0.01554	0.00080
RICHTER	1.7458	0.0721	0.01184	0.00012



26. táblázat. A hely- és skálaparaméter robusztus és normális eloszlás szerinti becslései a logaritmusos hozamokból

Részvény	átlag	medián	szórás	MAD
BUX	0.000286	0.000531	0.017436	0.012381
EGIS	0.000327	0.000000	0.021753	0.015366
MOL	0.000448	0.000000	0.024080	0.017187
MTELEKOM	-0.000341	0.000000	0.017790	0.012699
OTP	0.000182	0.000160	0.028241	0.019900
RICHTER	0.000171	0.000000	0.019769	0.014426

27. táblázat. A logaritmusos hozamok illeszkedésvizsgálatára végzett  $\chi^2$  próba p-értékei, tesztstatisztika értékei és szabadsági fokai

Részvény	$N_0$ : normális eo.			$N_0$ : stabil eo.		
	p-érték	tesztstat.	sz. fok	p-érték	tesztstat.	sz. fok
BUX	0.0000	52.2043	2	0.0259	12.7440	5
EGIS	0.0000	121.2358	2	0.0025	18.3700	5
MOL	0.0000	103.3843	2	0.1052	9.0975	5
MTELEKOM	0.0000	51.7739	2	0.0001	26.9211	5
OTP	0.0000	108.4685	2	0.5204	4.2039	5
RICHTER	0.0000	82.8593	3	0.0001	25.7553	5

ket kaptam, ami összhangban van az alacsonyabb  $\alpha$  paraméterrel. Valamint az alakparaméter alapján stabilabb, kevésbé kockázatos Richternél és M. Telekomnál a szórás és MAD is alacsonyabban alakult. A medián három részvélynél (Egis, Mol, M. Telekom) is nulla lett és a többi esetben is nagyon közeli a nullához, így a STABLE program  $\beta$  becslésével is összevetve a szimmetria mellett szóló eredményt kaptam.

A mintának a becsült stabil paraméterű eloszláshoz való illeszkedését Kolmogorov-Smirnov próbával és  $\chi^2$  goodness-of-fit tesztekkel értékeltem. A 27. táblázat a normalitás hipotézise és a becsült  $\alpha$ -stabil eloszláshoz való  $\chi^2$  illeszkedés vizsgálat eredményét mutatja. A próbákat a MATLAB szoftvercsomag *chi2gof* függvényével végeztem el. A táblázatban a tesztstatisztika értékét, a hozzá tartozó p-értéket, valamint a szabadsági fokot tüntettem fel. Az intervallumok száma mind a két esetben kezdetben 10 volt, de a normális nullipotézis esetében össze kellett vonni intervallumokat.

A normális eloszlás esetében a becsült paraméterek száma 2, a stabil eloszlás esetében 4, így a kritikus értékek 95%-os megbízhatósági szinten  $\chi_3^2 = 11.0705$ , illetve  $\chi_3^2 = 7.8147$ , és  $\chi_2^2 = 5.9915$ .

28. táblázat. A logaritmusos hozamok illeszkedésvizsgálatára végzett Kolmogorov-Smirnov próba p-értékei és tesztstatisztika értékei

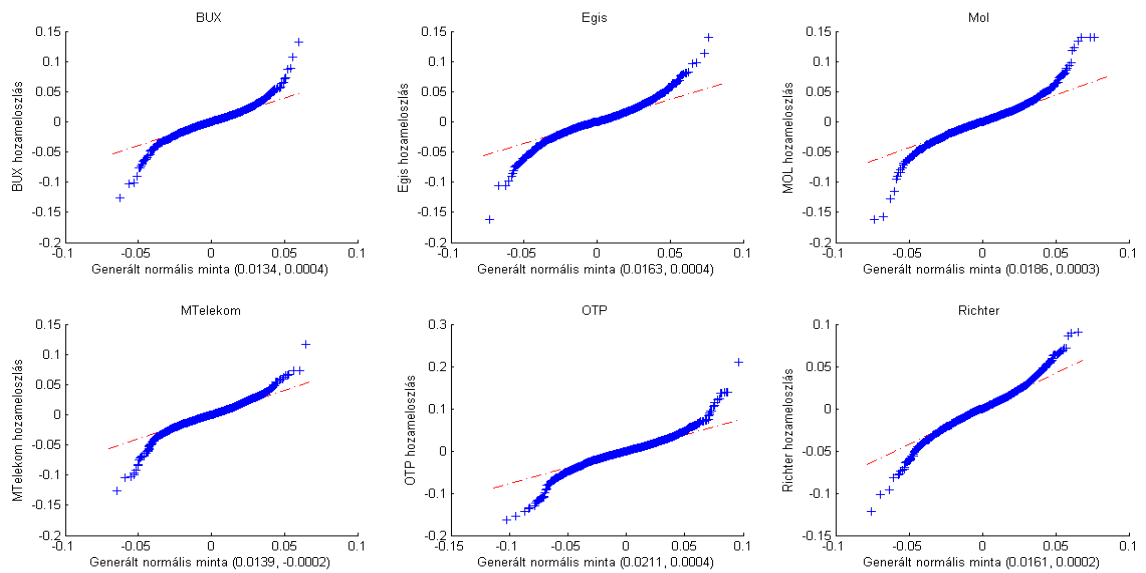
Részvény	Normális eloszlás		Stabil eloszlás	
	p-érték	tesztst.	p-érték	tesztst.
BUX	0.0000	0.0566	0.7804	0.0138
EGIS	0.0000	0.0658	0.2179	0.0221
MOL	0.0000	0.0585	0.8810	0.0123
MTELEKOM	0.0000	0.0563	0.0460	0.0288
OTP	0.0000	0.0641	0.8693	0.0125
RICHTER	0.0002	0.0455	0.2748	0.0209

A teszt elvégzéséhez a hipotetikus stabil eloszlásfüggvények értékeit  $\alpha = 1, \alpha = 1.1, \dots, \alpha = 2$  paraméterű, nagy elemszámú véletlen minták generálásával a minták percentiliseiből határoztam meg. A  $\chi^2$  teszt során a PIT módszerrel becsült  $\alpha$  paramétert kerekítettem a rendelkezésre álló egy tizedesjegy pontos  $\alpha$  értékhez, és ehhez a hipotetikus eloszláshoz képest vizsgáltam az illeszkedést. Az tesztek alapján azt mondhatjuk, hogy a stabil modellt több részvény esetében elfogadhatónak értékelték a tesztek, míg a normalitást egyértelműen el kell utasítanunk.

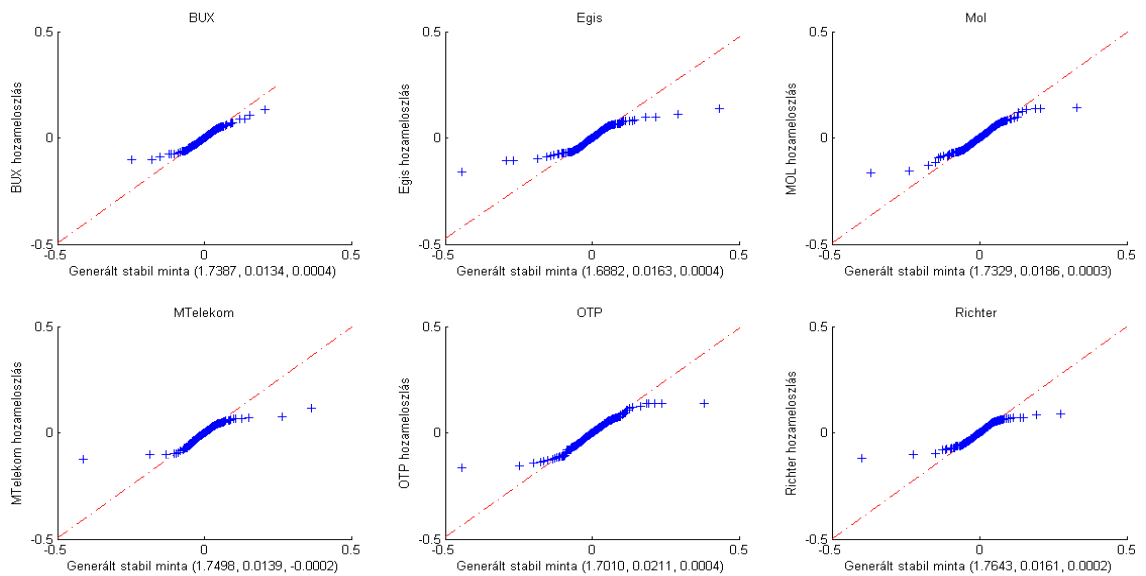
A 28. táblázat tartalmazza az elvégzett Kolmogorov-Smirnov tesztek p-értékeit és a tesztstatisztika értékeket. A 95%-os megbízhatósági szinthez tartozó kritikus érték a normális nullhipotézis esetén 0.00258, míg stabil eloszlású nullhipotézis esetére nem ismert a teszt aszimptotikus viselkedése. A stabil nullhipotézis esetében a p-értékek a normális eloszlásra ismert aszimptotikus eredmények alapján kerültek meghatározásra. A tesztek a Matlab *kstest* függvényével végeztem el. Az eredményekből kitűnik, hogy a hipotetikus  $\alpha$ -stabil eloszlásoktól való eltérés minden esetben kisebb, mint a normális eloszlástól való eltérés, és a teszt csak egy esetben (MTelekom) utasította el a stabil nullhipotézist.

A hozamok empirikus eloszlásfüggvényét a normális eloszláshoz (24. ábra), illetve a becsült  $\alpha$  paraméterű stabil eloszláshoz (25. ábra) illesztve ún. q-q ábrán ábrázoltam. A q-q ábra a megfigyelt és az elméleti, hipotetikus percentilis értékeket ábrázolja. Ha az illeszkedés megfelelő, azaz a hipotézisbeli és az empirikus eloszlás egyezik, akkor a pontok egy egyeneshez illeszkednek. Az elméleti eloszlásokat 10.000 elemű véletlen minták generálásával állítottam elő. Az ábrákon látható, hogy a normális eloszlás illesztése nem megfelelő, az eloszlás szélein jelentős eltérések figyelhetők meg az egyenestől. Az 25. ábrán kevesebb pont tér el az egyenestől az eloszlás széleinél, tehát grafikusan is a várakozásnak megfelelő eredményt kaptam.

24. ábra. Q-Q ábrák a logaritmusos hozamok empirikus eloszlása és a normális eloszlás között



25. ábra. Q-Q ábrák a logaritmusos hozamok empirikus eloszlása és a becsült stabil eloszlás között



### 6.3. A hozamok eloszlásának vizsgálata mozgóablakok segítségével

Elemzésem második felében azt vizsgáltam, hogyan alkalmazhatók az eredmények előrejelzésre. Ehhez változtatható méretű mozgóablakot készítettem, és ábrázoltam, hogyan változtak a paraméterek a vizsgált időperiódusban. Arra a kérdésre is próbáltam választ adni, hogy a 2008–2009-es pénzügyi válság hatása hogyan mutatkozott meg a hozamok eloszlásában. Alapvetően a részvények kockázatosságát egy adott időpontban a hozameloszlásának alakparamétere jellemzi. A hozameloszlás skála- és helyparaméterének ismeretében pedig konfidenciaintervallumot készíthetünk a hozam alakulására. A konfidenciaintervallum nagyságát a skálaparaméter határozza meg.

Minden részvény esetében 50 elemű mozgóablak segítségével becsültem a paramétereket, és ábrázoltam a paraméterek változását az időben. Ez az ablakméret megfelelően kicsi ahhoz, hogy a változások gyorsan érvényesüljenek a becslésértékekben. Hátránya viszont, hogy előfordult olyan eset, amikor a paraméterek nem voltak meghatározhatók a PIT becsléssel az alacsony mintaelemszám miatt.

A 26. ábrán a BUX index adatsorából becsült paraméterek láthatók, egy-egy al-ábrába foglalva, tehát a négy ábra ugyanannak az adatsornak a becsült paramétereit mutatja külön-külön. A legfelső ábra mutatja a logaritmikus hozam adatokat kék színnel, és a hozameloszlásokhoz kiszámolt 95%-os konfidencia intervallumokat. A normális eloszlás szerint számolt intervallumok pirossal, a stabil eloszlás szerint számolt intervallumok zöld színnel vannak feltüntetve, a konfidencia intervallumok természetesen a becsült paraméterek segítségével készültek. Felülről a második ábra az  $\alpha$  alakparaméter változását mutatja. Ha a PIT becslés nem adott eredményt, akkor az alakparaméter értéke hiányzik, ezért szaggatottá válik a becslésértékeket mutató vonal. A harmadik ábrán a PIT módszerrel számított skálaparamétert zöld színnel, a minta tapasztalati szórását a normális eloszlás szerint piros színnel jelöltem. A negyedik, legalsó ábra a PIT becsléssel becsült helyparamétert és a minta átlagát mutatja, a színek megegyeznek a korábbiakkal.

A pénzügyi válság hatása kb. az 1200. megfigyeléstől kezdődően, az ábrák középső részén látható. A négy ábrán egyszerre követhető nyomon az adatsor, és az összes paraméter változása. Mivel a választott ablakméret viszonylag kicsi, ezért a változások gyorsan érvényesülnek a becslésértékekben. A BUX adatsor  $\alpha$  paraméter értéke a válság időszakában a bemutatott ábrák alapján nagyon meredeken csökkent. Az időszakban elért legalacsonyabb értékek 1.3 körüliek voltak, míg az időszak végére az alakparaméter 2 közelébe került.

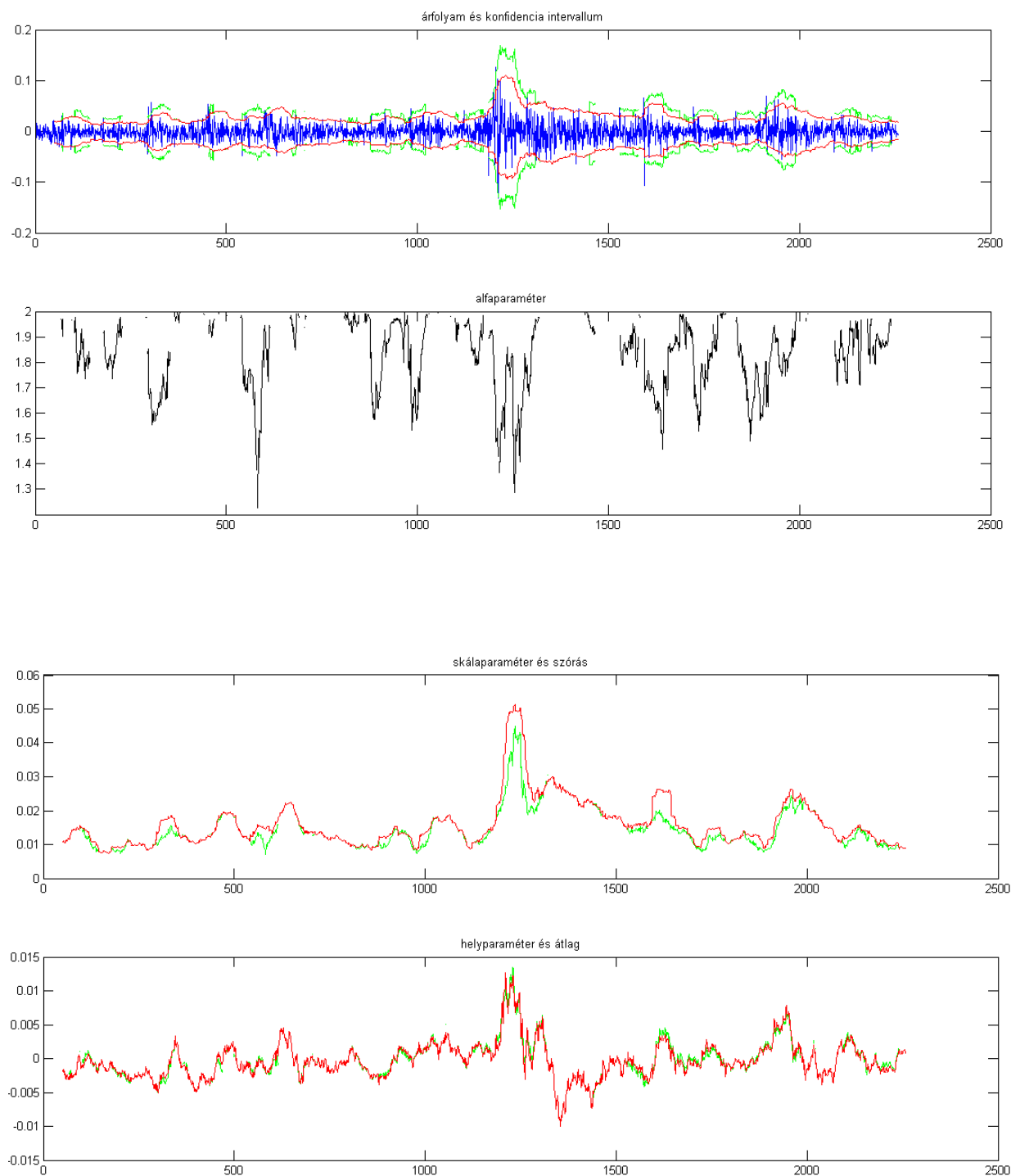
A többi részvény ábráját is elkészítettem, ezek alapján a BUX-hoz hasonlóan alakult az OTP és az Egis hozameloszlásának  $\alpha$  paramétere, míg a többi vizsgált papírnál az alakparaméter nem csökkent kiugróan. Megállapítható, hogy az  $\alpha$  paraméter becslése meglehetősen gyorsan változik. Ebben közrejátszhat a becslési eljárás érzékenysége egy-egy nagyon extrém megfigyelésre, bár ez az érzékenység jóval kisebb, mint a hagyományos módszerek esetében.

A vizsgált időszakban megfigyelhető volt mindegyik részvélynél a szórás és a skálaparaméter növekedése a válság évében, és ezzel egyidejűleg a helyparaméter, azaz a hozam várható értékének csökkenése. A skálaparaméter, illetve szórás értékének emelkedése nem volt hosszútávú, rövid idő alatt visszaálltak a 2008 előtti skála és szórás értékek. A stabil modellel becsült skálaparaméter természetesen mindig alacsonyabb, mint a normális eloszlásból becsült szórás értéke. Azokban az időszakokban a legjelentősebb az eltérés, amikor az alakparaméter jelentősen csökkent. Tehát minél inkább közelítünk a véges szórású esethez ( $\alpha = 2$ ), annál inkább érvényes a normális modell, míg ha az alakparaméter csökken, úgy válik az árfolyam volatilibbá, és emelkednek a szóródást leíró paraméterek értékei.

A mintaelemek függetlenségét a mozgóablakos elemzésben is teszteltem, ugyanazzal a módszerrel, mint a vizsgálat első felében, azaz páros és páratlan index alapján két mintát képeztem az alapmintából, és  $\chi^2$  teszttel vizsgáltam a függetlenségüket. Az 50 elemű ablakoknál általában teljesült a függetlenség kritériuma.

**3. TÉZIS.** *A PIT becslési eljárás alkalmazható valós adatok elemzésére. A Budapesti Értéktőzsde kiválasztott részvényei logaritmikus hozameloszlásának paraméterbecslése alapján megállapítottam, hogy a becsült paraméterekkel rendelkező stabil eloszlás jobban illeszkedik a hozam adatokra, mint a normális eloszlás. A paraméterek időbeni változását elemezve megállapítottam, hogy a becsült alak- és skálaparaméterek jól tükrözik az egyes részvények kockázatosságának alakulását, amely a 2008-2009-es pénzügyi válság hatására jelentősen megnövekedett ebben az időszakban.*

26. ábra. Konfidencia intervallum normális és stabil eloszlás alapján -  $n = 50$ , BUX, 2004.01.01. – 2012. 12. 31.



## 7. A PIT becslés implementációja és a segédprogramok

A következő fejezetben áttekintem a bemutatott kutatási eredmények eléréséhez szükséges informatikai munkát, és összefoglalom a statisztikai ötlet megvalósítását lehetővé tevő programozási megoldásokat. Az elkészült programok mindegyike MATLAB függvény (.m fájl), amelyek egymáshoz csak lazán kapcsolódnak, azok tetszőlegesen csoportokba (mappákba) szervezhetők. A függvényeket két nagyobb csoportra bontva ismertetem.

Az első csoport a PIT eljárást kiszámító függvények csoportja, tehát azok a függvények, amelyek a becslés értékek meghatározásában közvetlenül részt vesznek. Ezek a MATLAB függvények olvashatók dolgozatom Függelékében.

A második csoport a segédfüggvények – az előzőnél sokkal nagyobb – csoportja, amelyben minden más elkészült függvényt összegyűjtöttem. Ezek a függvények lehetővé tették az algoritmus gyorsítását és az implementáció megkönnyítését célzó törzfüggvényes közelítés meghatározását, amelynek módszere, lépései a dolgozatom 4.6. szakaszában található. A segédfüggvények csoportja tartalmazza a stabil eloszlású egyváltozós és többváltozós véletlen számok generálásához készített függvényeket, a véletlenszám generálás algoritmusait dolgozatom 5.1. szakaszában mutattam be. Ide sorolhatók a PIT módszer megbízhatóságának, pontosságának, és a becslés aszimptotikus normalitásának teszteléséhez elvégzett statisztikai vizsgálatokhoz (Monte-Carlo szimulációk, dolgozatom 5.2. - 5.4. szakasza) készített programkódok. Valamint a segédfüggvények között áttekintem a BUX adatsorok elemzéséhez (alkalmazási példa, dolgozatom 6. fejezete) írt automatizált elemző szkripteket is. Ezek közül a MATLAB függvények közül néhány kiválasztott programkód olvasható a Függelékben.

A MATLAB programcsomag szkriptnyelve nem típusos, azaz minden változót alapértelmezésként mátrixként kezel, amelynek tetszőleges típusú elemei lehetnek. Az .m fájlok ismertetésénél külön jelzem, ha valamelyik paraméter csak oszlopvektor, vagy skaláris érték lehet. A programkódok alapvetően a kutatásomban kitűzött statisztikai célok megvalósítását szolgálják, a programok nem felhasználóknak készültek, ezért a nem megfelelő paraméterekkel való függvényhívások következményeit nem teszteltem, nem zártam ki az ebből eredő lehetséges hibákat.

Az elkészített MATLAB függvényeket dolgozatomhoz mellékeltem. A függvények a fenti felsorolásnak megfelelően, funkció szerint vannak csoportosítva a mellékelt CD/DVD-n. Ezen a függvényeken kívül elkészült még számos olyan függvény, amely az adatok MATLAB-ba történő gyors beolvasását, mozgatását, mentését, konvertálását tette lehetővé, de ezeket a programokat nem ismertetem.

### 7.1. A PIT paraméterbecslést kiszámító MATLAB függvények

A 4.4. szakaszban bemutatott PIT algoritmus megvalósítását (implementációját), azaz a becslésértékek kiszámítását lehetővé tevő függvényeket tartalmazza a 29. táblázat. Az *alphasearch* függvény oszlopvektorban megadott véletlen mintaelemekből kiszámítja a három becsült stabil paramétert ( $\alpha$ ,  $\gamma$ ,  $\delta$ ), azaz az intervallumfelezéses eljárást valósítja meg. Összesen 18 iterációt hajt végre, azaz a skálaparaméter görbék metszéspontját  $2^{-18}$  pontossággal adja meg. Az utolsó iterációban a fennmaradó intervallum hossza  $2^{-18}$ . Ha az intervallumfelező algoritmus kezdőpontjaiban ( $\alpha = 1$  és  $\alpha = 2$ ) meghatározott skálaparaméter becslések különbségei nem váltanak előjelet, az azt jelenti, hogy az  $[1, 2]$  intervallumban nincs metszéspont, ezért az iteráció nem folytatódik, ebben az esetben a paraméterek értéke -1. Tehát a -1 érték jelzi, hogy az

input oszlopvektorban tárolt mintákra a PIT becslés nem ad érvényes paraméterbecsléseket.

A *pingpong1* és *pingpong2* függvények a ping-pong módszerrel a Cauchy és a normális eloszlás használatával meghatározzák rögzített  $\alpha$  esetén a hely- és skálaparaméter értékét. Az iterációból való kilépési feltétel a skálaparaméter konvergenciája, a megkövetelt pontosság  $\epsilon = 10^{-8}$ . A pontosság jelenleg nem változtatható paraméterben. Ha a skálaparaméter becslése nem konvergál, akkor 100 iteráció után mindenképp kilépünk a ciklusból, ekkor -1 értékű lesz mindkét becült paraméterérték.

A *B1* és *B2* függvények a törtfüggvény közelítést tartalmazzák, a 4.6 szakaszban bemutatott együtthatók alapján. A polinomok kiszámítása a MATLAB *polyval* függvényével történik, amely a polinomok kiértékelését a Horner séma szerint számítja. Az algoritmus implementálásához szükség van a normális eloszlásfüggvény értékének pontos kiszámítására. A normális eloszlásfüggvény értékei az ismert algoritmusok, például az inverz módszer alapján meghatározhatók, de a MATLAB programcsomag beépített függvénye (*normcdf*) alkalmasnak bizonyult a megfelelő pontosságú értékek kiszámítására.

29. táblázat. A PIT algoritmus implementációjának függvényei

függvény	funkció	input paraméterek	output paraméterek
<i>alphasearch</i>	intervallumfelezéses eljárás	mintaelemek oszlopvektorban	$\alpha, \gamma, \delta$ paraméterek becslései
<i>pingpong1</i>	ping-pong módszer a Cauchy eloszlással, rögzített $\alpha$ esetén megadja a hely- és skálaparamétert	mintaelemek oszlopvektorban, rögzített $\alpha$	becsült hely- és skálaparaméter
<i>pingpong2</i>	ping-pong módszer a normális eloszlással, rögzített $\alpha$ esetén megadja a hely- és skálaparamétert	mintaelemek oszlopvektorban, rögzített $\alpha$	becsült hely- és skálaparaméter
<i>B1</i>	$\mathcal{B}_1$ függvény törtfüggvény közelítése	$\alpha$ alakparaméter	$\mathcal{B}_1(\alpha)$
<i>B2</i>	$\mathcal{B}_2$ függvény törtfüggvény közelítése	$\alpha$ alakparaméter	$\mathcal{B}_2(\alpha)$



## 7.2. A törtfüggvény közelítés meghatározásának segédfüggvényei

A törtfüggvény közelítés meghatározását lehetővé tévő segédfüggvényeket a 30. táblázatban foglaltam össze. A törtfüggvény közelítés lépéseiről dolgozatom 4.6. szakasza szól.

A *B1kozelites* és *B2kozelites* függvények segítségével a törtfüggvény közelítés alappontjaiban felvett függvényértékeket számítottam. Mindkét függvény 50-szer ismételt 5 milliós nagyságú  $\alpha$ -stabil mintával egy  $\mathcal{B}(\alpha)$  függvény közelítést. A választott alappontok, amelyekben a közelített értékeket meghatároztam 1.05, 1.1, ..., 1.95 voltak. Az  $\alpha = 1$  és  $\alpha = 2$  pontokban ismert a függvények értéke, amely  $1/12$ , illetve a MAPLE programmal nagy pontossággal meghatározható, lásd 4.6. szakasz.

A *ractfgv\_kereses* függvény a racionális törtfüggvény közelítések előállítását szolgálja. Az input paraméterek az alappontokban meghatározott függvényértékek mátrixa és egy olyan mátrix, amely a kipróbálásra váró törtfüggvény közelítések jellemzőit írja le. A függvényértékek mátrixa három oszlopot tartalmaz, az első az alappontokat, a második a Cauchy eloszlás használatával, a harmadik a normális eloszlás használatával meghatározott függvényértékeket. A második mátrix, amelyet a függvény vár, a készítendő 12 törtfüggvény esetet írja le, első oszlopában található az ismeretlenek (együtthatók), második oszlopában a kiválasztott alappontok (egyenletek) száma (lásd dolgozatom 2. táblázata).

A *ractfgv\_kereses* függvény a kiszámított közelítések együtthatóit, és az alappontonkénti eltérések mátrixát esetenként csoportosítva fájlba írja ki. A függvény visszatérési értéke a 12 törtfüggvény közelítés  $\mathcal{B}_1$ -re és  $\mathcal{B}_2$ -re egyaránt meghatározott abszolút értékben vett maximális eltérések értéke, azaz a hibafüggvény normájának közelítése. A hibafüggvényt valójában nem ismerjük, csak a szimulált alappontbeli függvényértékekhez képest vizsgáltam a közelítő függvények eltérését.

30. táblázat. A törtfüggvény közelítés meghatározásának segédfüggvényei

függvény	funkció	input paraméterek	output paraméterek
<i>B1kozelites</i>	az alappontokban felvett $\mathcal{B}_1$ függvényértékeinek közelítése	-	fájlba írja az eredményt
<i>B2kozelites</i>	az alappontokban felvett $\mathcal{B}_2$ függvényértékeinek közelítése	-	fájlba írja az eredményt
<i>ractfgv_kereses</i>	a $\mathcal{B}$ függvények legmegfelelőbb racionális törtfüggvény közelítését 12 variációval automatizáltan keresi	kozMatrix: alappontbeli szimulált függvényértékek, tervMatrix: törtfüggvény variációk fokszáma és a választott alappontok száma	hibafüggvény becslések (maximális eltérések) mátrixa

## 7.3. Függvények véletlenszámok generálásához

Az egyváltozós és többváltozós véletlenszám generálás függvényeit foglalja össze a 31. táblázat. Az implementált módszerek és képletek részletesen dolgozatom véletlenszám generálással foglalkozó 5.1. szakasza tartalmazza.

A *zol* függvény a Zolotarev-féle véletlenszám generálás képletét tartalmazza, amely szimmetrikus eloszlás generálására használható a (60) képlet alapján. A *zol* függvény standardizált változókat készít, a (60) képlet után szereplő megjegyzés alapján. Ha véletlen vektort szimulálunk, akkor a komponensek azonos  $\alpha$  paraméterrel rendelkeznek és függetlenek. A *zol* függvény paraméterei a generálandó minta elemszáma, véletlen vektor készítéséhez a komponensek száma, az alakparaméter, a hely- és skálaparaméter értéke. A visszatérési értéke a generált minta mátrixa.

A többváltozós véletlenszám generálás módszerei közül elkészítettem az Uchaikin és Zolotarev [79] által bemutatott három dimenziós gömbszimmetrikus eloszlás generálását megvalósító programot a (65) formulák alapján. A *mintakeszit* függvény adja meg a háromdimenziós mintát egy *elemszam*  $\times$  3 típusú mátrixban. A *mintakeszit* függvény a *gombszim* függvény által generált véletlen irányok (pontok egyenletes eloszlás szerint az egységgömb felszínén) és a *rfuggveny* függvény által számított sugárhosszok elemenkénti szorzatát számítja ki. A *gombszim* függvény a *pontszam* input paraméterben megadott elemszámú polárkoordinátás gömbszimmetrikus szimulálás szerinti irányt készíti el. Az *rfuggveny* függvény a polárkoordinátás gömbszimmetrikus szimulálásban használt, *pontszam* paraméterben megadott darabszámú sugárhosszt szimulálja, input paraméter még az  $\alpha$  értéke, mivel a hosszúságok függenek az alakparamétertől, lásd dolgozatom 5.1. szakasza, (65) és (66) képletek. Az *etal* és *eta2* .m fájlok a sorfejtés közelítését leíró segédfüggvények.

31. táblázat. Függvények véletlenszám generáláshoz (.m fájlok)

függvény	funkció	input paraméterek	output paraméterek
<i>zol</i>	véletlen minta generálása a Zolotarev formulával	n: minta elemszám, dim: komponensek száma, alpha: minta alakparamétere, scale: skálapar., location:helypar.	a generált $n \times dim$ mátrix
<i>mintakeszit</i>	3 dimenziós gömbszimmetrikus eloszlás szimulálása Uchaikin és Zolotarev alapján	pontszam: készítendő minta-elemszám, alpha: $\alpha$ paraméter	3 dimenziós minta mátrixa
<i>gombszim</i>	gömbszimmetrikus szimulálás irányait adja	pontszam: készítendő irányok száma	véletlen irányok mátrixa
<i>rfuggveny</i>	gömbszimmetrikus szimulálásban az $R$ sugarak hosszát adja	pontszam: készítendő sugarak száma, alpha: $\alpha$ alakparaméter	sugarak mátrixa
<i>etal</i>	segédfüggvények a háromdimenziós generáláshoz	$u$ : függvény argumentum, alpha: $\alpha$ alakparaméter	$\eta_1(u, \alpha)$
<i>eta2</i>	segédfüggvények a háromdimenziós generáláshoz	$u$ : függvény argumentum, alpha: $\alpha$ alakparaméter	$\eta_2(u, \alpha)$

#### 7.4. A valós árfolyamok elemzésére készített MATLAB függvények

Ebben a szakaszban ismertetésre kerülő függvények a BÉT részvény árfolyamok elemzéséhez (dolgozatom 6. fejezete) voltak szükségesek. A függvények lehetővé tették az adatok beolvasását, konvertálását, az elemzés elvégzését a stabil paraméterek meghatározásán keresztül, majd az ábrák, összefoglaló táblázatok elkészítését. A következőkben csak néhány kiválasztott MATLAB függvényt ismertetek.

A *BUXelemez* függvény a BUX alkalmazási példa részvény adatait elemzi. A vizsgált részvények az BUX, Egis, Mol, MTelekom, OTP, Richter voltak, a program 6 mintát tartalmazó cellamátrixot vár bemenő adatként, a cellamátrixok tartalmazzák az előzetesen kiszámolt logaritmikus vagy diszkrét hozamok adatsorait. (A cellamátrix olyan MATLAB adatstruktúra, amelynek elemei maguk is mátrixok lehetnek.) A függvény a három stabil paramétert becsüli, a kiszámított paramétereket soronként tárolja, az eredmény egy  $6 \times 3$  mátrix a becsült paraméterekkel.

A *diszkretreturn* és *logreturn* függvények a (71) és (72) képletek alapján kiszámítják a diszkrét és a logaritmikus hozamokat a részvények napi árfolyamadataiból. A *www.portfolio.hu* weboldalról letöltött adatsorokban a második oszlop tartalmazta a napi záróár adatokat, ezért a hozamszámító függvények egyelőre csak ezekből az oszlopokból dolgoznak.

A *hozamstat* függvény a mediánt, átlagot, szórást és a medián abszolút eltérést számolja ki a diszkrét és logaritmikus hozamok cellamátrixaiból. A MATLAB beépített függvényeit (*mean*, *std*, *median*, *mad*) használja a statisztikai jellemzők kiszámításához. Eredményként egy latex formátumú táblázatot készít.

A *fuggetlenseg* függvény azt teszteli, hogy egy hozamadatokból álló minta elemei egymástól függetlennek tekinthetők-e. Ez annak a feltétele, hogy paraméterbecslésre használhassuk az adatsort, mint a hozamok eloszlására vonatkozó, független azonos eloszlású megfigyelésekből álló mintát. Az ellenőrzést úgy végzi, hogy a mintavektort két részmintára bontja páros és páratlan indexek alapján, majd a két mintának a függetlenségét teszteli  $\chi^2$  illeszkedésvizsgálat próbával. Kiszámítja a kritikus értéket, és a hozzá tartozó p-értéket. A p-érték számításához a MATLAB beépített *chi2cdf* függvényével meghatározza a statisztikához tartozó eloszlásfüggvény értékét, és kivonja 1-ből.

A *mozgoablak* függvény a stabil paramétereket becsüli az *alphasearch* függvény hívásával egy hozam adatsorból úgy, hogy meghatározott *ablakhossz* egymást követő mintaelemet tekint egy mintának, majd egy mintaelemmel mindig előre tolva a mozgóablakot újra becsüli a paramétereket. A *mozgoablak* függvény a stabil paramétereken kívül ugyanezen minták átlagát és szórását is kiszámítja. A becsült stabil paraméterek alapján és a normális eloszlás szerint becsült átlag és szórás paraméterek alapján kiszámítja a 95%-os konfidenciaintervallum alsó és felső határát. Végül összesen négy ábrát készít el. A függvényvel készült konfidenciaintervallumokat tartalmazó ábra dolgozatom 26. ábrája.

#### 7.5. A statisztikai vizsgálatokhoz, szimulációkhoz készített MATLAB függvények

A 33. táblázatban foglaltam össze azokat a legfontosabb függvényeket, amelyeket a statisztikai vizsgálatok és szimulációk elkészítéséhez készítettem. A módszerek implementálásánál csak kevés esetben támaszkodhattam a MATLAB programcsomag rendelkezésre álló (beépített) függvényeire, ugyanis a többváltozós normális eloszlás teszteléséhez nem tartalmaz a MATLAB ilyen függvényt. Az egyváltozós normalitás teszteléséhez használt Sarkadi próbát szintén

32. táblázat. A valós árfolyamok elemzésére készített MATLAB függvények

függvény	funkció	input paraméterek	output paraméterek
<i>BUXelemez</i>	A három stabil paramétereket becsüli a részvény árfolyamadatak alapján	a hozamok cellamátrixa	a becsült paraméterek $6 \times 3$ típusú mátrixa
<i>diszkretreturn</i>	A részvényárfolyamok záróáradataiból kiszámítja a diszkrét hozamokat	A: árfolyamadat mátrix	a záróárak diszkrét hozama
<i>logreturn</i>	A részvényárfolyamok záróáradataiból kiszámítja a logaritmi-kus hozamokat	A: árfolyamadat mátrix	a záróárak logaritmi-kus hozama
<i>hozamstat</i>	A mediánt, átlagot, szórást és a medián abszolút eltérést számolja ki a hozamok mátrixaiból	logR: logaritmi-kus hozamok, disR: diszkrét hozamok cellamátrixai	eredmény latex táblázatos formában
<i>fuggetlenség</i>	A hozam adatok elemeinek függetlenségét teszteli	x: minta, bin: intervallumok száma a $\chi^2$ -próba során	a $\chi^2$ -próba eredménye: döntés, szabadsági fok, p-érték
<i>mozgoablak</i>	Mozgóablakos módszerrel ábrát készít a becsült paraméterekről, az árfolyam adatokról, és konfidenciaintervallumot rajzol hozzá	adat: hozam adatsor, ablakhossz: a mozgóablak mérete	ábrák

implementáltam.

A szimulációkhoz végrehajtásához készített programkódokat nem részletezem, mert ezek többnyire egyszerű, sok egymásba ágyazott ciklust tartalmazó programok. Ezeknél az eljárásoknál a futtatások száma, és a rendkívül magas számítási igény jelenti a nehézséget. Egy-egy szimuláció elvégzése több órát vesz igénybe egy átlagos teljesítményű PC-n.

A *multivnormality* és a *DoornikHansen* függvények a statisztikai vizsgálatokat leíró 5.3. szakaszban bemutatott Mardia-féle ferdeség és lapultság mutatókon alapuló tesztet, valamint a Doornik és Hansen szerzők által bemutatott módszert valósítják meg. Mindkét függvénynek a többváltozós megfigyelések mátrixa az input paramétere.

A Mardia-féle teszt esetében először meg kell határozni a Mahalanobis távolságokat ( a (67) formula szerint), majd ezek segítségével a többváltozós ferdeség és a többváltozós lapultság értékét ( (68) és (69) formulák). A ferdeség tesztstatisztikáját a megfelelő szabadsági fokú  $\chi^2$  eloszlással szemben, a lapultság tesztstatisztikáját a standard normális eloszlással szemben teszteljük. A megfelelő eloszlásfüggvény értékeket (*chi2cdf* – a  $\chi^2$  eloszlás eloszlásfüggvénye, a *chi2inv* – a  $\chi^2$  eloszlás eloszlásfüggvényének inverze, *normcdf* – normális eloszlásfüggvény, *norminv* – normális eloszlásfüggvényének inverze) a MATLAB beépített függvényeivel számoltam.

A Doornik és Hansen által bemutatott eljárás esetén a többváltozós minta transzformáltját készítjük el a változók korrelációs mátrixának sajátértékei és sajátvektorai segítségével a (70) formula alkalmazásával. A sajátértékek és sajátvektorok számolásához a MATLAB beépített *eigs* függvényét használtam. A tesztstatisztika ennél a módszernél is  $\chi^2$  eloszlást követ, így a  $\chi^2$  eloszlás MATLAB függvényeinek segítségével meghatároztam a próba eredményét: a tesztsta-

tisztika értékét, a döntést és a p-értéket.

33. táblázat. A PIT becslés normalitásvizsgálatának néhány MATLAB függvénye

függvény	funkció	input paraméterek	output paraméterek
<i>sarkadiID</i>	Sarkadi-próba mintaelem transzformációját határozza meg	$X$ vektor: egyváltozós minta	$a$ transzformált minta
<i>multivnormality</i>	a Mardia-féle többváltozós ferdeség és lapultság értékeket adja meg	$X$ : $n \times p$ típusú mátrix, a többváltozós minta	fájlba írja a tesztstatisztika értékét, a döntést és a p-értéket
<i>DoornikHansen</i>	Doornik és Hansen módszer többváltozós normalitás tesztelés	$X$ : $n \times p$ típusú mátrix, a többváltozós minta	eredmények: döntés, a statisztika értéke, p-érték

**4. TÉZIS.** *Elkészítettem a PIT paraméterbecslési módszert megvalósító MATLAB függvényeket, valamint a módszer tesztelését, szimulációkkal történő statisztikai vizsgálatát, és valós adatsorokra való alkalmazásának lehetőségét megteremtő programkódokat, segédfüggvényeket. A módszer hatékonyságának, megbízhatóságának elemzésével megmutattam, hogy az eljárás hasonló performancia tulajdonságokkal rendelkezik, mint a szakirodalomban ismert módszerek. A paraméter becslések külön-külön egyváltozós és együttes normalitását szimuláció sorozatban vizsgáltam.*

## 8. Összefoglalás, tézisek

### 8.1. Összefoglalás és javasolt kutatási irányok

Elkészítettem egy az M- becslések csoportjába tartozó új becslési eljárást, amely szimmetrikus stabil eloszlások együttes paraméterbecslésére használható.

A bemutatott új módszer pozitívumai, használatának előnyei:

- A módszer nem használja közvetlenül az ismeretlen  $\alpha$ -stabil sűrűség-, eloszlás-, vagy karakterisztikus függvényt.
- Futási időben az algoritmus nem használ numerikus integrálást, így gyorsabb mint az integrálást használó módszerek, pl. a maximum likelihood módszer.
- Egyszerűen implementálható bármilyen programozási nyelven, amely jelentős előnye a többi paraméterbecslési eljárásához hasonlítva.
- A módszer gyakorlati hasznosításához minden közelítés, függvény, algoritmus rendelkezésre áll, nincs szükség további segédkonstansokra, táblázatokra.
- A pontosság beállítható, így ha nincs szükség nagyon pontos eredményre, akkor az algoritmus gyorsítható.
- A bemutatott becslési eljárás valós feladatokra alkalmazható.

A módszer hatékonyságát szimulációs vizsgálatok segítségével igazoltam, valamint összehasonlítottam a létező módszerekkel a hatékonyság (performance) jellemzőket.

A bemutatott eljárást valós adatokra alkalmaztam. A becslt paraméterek alapján következtéseket tudtam levonni a pénzügyi válság részvényárfolyamokra gyakorolt hatására vonatkozóan. A bemutatott eljárás hasznos eszköznnek bizonyult előrejelzéshez, modell illesztéshez.

A lehetséges további kutatási irányokat az alábbiakban foglalom össze.

- A PIT becslés aszimptotikus normalitásának igazolása: a szimulációs vizsgálatokkal empirikusan igazoltam, hogy nagy minták esetén az aszimptotikus normalitás teljesül. Az állítás elméleti bizonyítása lehetséges kutatási irány.
- A szimulációs vizsgálatokkal kapott kovariancia mátrixok elemzésével lehetséges a paraméterek becslésének együttes eloszlásának további vizsgálata.
- A skálaparaméter görbék helyzete, egymáshoz való viszonya a tapasztalatokból származik. A skálaparaméter görbék, a monotonitás, a metszéspont létezésének elméleti bizonyítása megoldandó feladat.

- A racionális törtfüggvényeket az  $\alpha \in [1, 2]$  intervallumon közelítettem, mert a stabil eloszlásnak ebben az intervallumban léteznek a várható értékei. A numerikus közelítés meghatározható alacsonyabb alakparaméterű stabil eloszlások esetén is.
- A szakirodalomban ismert paraméterbecslési módszerek implementálása lehetővé tenné az összehasonlító szimulációk egységes futtatási környezetben, egységes véletlenszám generálással való vizsgálatát. Ezzel a módszerek futási idő, bonyolultság, implementálhatóság szempontjából összehasonlíthatóvá válnának.
- A paraméterbecslési eljárás bármilyen valós, feltételezeten stabil eloszlású adatsorra alkalmazható, ezért a módszer más tudományterületek alkalmazott kutatásaiban használható.
- Az illeszkedésvizsgálatra használt módszerek (pl.  $\chi^2$ -próba, KS. próba) feltételezik a tesztstatisztika eloszlásának ismeretét. Ha stabil eloszlás szerepel a nullhipotézisben, akkor a normális eloszlásra meghatározott kritikus érték táblázatok nem érvényesek. A stabil alternatívákkal szembeni illeszkedésvizsgálat témájában született szimulációs tanulmányok a közelmúltban. A különböző illeszkedésvizsgálat próbák hatékonysága stabil eloszlások esetén még nem kidolgozott terület. A bemutatott eljárást kritikus értékek meghatározására, erőfüggvény szimulációs vizsgálatokra lehetne használni. A témában született szimulációs tanulmány Csendes [S11].
- A részvényárfolyamok elemzésénél érdekes kérdés, hogy mennyi ideig érvényesek a becsült paraméterek alapján meghatározott konfidenciaintervallumok, azaz hány nap után lép ki az árfolyam a meghatározott konfidenciaintervallumból. Érdekes lenne vizsgálni továbbá azt is, hogy az érvényességi idő és a becsült paraméterek között milyen kapcsolat mutatható ki.
- A pénzügyi válság hatása vizsgálható részletesebben is több hazai és külföldi részvény, illetve más pénzügyi instrumentum árfolyama alapján.
- A meghatározott konfidenciaintervallum alsó végpontja (az ellentettje) tulajdonképpen a VaR kockázatosított értéket adja meg. A paraméterek ismeretében meghatározható a részvény vagy portfólió CVaR mutatója is.
- A meghatározott paraméterek alapján a portfólió kiválasztási feladat megoldása. Olyan program készítése, ami az árfolyamadatok függvényében a paraméterek változása esetén az időben újrasúlyozza a portfóliót.

## 8.2. Tézisek

**1. TÉZIS.** *Ismert  $\alpha$  alakparaméterű eloszlástípus esetén az  $M$ -becslés használható a stabil eloszlások hely- és skálaparaméterének becslésére. Ha más alakparaméternek megfelelő  $F_{0,\alpha}$  eloszlásfüggvényt választunk a súlyfüggvényben, a segítségével meghatározott  $\mathcal{B}$  érték felhasználásával a hely- és skálaparaméter torzítatlan becslését adjuk ( $1 \leq \alpha \leq 2$ ).*

**2. TÉZIS.** *Ha az alakparaméter sem ismert, akkor a Cauchy és a normális eloszlás alapján a becsült skálaparaméterek összehasonlításával megadható az alakparaméter becslése és így a szimmetrikus stabil eloszlás alak-, hely- és skálaparamétere egyszerre becsülhető.*

**3. TÉZIS.** *A PIT becslési eljárás alkalmazható valós adatok elemzésére. A Budapesti Értéktőzsde kiválasztott részvényei logaritmikus hozameloszlásának paraméterbecslése alapján megállapítottam, hogy a becsült paraméterekkel rendelkező stabil eloszlás jobban illeszkedik a hozam adatokra, mint a normális eloszlás. A paraméterek időbeni változását elemezve megállapítottam, hogy a becsült alak- és skálaparaméterek jól tükrözik az egyes részvények kockázatosságának alakulását, amely a 2008-2009-es pénzügyi válság hatására jelentősen megnövekedett ebben az időszakban.*

**4. TÉZIS.** *Elkészítettem a PIT paraméterbecslési módszert megvalósító MATLAB függvényeket, valamint a módszer tesztelését, szimulációkkal történő statisztikai vizsgálatát, és valós adatsorokra való alkalmazásának lehetőségét megteremtő programkódokat, segédfüggvényeket. A módszer hatékonyságának, megbízhatóságának elemzésével megmutattam, hogy az eljárás hasonló performancia tulajdonságokkal rendelkezik, mint a szakirodalomban ismert módszerek. A paraméter becslések külön-külön egyváltozós és együttes normalitását szimulációval vizsgáltam.*



## 9 Summary

### 9.1 Introduction

Stable distribution family has been widely investigated since its introduction by Levy [40]. The theoretical importance of stable laws is inevitable since the distribution family provides the only possible solution to the generalization of the central limit theorem (CLT). With the assumption of a common finite variance, limiting distribution for the sum of independent, identically distributed (iid) random variables is the normal law. The generalization of CLT called the domain of attraction problem arises by summing iid variables with infinite variance. Detailed works on the field are Gnedenko and Kolmogorov [25], Petrov[65], Feller [21].

In data analysis heavier tails are usually realized when a great number of observations are aggregated with very high or nearly infinite variance. It seems reasonable to assume a relation between these characteristics of data sets and the generalized CLT. If such data are modelled application of stable laws should be considered. Data sets of this sort are collected in studies in the field of financial mathematics, e.g. price changes in high frequency trading, signal processing, or measuring the data transfer of Internet traffic.

Stable distributions have heavy tails which means that the probability of extreme observations are higher than the probability considering the normal distribution. At stock exchange the distribution of price movements of the assets have also heavy tails representing the shocks that occasionally happen. However, in the case of a stable distribution, the probability close to the mean is higher than by the normal distribution, so the shape of the density is peaked. To capture this behaviour from the 1960's many author has examined the use of stable distributions for modelling financial time series, a few of them are Mandelbrot [45], Fama [14], [15], [16], Press [66]. More recent works on the topic are Adler, Feldman and Taqqu [3], Bradley and Taqqu [6], Rachev [68], Rachev and Mittnik [69].

However, stable distributions are not commonly accepted and widely used, because there are some issues which are hard to handle. The general probability density function (pdf) and cumulative distribution function (cdf) are not known in exact form. The pdf and cdf of a general stable variable is available in so-called integral representation. Numerical integration has a very high computational demand and the convergence of the formula is very slow, it is possible that around a thousand terms are necessary. A general stable distribution can be characterized by four parameters of the characteristic function (chf). The shape parameter (characteristic exponent, index of stability)  $\alpha \in (0, 2]$  describes peakedness around the mean and heaviness of the tails, skewness  $\beta \in [-1, 1]$  is the symmetry parameter, scale  $\gamma > 0$  is a measure of dispersion and location  $\delta \in \mathbb{R}$  is the mean (if it exists). The chf in general is a complex valued function and cannot be easily used in applications.

The variance is infinite of all stable distributions, except the normal distribution which is a member of the family with characteristic exponent  $\alpha = 2$ . Another fact is that all moments  $E|X|^p$  with  $p \geq \alpha$  are infinite, i.e. do not exist. Basic works dealing with the distribution family are Uchaikin and Zolotarev [79] and Zolotarev [84] on general characteristics, Adler *et. al* [3], Samorodnitsky and Taqqu [71], and Nolan [56] on simulation and statistical diagnostics.

Empirical studies has pro and contra results in connection with application of stable distributions for modelling asset price changes, and intensive research work is still in progress in this field. The inefficient and cumbersome procedures, and computationally intensive algorithms prevented wide-spread usage of the stable family in data analysis until the last decades. These

difficulties tend to vanish with the increased computational capacity of computer hardware and achievements in numerical methodology. There is an increasing need for feasible and reliable methods which can facilitate these statistical examinations, e.g. numerically determine values of the pdf or cdf, goodness-of-fit tests and parameter estimation.

Parameter estimation of the four stable parameters is a compounded problem. The above mentioned properties of the stable family require extraordinary procedures. Methods that are usually used, such as the maximum likelihood method or method of moments can not be applied in a usual way. Many different approaches have been introduced to estimate the parameters. Nolan [55] presented Maximum likelihood method calculated via the numerical integration of stable pdf with Fast Fourier Transform algorithm. There exist tail index estimators which estimates parameter  $\alpha$  of the distribution based on asymptotic Pareto behaviour of the tails, e.g. Csörgő [9], Csörgő and Viharos [10], Hall [26], Hill [31], Szeidl [77], Viharos [80]).

Historical importance is in the methods based on empirical quantiles by Fama and Roll [17], [18], and McCulloch [50]. The characteristic function and its transformations are used in the methods by Press [66]. A regression type estimator based on the characteristic function has been presented by Koutrouvelis [39], and an improved version of his method by Kogon and Williams [38]. Very recent works are Garcia et. al [24] which is a Bayes statistical approach and Matsui and Pawlas [49] where formulas of stable fractional moments are derived for parameter estimation and prediction.

Main motivation of my research was to introduce a new parameter estimation method which is easy to implement, fast, gives accurate, reliable results and useful in practical applications. In this Thesis a new robust parameter estimation method to symmetric ( $\beta = 0$ ) stable distributions will be presented. The procedure is a variation of maximum likelihood type M-estimators presented originally by Huber [32] and its weight functions are derived from Probability Integral Transformation. The estimator provides joint estimation of shape parameter  $\alpha$ , scale parameter  $\gamma$  and location parameter  $\delta$ . The proposed method possesses all known good robustness performance properties. The estimation procedure does not use the probability density function or the characteristic (chf) function directly, hence is faster than the maximum likelihood or chf based methods.

A simulation study will be presented which was performed to compare the proposed estimator with other methods based on performance properties and assess convergence of the estimators. I will also demonstrate that the method could be used to model Hungarian Stock Exchange price change data and the goodness-of-fit of distributions with the estimated parameters was assessed.

The structure of my Thesis is the following. Chapter 2 deals with definitions and basic properties of univariate and multivariate stable distributions. In Chapter 3 the main results of modern portfolio theory are summarized. Chapter 4.1 and 4.2 are overviews on existing parameter estimation methods and achievements of robust statistics, respectively. In Chapter 4 the new parameter estimation method is proposed. The computation of the estimators is fastened with pre-calculated rational fraction approximations with high accuracy. Chapter 4.6 is devoted to present the steps of determining the rational fraction approximation.

Chapter 5 contains simulation results on PIT estimation of the stable parameters. In Chapter 5.1 random variable generation techniques are summarized which were necessary to be implemented for the Monte-Carlo simulations. The simulation study is presented in Chapter 5.2. In Chapter 5.4 a simulation study is described where performance of some known parameter

estimation methods and the new method are compared. The study also investigates the convergence of joint estimators of the parameters and performance of the method. Results can be found in Chapter 5.3.

Chapter 6 is devoted to present the results of modelling Hungarian Stock Exchange price change data with stable distributions and estimate parameters with the PIT estimation procedure. Investigation involves not only parameter estimation of logarithmic price change time series of some assets, but also visualization and goodness-of-fit testing.

In Chapter 7 I summarized all of the MATLAB program codes which were used to reach the scientific results presented in the Thesis. The most important algorithms are presented in the Appendix.

At last I briefly summarize the presented work, the advantages and disadvantages (limitations) of the new method and define further research directions in Chapter 8.

## 9.2 Main results

**THESIS 1** *The M-estimator can be used to estimate stable parameters scale and location if the  $\alpha$ -stable distribution type is known. If the chosen distribution type  $F_{0,\alpha}$  in the weight function is not the one that coincides the sample's  $\alpha$ -stable distribution then a predetermined value  $\mathcal{B}$  that is derived from the chosen  $F_{0,\alpha}$  distribution is used to give an unbiased estimator of scale and location parameter ( $1 \leq \alpha \leq 2$ ).*

**THESIS 2** *If the shape parameter  $\alpha$  is not known but is also to be estimated from the sample, then by comparing scale estimators calculated by considering the normal and the Cauchy distribution as  $F_{0,\alpha}$ , the three parameters of a symmetric stable distribution can be simultaneously estimated.*

**THESIS 3** *The PIT parameter estimation method can be applied for modelling real data sets. By investigating the parameters of logarithmic returns of some assets at Budapest Stock Exchange with PIT method, I have stated that a stable distribution with the estimated parameters fits better to the modelled data set than the normal distribution. By analysing alteration of the parameters in time, I have identified the effects of the world financial crisis in 2008-2009 to the returns and the volatility (risk) of the assets which had a remarkable increase in that time period.*

**THESIS 4** *I have implemented the algorithms that can calculate the PIT estimators of a data set. Moreover, I have written program codes that are used to test accuracy and performance of the new method, accomplish statistical investigation of the method via simulation. Auxiliary MATLAB functions were created to facilitate the application of the PIT method to real financial data. By analysing efficiency and reliability of the PIT method I have proved that the presented new method has similar performance properties as the existing methods. The univariate and multivariate normality of the estimators of the three parameters has been also investigated by a simulation study.*

## 10 Függetlék

### Program 1 *alphasearch*

```
function [alpha, gamma, delta] = alphasearch(sample)
%finding estimates with PIT cut-and-try algorithm
%alpha: characteristic exponent
%gamma: scale parameter
%delta: location parameter

[n,m]=size(sample);

initialalpha1=1;
initialalpha2=2;

[Scale1_1, Location]=pingpong1(sample,initialalpha1); %Cauchy, alpha=1
[Scale1_2, Location]=pingpong1(sample,initialalpha2); %Cauchy, alpha=2
[Scale2_1, Location]=pingpong2(sample,initialalpha1); %normal, alpha=1
[Scale2_2, Location]=pingpong2(sample,initialalpha2); %normal, alpha=2
%location is not used in alphasearch

Sn_mid1=0;
location1=0;
Sn_mid2=0;
location2=0;

alpha=0;
gamma=0;
delta=0;

if(( Scale2_1 > Scale1_1) && (Scale2_2 < Scale1_2))
    %there is an intersection point
    step=0;

    down=1;
    up=2;
    while(step < 18)
        mid=(up-down)/2 + down;
        [Sn_mid1,location1]=pingpong1(sample,mid);
        [Sn_mid2,location2]=pingpong2(sample,mid);

        if ((Sn_mid1 ~= -1) && (Sn_mid2 ~= -1))
            if(Sn_mid2 > Sn_mid1)
                down=mid;
            else
                up=mid;
            end
            step=step+1;
        else
            step=19; %exit while loop
        end
    end
end
```

```

alpha=(up+down)/2;
gamma=(Sn_mid1+Sn_mid2)/2;
delta=(location1+location2)/2;

else
alpha=-1;
gamma=-1;
delta=-1;

%log
logfile = fopen('alphasearchlog.txt','a');
fprintf(logfile, 'there is no intersection point \n');
fclose(logfile);
end
end

```

### **Program 2** *B1value*

```

function B1value=B1(x)
%Evaluate function B1 in point x

a3= 0.00343012930092104;
a2= 0.00605670349030523;
a1= 0.0470997840039568;
a0= 0.00972618179334997;
b2= 1.0;
b1=-0.380875898248913;
b0= 0.176639169350131;

B1value=polyval([a3 a2 a1 a0], x)/ polyval([ b2 b1 b0], x );

end

```

### **Program 3** *B2value*

```

function B2value=B2(x)
%Evaluate B2 function in point x

a3= 0.00631314527817693;
a2= 0.019439037177846;
a1= 0.0933248118805677;
a0= 0.0161987713563325;
b2= 1.0;
b1=-0.0934509457370816;
b0= 0.160295690343675;

B2value=polyval([a3 a2 a1 a0], x)/ polyval([ b2 b1 b0], x );

end

```

### **Program 4** *pingpong1*

```

function [Sn, Tn]=pingpong1(sample, alpha)

```

```

% Finding location and scale estimate with Cauchy CDF at fixed alpha
% Sn: scale estimate
% Tn: location estimate

[n,m]=size(sample);
Sn=mad(sample,1); %1 for median absolute deviation
Tn=median(sample);

Sn2=Sn^2;
Sn2_new=0;
Sn_new=0;

accuracy=10e-8;
distance=10^7;

step=0;

while ((distance > accuracy ) && (step < 100))

    summandT=sum( atan((sample-Tn)/Sn) )/ pi;
    Tn=Tn+(summandT*Sn)/n;

    summandS= sum((atan((sample-Tn)/Sn) /pi).^2);
    Sn2_new=(summandS*Sn2) / ((n-1)*B1(alpha));

    Sn_new=sqrt(Sn2_new);
    distance=abs(Sn-Sn_new);
    Sn=Sn_new;
    Sn2=Sn2_new;

    step=step+1;

end
if step==100
    logfile=fopen('logpingpong1.txt', 'a');
    fprintf(logfile, 'pingpong1: estimation do not converge,
        distance: %1.8f \n', distance );
    fclose(logfile);
    Sn=-1;
    Tn=-1;
end
end

```

## Program 5 *pingpong2*

```

function [Sn, Tn]=pingpong2(sample, alpha)
% Finding location and scale estimate with Normal CDF at fixed alpha
% Sn: scale estimate
% Tn: location estimate

[n,m]=size(sample);

```

```

Sn=mad(sample,1); %1 for median absolute deviation
C=0.6744897501960; % correction with F^-1(3/4)
Sn=Sn/C;
Tn=median(sample);

Sn2=Sn^2;
Sn2_new=0;
Sn_new=0;

accuracy=10e-8;
distance=10^7;

step=0;
while ((distance > accuracy) && (step < 100) )

    summandT=sum( normcdf((sample-Tn)/Sn) - 0.5);
    Tn=Tn+(summandT*Sn)/n;

    summandS= sum((normcdf((sample-Tn)/Sn) -0.5 ).^2);
    Sn2_new=(summandS*Sn2) / ((n-1)*B2(alpha));

    Sn_new=sqrt(Sn2_new);
    distance=abs(Sn-Sn_new);
    Sn=Sn_new;
    Sn2=Sn2_new;

    step=step+1;
end
if step==100
    logfile=fopen('logpingpong2.txt', 'a');
    fprintf(logfile, 'pingpong 2: estimation do not converge,
        distance: %1.8f \n', distance );
    fclose(logfile);
    Sn=-1;
    Tn=-1;
end
end
end

```

### Program 6 *multivnormality*

```

function eredmeny=multivnormality(X);
%többváltozós Mardia féle ferdeség és lapultság értékeket ad meg
%X: tartalmazza az adatokat, n*p típusú, n: megfigyelések száma, p elemű vektorok

[n,p]=size(X);
Xatlag=mean(X);
Xhullam=zeros(n,p);
oszlop=ones(n,1);

Xhullam=X-oszlop*Xatlag;
S=(1/n)*Xhullam'*Xhullam;
Sinv=inv(S);

```

```

D=Xhullam*Sinv*(Xhullam');

skewness=(1/n^2)*sum(sum(D.^3));
skewness_=n*skewness/6;
%chi^2 eloszlású inntől p(p+1)(p+2)/6 szabadsági fokkal df=10

d=diag(D).^4;
kurtosis=(1/n)*sum((diag(D).^2));
kurtosis_=(sqrt(n)*(kurtosis-p*(p+2))^2)/sqrt(8*p*(p+2));

% standard normálissal kell tesztelni inntől
omnibus=skewness_+kurtosis_;

% A ferdeség (skewness) 10 szabadsági fokú chi^2 eloszlás
if (skewness_ < chi2inv(0.95,10) )
    dontes_s=0;
else
    dontes_s=1;
end

pskewness=1-chi2cdf(skewness_,10);

% A lapultságot (kurtosis) standard normálissal szemben kell tesztelni
if (kurtosis_ < norminv(0.975) )
    dontes_k=0;
else
    dontes_k=1;
end

pkurtosis=1-normcdf(kurtosis_);

eredmeny=[dontes_s, skewness_, pskewness, dontes_k, kurtosis_, pkurtosis];
end

```

### Program 7 sarkadi1D

```

function [Y]=sarkadi1D(X)
%Sarkadi próba 1dimenzióban
%input: X egyváltozós minta, X vektor
%output: transzformált minta

[n,m]=size(X);
x_osszeg=sum(X);
x_modosított_atlag=(x_osszeg+sqrt(n/2)*(X(n-1)+X(n)))/(n+sqrt(2*n));
S=sqrt(sum(X.^2)-(x_osszeg^2)/n-((X(n-1)-X(n))^2)/n );
t=abs(X(n-1)-X(n))*sqrt(n-2)/S*sqrt(2);
nu=n-2;

%pszi értéke
jobboldal= 2*tcdf(t,nu)-1;

also=0;
felso=4000;

```



```

kozep=0;
kulonbseg=10e+300;
pontosság=0.00001;

ertek_also=1-chi2cdf(also,nu)-jobboldal;
ertek_felső=1-chi2cdf(felső,nu)-jobboldal;
%intervallumfelezés
while (kulonbseg>pontosság)

    kozep=(also+felső)/2;
    ertek_kozep=1-chi2cdf(kozep,nu)-jobboldal;

    if( ertek_kozep>0 )
        also=kozep;
    else
        felső=kozep;
    end
    kulonbseg=abs(ertek_kozep);
end

pszi_ertek=sqrt(kozep);
Y=(X(1:n-2,:)-x_modositott_atlag)./S;
end

```

### Program 8 *DoornikHansen*

```

function [dontes, stat, pertek]=DoornikHansen(X)
% többváltozós normalitás tesztelés Doornik és Hansen alapján
% n megfigyelés, p dimenziós vektorok
% X egy sora egy megfigyelés összetartozó adatai

[n,p]=size(X);
X_osszeg=sum(X); %sorvektor az oszlopok összegeivel
X_atlag=X_osszeg./n; %átlag sorvektor

S=cov(X);
d=1./(sqrt(diag(S))); %diagonális elemek gyökének reciproka
V=zeros(p,p);
for i=1:p
    V(i,i)=d(i);
end
C=V*S*V; % korrelációs együtthatók mátrixa

[H,L]=eigs(C);

Xhullam= X-(ones(n,1)*X_atlag);
Lmod=L^(-1/2);
Rt=H*Lmod*H'*V*Xhullam';
R=Rt';

z1=skewt(R);

```

```
z2=kurtt(R);

E=z1*z1'+z2*z2';
%chi négyzet eloszlással kell összehasonlítani,
% df = 6 nál 12.59 (95%), 16.81 (99%), 22.46 (99,9%)
if E < chi2inv(0.95, 6)
    dontes=0;
else
    dontes=1;
end

stat=E;
pertek=1-chi2cdf(E, 6);

end
```

## Irodalomjegyzék

### A jelölt témában született munkái.

#### Nemzetközi, lektorált folyóirat

- [S1] Csendes, Cs., Joint Robust Parameter Estimation for Symmetric Stable Distributions. *Journal of Statistical and Econometric Methods* **2** (2013) 85–106

#### Nemzetközi, lektorált könyvfejezet

- [S2] Csendes, Cs., Fegyverneki, S., Parameter Estimation for Symmetric Stable Distributions by Probability Integral Transformation. *Applied Information Science, Engineering and Technology, Topics in Intelligent Engineering and Informatics* **7** (2014) 1–18  
DOI : 10.1007/978 – 3 – 319 – 01919 – 2

#### Hazai lektorált folyóirat

- [S3] Árfolyamingadozások vizsgálata szimmetrikus stabil modellben, *Sigma*, **XLV** 3-4. 1–26 (várható megjelenés: 2015)

#### Konferenciakiadvány

- [S4] Csendes, Cs., Fegyverneki, S.: Parameter Estimation to the Stable Portfolio Analysis, *Proceedings of XXIII. microCAD International Scientific Conference, Sec. G, Miskolc, Hungary*, (2009) 1–8.
- [S5] Csendes, Cs.: Random Number Generation to Multivariate Stable Distributions, *Proceedings of Spring Wind International Conference, Pécs, Hungary*, (2010) 79–85.
- [S6] Csendes, Cs.: Random Number Generation to Multivariate Stable Distributions, *Proceedings of XXIV. microCAD International Scientific Conference, Sec. G, Miskolc, Hungary*, (2010) 7–13.
- [S7] Csendes, Cs.: Parameter Estimation and Random Number Generation to Stable Distributions, *Proceedings of 8th International Conference on Applied Informatics, Eger, Hungary*, (2010) 239–246.
- [S8] Csendes, Cs.: Normality Testing on PT Estimation of Parameters of Stable Distributions, *Proceedings of International Conference of Ph.D. Students, Miskolc, Hungary*, (2010) 35–40.
- [S9] Csendes, Cs.: Multivariate Normality Testing, *Proceedings of XXV. microCAD International Scientific Conference, Sec. G, Miskolc, Hungary*, March (2011) 13–18.
- [S10] Csendes Cs.: Parameter Estimation and Hypothesis Testing to Stable Distributions, *Proceedings of 17th European Young Statisticians Meeting, Lisbon, Portugal*, (2011) 69–73.

[S11] Csendes Cs.: A Simulation Study about Stable Distributions, *Proceedings of XXVI. microCAD International Scientific Conference, Miskolc, Hungary, March (2012)*

### **Egyéb publikációk**

[S12] Csendes, Cs.: Stabil portfólió analízis, Tudományos Diákköri Dolgozat, (2008) 59 p.

[S13] Csendes, Cs.: Többváltozós stabil eloszlású véletlen számok generálása, *Doktoranduszok Fóruma Kiadvány*, Miskolci Egyetem (2009)

[S14] Csendes, Cs.: Normality Testing on PT Estimation of Parameters of Stable Distributions, *Doktoranduszok Fóruma Kiadvány*, Miskolci Egyetem, (2010)

## Irodalomjegyzék

- [1] Aban, I. B., Meerschaert, M. M., Shifted Hill's Estimator for Heavy Tails. *Communications in Statistics - Simulation and Computation*, **30**(4) (2001) 949–962
- [2] Abdul-Hamid, H., Nolan, J. P., Multivariate stable densities as functions of their one dimensional projections. *J. Multivariate Analysis*, **67** (1998) 80–89
- [3] Adler, J. R., Feldman, R. E., Taqqu, M. S., (Editors), *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, Birkhauser, Boston (1998)
- [4] Borak, Sz., Härdle, W., Weron, R., Stable Distributions. SFB 649 Discussion Papers, SFB649DP2005-008, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany (2005)
- [5] Box, G. E. P, Non-normality and Tests on Variances. *Biometrika*, **4** (1953) 318–335
- [6] Bradley, B. O., Taqqu, M. S., *Financial Risk and Heavy Tails*, in Hadbook of Heavy-tailed Distributions in Finance, (ed. Rachev, S. T.), North-holland (2003) 35–103
- [7] Brealey, R. A., Myers, R. C., *Modern vállalati pénzügyek*, Panem, (2005)
- [8] Chambers, J. M., Mallows, C. L., Stuck, B. W., A method for simulating stable random variables. *J. Amer. Statist. Assoc.*, **71** (1976) 340–344
- [9] Csörgő, S., Adaptive Estimation of the Parameters of Stable Laws. *Colloquia Math. Soc. J. Bolyai*, **36**. Limit Theorems in Probability and Statistics (szerk. Révész P.), North-Holland, Amsterdam (1984) 305–368
- [10] Csörgő, S. , Viharos, L., *Estimating the Tail Index*, in: Asymptotic Methods in Probability and Statistics (ed. Szyszkowicz, B.), Elsevier Science, North-Holland (1998) 833–881
- [11] Davies, P. L., On Locally Uniformly Linearizable High Breakdown Location and Scale Functionals. *Ann. Stat.*, **26** (1998) 1103–1125
- [12] Doornik, J.A., Hansen, H., An Omnibus Test for Univariate and Multivariate Normality. *Nuffield Economics Working Papers*, (1994)
- [13] Dutter R., Huber, P. J., Numerical Methods for the Nonlinear Robust Regression Problem. *J. statist. comput. simul.*, **13** (1981) 79–113
- [14] Fama, E. F., Portfolio Analysis in Stable Paretian Markets. *Management Science*, **11**(3) (1965) 404–419
- [15] Fama, E. F., The Behavior of Stock-Market Prices. *Journal of Business*, **38**(1) (1965) 34–105
- [16] Fama, E. F., Risk, Return, and Equilibrium. *Journal of Political Economy*, **79**(1) (1971) 30–55
- [17] Fama, E. F., Roll, R., Parameter Estimates of Symmetric Stable Distributions. *J. Amer. Statis. Assoc.*, **66** (1971) 331–338

- [18] Fama, E. F., Roll, R., Some Properties of Symmetric Stable Distributions. *Journal of the American Statistical Association*, **63** (1968) 817–836
- [19] Fegyverneki, S., Robust Estimators and Probability Integral Transformations. *Math. Comput. Modelling*, **38** (2003) 803–814
- [20] Feldheim, M. E., Étude de la stabilité des lois de probabilité. Thèse de la Faculté des Sciences de Paris (1937)
- [21] Feller, W., *An Introduction to Probability Theory and Its Applications. II.* Wiley, New York (1966)
- [22] Gáll, J., Pap, Gy., *Bevezetés a hasznosság alapú portfólió-menedzsmentbe.*, egyetemi jegyzet, mobiDIÁK könyvtár, Debreceni Egyetem (2004)
- [23] Gather, U.; Davies, P. L., Robust Statistics, Papers / Humboldt-Universität Berlin, Center for Applied Statistics and Economics (CASE), No. 2004,20, <http://hdl.handle.net/10419/22194> (2004)
- [24] Garcia, R., Renault, E., Veredas, D., Estimation of Stable Distributions by Indirect Inference. *J. Econometrics*, **161** (2011) 325–337
- [25] B. V. Gnedenko, A. N. Kolmogorov, *Független valószínűségi változók összegeinek határeloszlásai.*, Akadémiai kiadó (1951)
- [26] Hall, P., On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc., Ser. B*, **44** (1982) 37–42
- [27] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., Stahel, W. A., *Robust Statistics - The Approach Based on Influence Functions*, Wiley, New York (1986)
- [28] Hampel, F. R., Robust Estimation: a Condensed Partial Survey. *Z. Wahrsch. verw. Geb.*, **27**, (1973) 87–104
- [29] Huber, P. J., Dutter, R., Numerical Solution of Robust Regression Problems. *COMPSTAT 1974, Proc. in Comput. statist.*, ed. G. Bruckmann, Physica Verlag, Vienna (1974)
- [30] Henze, N., Invariant Tests for Multivariate Normality: a Critical Review. *Springer Statistical Papers*, **43**(4), (2001) 467–506
- [31] Hill, B. M., A Simple General Approach to Inference about the Tail of a Distribution. *Ann. Stat.*, **3** (1975) 1163–1174
- [32] Huber, P. J., *Robust Statistics*. Wiley, New York (1981)
- [33] Huber, P. J., Robust Estimation of a Location Parameter. *Ann. Math. Statist.*, **35** (1964) 73–101
- [34] Kendall, M. G., Buckland, W. R., *A Dictionary of Statistical Terms.*, 3rd ed., Hafner Pub. Co., (1971)
- [35] Kerékfy, P., A robusztus becslésekről. *Alkalmazott Matematikai Lapok*, **4** (1978) 327–357

- [36] Khindanova, I., Rachev, S., Schwartz, E., Stable modeling of value at risk. *Math. Comput. Modelling*, **34** (9-11) (2001) 1223–1259
- [37] Kincaid D., Cheney, W., *Numerical Analysis: Mathematics of Scientific Computing*. 3rd Edition, American Mathematical Society, Providence (2002)
- [38] Kogon, S. M., Williams, D. B. *Characteristic function based estimation of stable parameters*, in Adler, R., Feldman, R., Taqqu, M., (eds.), *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, Birkhauser, Boston (1998) 311–335
- [39] Koutrouvelis, I. A., Regression-type Estimation of the Parameters of Stable Laws. *J. Amer. Statist. Assoc.*, **75** (1980) 918–928
- [40] Levy, P., *Calcul des Probabilités*. Gauthier-Villars, Paris (1925)
- [41] Levy, P., Theorie des erreurs la loi de Gauss et les lois exceptionnelles. *Bulletin Soc. Math. France*, **52** (1924) 49–85
- [42] Lintner, J., The Valuation of Risky Assets and the Selection of Risky Investment in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics*, **47** (1965) 13–37
- [43] Lukács, P., Portfólió optimalizálása várható hozam - varianca és várható hozam - CVaR módszerrel. *Vezetéstudomány*, **35** (2) (2004) 34–41
- [44] Lux, T., Varga, J., A Pareto hipotézis vizsgálata: értékpapírpiaci hozamok és az extrémális hozamok eloszlása. *Sigma*, **27** (1996) 1–23
- [45] Mandelbrot, B., The Variation of Certain Speculative Prices. *The Random Character of Stock Market Prices* (ed. Cootner, P. H.), Cambridge, The M.I.T. Press (1964)
- [46] Mardia, K. V., Tests of Univariate and Multivariate Normality. *Handbook of Statistics*, North-Holland (1980) 279–320
- [47] Mardia, K. V., Applications of some measures of multivariate skewness and kurtosis for testing normality and robustness studies. *Sankhya*, **A 36** (1974) 116–128
- [48] Markowitz, H., Portfolio Selection. *Journal of Finance*, **7** (1952) 77–91
- [49] Matsui, M., Pawlas, Z., Fractional Absolute Moments of Heavy Tailed Distributions. online verzió: <http://arxiv-web3.library.cornell.edu/abs/1301.4804> (2014)
- [50] McCulloch, J. H., Simple Consistent Estimators of Stable Distribution Parameters. *Commun. Statist. - Simula.*, **15**(4) (1986) 1109–1136
- [51] McCulloch, J. H., Measuring Tail Thickness to Estimate the Stable Index: A Critique. *Journal of Business and Economic Statistics*, **15** (1997) 74–81
- [52] Mittnik, S., Rachev, S. T., Paolella, M. S., *Stable Paretian Modeling in Finance*. In Adler, R. J., Feldman, R. E., and Taqqu, M.S., (szerk.), *A Practical Guide to Heavy Tails: Statistical Techniques for Analyzing Heavy Tailed Distributions*, Birkhäuser, Boston (1998)

- [53] Mittnik, S., Rachev, S. T., Modeling Asset Returns with Alternative Stable Distributions. *Econometric Reviews*, **12**(3) (1993) 261–330
- [54] Modarres, R., Nolan, J. P., A Method for Simulating Stable Random Vectors. *Computational Statistics*, **9** (1994) 11–19
- [55] Nolan, J. P., Maximum Likelihood Estimation of Stable Parameters in Barndorff-Nielsen, O. E. , Mikosch, T., and Resnick, S. I., (eds.), *Levy Processes: Theory and Applications*, Birkhäuser, Boston (2001) 379–400
- [56] Nolan, J. P., Numerical Calculation of Stable Densities and Distribution Functions. *Comm. in Stat. - Stoch. Model.*, **13** (1997) 759–774
- [57] Nolan, J. P., An Overview of Multivariate Stable Distributions. *Online elérhető: <http://academic2.american.edu/jpnolan/stable/overview.pdf>* (2008)
- [58] Nolan, J. P., Multivariate Stable Densities and Distribution Functions: General and Elliptical Case. Deutsche Bundesbank's 2005 Annual Fall Conference (2005)
- [59] Nolan, J. P., Multivariate elliptically contoured stable distributions: theory and estimation. *Online elérhető: <http://academic2.american.edu/jpnolan/stable/EllipticalStable.pdf>* (2006)
- [60] Nolan, J. P., Panorska, A. K., McCulloch, J. H., Estimation of Stable Spectral Measures. *Math. Comput. Modelling*, **34** (2001) 1113–1122
- [61] Nolan, J. P., stabil eloszlásokkal foglalkozó weboldal: <http://academic2.american.edu/jpnolan/stable/stable.html>
- [62] Oral, E., Erdemir C., A Bayesian Estimation of Stable Distributions. *Journal of Statistical and Econometric Methods*, **1** (3) (2012) 39–52
- [63] Palágyi, Z., Árfolyamingadozások és kockázatbecslés a Budapesti Értéktőzsdén. *Sigma*, **30**(1-2) (1999) 27–33
- [64] Palágyi, Z., Pénzügyi idősorok elemzése Levy-hatvány GARCH modellel. *Statisztikai szemle*, **81** (7) (2003) 571–587
- [65] Petrov, V. V., Sums of Independent Random Variables. Springer (1975)
- [66] Press, S. J., *Applied Multivariate Analysis*. Holt, Rinehart and Winston, New York (1972)
- [67] Press, S. J., Multivariate Stable Distributions. *Journal of Multivariate Analysis*, **2** (1972) 444 – 462
- [68] Rachev, S. T. (ed.), *Handbook of Heavy-tailed Distributions in Finance*. North-Holland, Amsterdam (2003)
- [69] Rachev, S. T., Mittnik, S., *Stable Paretian Models in Finance*. Wiley, New York (2000)
- [70] Rousseeuw, P. J., A New Infinitesimal Approach to Robust Estimation. *Z. Wahrsch. verw. Geb.*, **56**, (1981) 127–132



- [71] Samorodnitsky, G., Taqqu, M., *Stable Non-Gaussian Random Processes*. Chapman and Hall, New York (1994)
- [72] Sarkadi, K., Tusnády, G., Testing for Normality and for the Exponential Distribution. *Proceedings of the 5th Conference on Probability Theory*, Brasov, (1977) 99–118
- [73] Sarkadi, K., On Testing for Normality. *Proc. Fifth Berkeley Symp. on Math. Statist. and Prob.*, **1** Univ. of Calif. Press, (1967) 373–387
- [74] Stoyanov, S., Samorodnitsky, G., Ortobelli, S., Rachev, S. T., Computing the portfolio Conditional Value-at-Risk in the alpha-stable case. *Probability and Mathematical Statistics*, **26**(1) (2006) 1–22
- [75] Sharpe, W. F., Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance*, **19** (1964) 425–442
- [76] Stoyanov, S., Samorodnitsky, G., Ortobelli, S., Rachev, S. T., Computing the portfolio Conditional Value-at-Risk in the alpha-stable case. *Probability and Mathematical Statistics*, **26**(1) (2006) 1–22
- [77] Szeidl, L., Non-normal Limit Theorem for a New Tail Index Estimation. *Annales Univ. Sci. Budapest. Sect. Comp.*, **24**, (2004) 307–322
- [78] Thode, H. C., *Testing for normality*. Marcel Dekker, New York (2002)
- [79] Uchaikin, V. V., Zolotarev, V. M., *Chance and Stability - Stable Distributions and their Applications*, VSP, Utrecht (1999)
- [80] Viharos, L., Tail index estimation based on linear combinations of intermediate order statistics. *Statistica Neerlandica*, **51** (1997) 164–177
- [81] Weron, R., Performance of the Estimators of Stable Law Parameters. Hugo Steinhaus Center for Stochastic Methods, *Research Report HSC/95/1* (1995)
- [82] Weron, R., On the Chambers-Mallow-Stuck method for simulating skewed stable random variables. *Statist. Probab. Lett.*, **28** (1996) 165–171
- [83] Weron, R., Computationally intensive Value at Risk calculations, Papers / Humboldt-Universität Berlin, Center for Applied Statistics and Economics (CASE), No. 2004, 32 (2004)
- [84] Zolotarev, V. M., *One-dimensional Stable Distributions*, Translations of Mathematical Monographs, **65**, American Mathematical Society, Providence (1986)