

Nagy károk modellezése

Diplomamunka

Írta: Szalai András

Alkalmazott matematikus szak

Témavezető:

Zempléni András, egyetemi docens

Matematikai intézet

Eötvös Loránd Tudományegyetem, Természettudományi Kar



Eötvös Loránd Tudományegyetem

Természettudományi Kar

2011

Tartalomjegyzék

Bevezetés	1
1. Határeloszlások	1
1.1. Maximumok eloszlása	1
1.1.1. Az általánosított szélsőérték-eloszlás	1
1.1.2. Max-vonzási tartományok, karakterizáció	4
1.1.3. Erősen stacionárius idősorok	6
1.2. Küszöbmeghaladások, általánosított Pareto eloszlás	7
2. Extrémérték - modellek	10
2.1. Blokk maximumok módszere	10
2.2. Küszöbmeghaladások	12
2.2.1. GPD modell	12
2.2.2. Farokeloszlások és kockázati mértékek becslése	14
2.2.3. Viszontbiztosítási szerződés árazása	15
2.3. A Hill-módszer	16
3. Adatelemzés	18
3.1. Dán tűzkárok	18
3.1.1. Küszöb-választás	18
3.1.2. Kvantilisek, díjkalkuláció, érzékenység	21
3.2. Magyarországi biztosítási adatok	24
3.2.1. Küszöb-választás	24
3.2.2. Érzékenység	27
3.3. Összegzés	29
Irodalomjegyzék	30
Függelék	30

Bevezetés

Egy bizonyos valószínűségi változó legnagyobb értékeinek előrejelzése természetesen felmerülő igény a modern élet sok területén. Az extrémérték-elmélet kezdeti felhasználási területei főleg a meteorológiai, hidrológiai előrejelzések voltak - legnagyobb vízszintek, hőmérsékleti értékek előrejelzésére használták. A huszadik század végén ismerték fel alkalmazhatóságát pénzügyi területen, ahol különböző veszteségek becslésére, díjkalkulációra, tartalékolási számításokra használják. A téma fő úttörőinek Leonard Tippett (1902–1985), Sir Ronald Aylmer Fisher (1890-1962), és Emil Julius Gumbel (1891-1966) számítanak.

Jelen szakdolgozat témája az elméleti eredmények bemutatása, majd azok biztosítási adatsorokra való alkalmazása. Az első fejezetben bemutatom a modellekhez felhasznált elméleti eredményeket, határeloszlásokat: az általánosított extrémérték-eloszlást és az általánosított Pareto-eloszlást. A másodikban a különböző modellekkel és bizonyos jellemző paraméterek becslésével foglalkozom: a blokk maximumokat használó módszerről, a bizonyos határt túllépő értékeket használó küszöbmeghaladásos módszerről, és a Hill által javasolt, Karamata tételén alapuló módszerről. Bemutatom két kockázati mérték: a value at risk és az expected shortfall becslésének módját is. Végül a harmadik fejezetben biztosítási adatokra próbálom alkalmazni az ismertetett módszereket. Két adatsort vizsgálok meg a küszöbmeghaladásos módszer segítségével: az egyik a témában klasszikusnak számító, a modell bemutatására jól használható dán tűzkárok adatai; a másik egy magyarországi biztosítási kárkifizetéseket tartalmazó adatsor. A vizsgálat fő célja a modell megfelelő megválasztása lesz, összehasonlítom a különböző illeszthető modelleket és megpróbálom a legmegfelelőbbet kiválasztani, az első esetben viszontbiztosítói, a másodikban direkt biztosítói szempontból.

1. fejezet

Határeloszlások

1.1. Maximumok eloszlása

Tekintsük $(X_i)_{i \in \mathbb{N}}$ független, azonos eloszlású (iid) valószínűségi változók sorozatát. Ezek lehetnek valamilyen szabályos időközönként mért adatok, pl. hőmérséklet-vagy vízállás-értékek, pénzügyi veszteségek, biztosítási kárfizetések. Természetesen a nyersen kapott adatok általában nem teljesítik az iid feltételt, gyakran valamilyen trend figyelhető meg, ezt az elemzés előtt ki kell szűrni (pénzügyi adatoknál ilyen lehet pl. az infláció, hőmérséklet-adatoknál a periodicitás). Ez alapján az egyetlen adatsor alapján szeretnénk becslést adni az X_i -k közös eloszlásának valamilyen magas rendű kvantilisére, valamint a legnagyobb események bekövetkezésének várható időpontjára.

1.1.1. Az általánosított szélsőérték-eloszlás

Motiváció - a centrális határeloszlás-tétel

A GEV szerepe a maximális értékek vizsgálata terén hasonló, mint a normális eloszlásé a valószínűségi változók összegénél. Vegyünk X_1, X_2, \dots független, azonos eloszlású valószínűségi változókat, véges szórással. Jelöljük S_n -el az $X_1 + X_2 + \dots + X_n$ összeget. Ekkor a centrális határeloszlás-tétel (CHT) szerint a normalizált $(S_n - a_n)/b_n$ valószínűségi változók sorozatának eloszlásbeli határértéke standard normális, ha n tart a végtelenbe. A megfelelő normalizáló konstansok $a_n = nE(X_1)$ és $b_n = \sqrt{nD(X_1)}$. Formálisan:

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - a_n}{b_n} \leq x\right) = \Phi(x), \quad x \in \mathbb{R}.$$

Maximális értékek, a Fisher-Tippett-Gnedenko tétel

Tekintsük most S_n helyett az $M_n = \max(X_1, \dots, X_n)$ maximumok sorozatát, X_i továbbra is iid megfigyelés F közös eloszlásfüggvénnyel. Ekkor természetesen M_n eloszlásfüggvénye: $F_{M_n}(x) = P(M_n \leq x) = P(X_1 \leq x)^n = F^n(x)$ Tegyük fel, hogy a CHT-hez hasonlóan $\bar{M}_n = (M_n - d_n)/c_n$ - nek létezik eloszlásbeli határértéke megfelelő c_n, d_n normalizáló sorozatok mellett. Formálisan:

$$\lim_{n \rightarrow \infty} P((M_n - d_n)/c_n \leq x) = \lim_{n \rightarrow \infty} F^n(c_n x + d_n) = H(x) \quad (1.1)$$

teljesül valamilyen nem elfajult $H(x)$ eloszlásfüggvényre.

1.1.1. Definíció (max-vonzási tartomány). Amennyiben (1.1) teljesül valamilyen nem elfajuló H eloszlásfüggvénnyel, akkor azt mondjuk, hogy F a H max-vonzási tartományában (maximum domain of attraction, MDA) van. Jelölése: $F \in MDA(H)$.

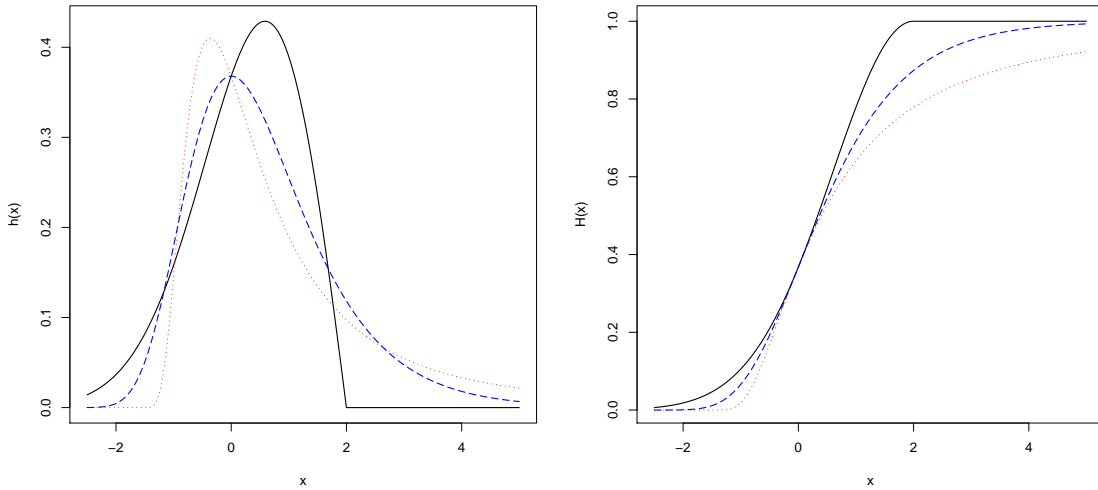
1.1.2. Definíció (GEV eloszláscsalád). A standard GEV eloszlás eloszlásfüggvénye:

$$H_\xi(x) = \begin{cases} \exp(-(1 + \xi x))^{-1/\xi}, & \xi \neq 0, \\ \exp(-e^{-x}), & \xi = 0, \end{cases}$$

ahol $1 + \xi x > 0$. ξ az eloszlás alakparamétere. További két paraméter bevezetésével kapjuk a háromparaméteres $H_{\xi, \mu, \sigma} := H_\xi((x - \mu)/\sigma)$ eloszláscsaládot, $\mu \in \mathbb{R}$ eltolási és $\sigma > 0$ skálaparaméterekkel. ξ értéke szerint háromféle, különböző tulajdonságokkal rendelkező eloszlásról beszélhetünk (1.1 ábra):

- $\xi > 0$ esetén Fréchet-eloszlás: végtelen tartójú, lassú lecsengésű
- $\xi = 0$ esetén Gumbel-eloszlás: végtelen tartójú, Fréchet-nél jóval gyorsabban lecsengő
- $\xi < 0$ esetén Weibull-eloszlás: felülről korlátos tartójú.

A család statisztikailag fontos tulajdonsága, hogy rögzített x esetén ξ -ben folytonos, azaz $\lim_{\xi \rightarrow 0} H_\xi(x) = H_0(x)$ mindkét oldalról.



1.1. ábra. A GEV eloszlás sűrűségfüggvényei (balra) és eloszlásfüggvényei (jobbra). Folytonos vonallal $\xi = -0.5$ (Weibull), szaggatottal $\xi = 0$ (Gumbel), pontozottal $\xi = 0.5$ (Fréchet) esetben. Mindhárom változatnál $\mu = 0$, $\sigma = 1$.

1.1.3. Tétel (Fisher-Tippett, 1928; Gnedenko, 1941). Ha $F \in MDA(H)$ valamilyen nem elfajuló H eloszlásfüggvényre, akkor H szükségszerűen a GEV eloszlás-családba tartozik, azaz $\exists \xi \in \mathbb{R}$, hogy $H = H_\xi$.

A tétel szerint csak ξ értéke egyértelműen meghatározott; az eltolási és skálaparaméterek a választott c_n, d_n sorozatoktól függenek. Ezek mindig megválaszthatóak úgy, hogy a határeloszlás standard alakban jelenjen meg, azaz $\mu = 0$ és $\sigma = 1$ teljesüljön.

Példák

Exponenciális eloszlás. Ha a kezdeti adataink $F(x) = 1 - \exp(-\lambda x)$ eloszlásfüggvényű exponenciális eloszlásból származnak, ($\lambda > 0$, $x \geq 0$), akkor $c_n = 1/\lambda$ és $d_n = \ln n/\lambda$ választással a maximumok határeloszlása kiszámolható:

$$F^n(c_n x + d_n) = \left(1 - \frac{1}{n} \exp(-x)\right)^n, \quad x \geq -\ln n,$$

$$\lim_{n \rightarrow \infty} F^n(c_n x + d_n) = \exp(-e^{-x}), \quad x \in \mathbb{R}$$

Azaz $F \in MDA(H_0)$.

Pareto eloszlás. Ha a kezdeti adataink $F(x) = 1 - (\kappa/(\kappa+x))^\alpha$ eloszlásfüggvényű $\text{Pa}(\alpha, \kappa)$ eloszlásból származnak, ($\alpha > 0$, $\kappa > 0$, $x \geq 0$), akkor $c_n = \kappa n^{1/\alpha}$ és

$d_n = \kappa n^{1/\alpha} - \kappa$ választással a maximumok határeloszlása kiszámolható:

$$F^n(c_n x + d_n) = \left(1 - \frac{1}{n} \left(1 + \frac{x}{\alpha}\right)^{-\alpha}\right)^n, 1 + \frac{x}{\alpha} \geq n^{-1/\alpha},$$

$$\lim_{n \rightarrow \infty} F^n(c_n x + d_n) = \exp\left(-\left(1 + \frac{x}{\alpha}\right)^{-\alpha}\right), 1 + \frac{x}{\alpha} > 0,$$

azaz $F \in MDA(H_{1/\alpha})$.

1.1.2. Max-vonzási tartományok, karakterizáció

Arra a kérdésre, hogy egy adott F eloszláshoz létezik-e H_ξ , hogy $F \in MDA(H_\xi)$, szinte az összes gyakorlatban használt folytonos eloszlás esetén igen a válasz. Kíváncsiságunk tárgya a határeloszlás fajtájának meghatározása, azaz annak eldöntése, hogy milyen F esetén kapunk Fréchet, Gumbel, vagy Weibull eloszlást.

Fréchet eset

$\xi > 0$ esetén a karakterizációhoz szükségünk lesz a lassú változású és reguláris változású függvényekre.

1.1.4. Definíció (lassú és reguláris változású függvények).

1. Egy pozitív, $(0, \infty)$ -en Lebesgue-mérhető L függvény lassú változású a végtelenben, ha

$$\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1, t > 0.$$

2. Egy pozitív, $(0, \infty)$ -en Lebesgue-mérhető h függvény reguláris változású a végtelenben ρ kitevővel, ha

$$\lim_{x \rightarrow \infty} \frac{h(tx)}{h(x)} = t^\rho, t > 0.$$

Például, bármely végtelenben határértékkel rendelkező függvény lassú változású, mivel ekkor $\lim_{x \rightarrow \infty} L(tx) = \lim_{x \rightarrow \infty} L(x) = b$, valamint $h(x) = x^\rho$ reguláris változású bármely $\rho \neq 0, t > 0$ esetén, hiszen $\lim_{x \rightarrow \infty} \frac{h(tx)}{h(x)} = \lim_{x \rightarrow \infty} t^\rho = t^\rho$.

1.1.5. Tétel (Gnedenko, 1941). $\xi > 0$ esetén,

$$F \in MDA(H_\xi) \iff \bar{F}(x) = 1 - F(x) = x^{-1/\xi} L(x)$$

valamilyen L lassú változású függvényre.

Tehát a Fréchet esethez vezető eloszlások farka reguláris változású, negatív indexszel. Ezek az eloszlások a legfontosabbak a pénzügyi modellezés területén, mivel lassan lecsengők, és magasabb rendű momentumaik végtelenek.

1.1.6. Állítás. *Legyen X egy nemnegatív valószínűségi változó amelynek F eloszlásfüggvényére $F \in MDA(H_\xi)$ valamilyen $\xi > 0$ -ra. Ekkor $E(X^k) = \infty$, ha $k > 1/\xi$.*

Példa - Pareto eloszlás

$$\bar{F}(x) = \left(\frac{\kappa}{\kappa + x} \right)^\alpha = x^{-\alpha} \frac{\kappa^\alpha}{\left(1 + \frac{\kappa}{x}\right)^\alpha}$$

Tehát $\alpha = 1/\xi$ és $L(x) = \kappa^\alpha / \left(1 + \frac{\kappa}{x}\right)^\alpha$ választással $F \in MDA(H_\xi)$.

Fréchet határeloszlást kapunk továbbá pl. az inverz gamma, Student t , loggamma, F , és a Fréchet eloszlások maximális értékeinek vizsgálatakor is.

Gumbel eset

Az ebbe a csoportba tartozó eloszlások meghatározása jóval nehezebb, mint a Fréchet esetben. Korábban láttuk, hogy pl. az exponenciális eloszlás ide tartozik. Általánosan elmondható, hogy a Gumbel határeloszlásra vezető eloszlások exponenciális lecsengésűek. Ezekből az eloszlásokból származó pozitív valószínűségi változók momentumai végesek, azaz $F_X \in MDA(H_0)$ esetén $E(X^k) < \infty$ minden $k > 0$ esetén. Ide tartoznak pl. a normális, lognormális, Gumbel, gamma, χ^2 .

Weibull eset

Felülről korlátos tartója miatt ez az eset kevésbé fontos a pénzügyi modellezés területén, mint a fentiek. A Weibull osztály a következőképpen karakterizálható:

1.1.7. Tétel (Gnedenko, 1941). $\xi < 0$ esetén, x_F -el jelölve az eloszlás tartójának jobboldali határát, azaz $x_F = \sup \{x \in \mathbb{R} : F(x) < 1\}$

$$F \in MDA(H_{1/\xi}) \iff \bar{F}(x^F - x^{-1}) = 1 - F(x) = x^{-1/\xi} L(x), \text{ és } x_F < \infty$$

valamilyen L lassú változású függvényre.

Ide tartozik például a béta (így speciális esetként az egyenletes) eloszlás is.

1.1.3. Erősen stacionárius idősorok

A függetlenség, iid minta feltételezése gyakran túl erősnek bizonyul, ezért szeretnénk a fenti eredményeket a pénzügyi modellezés területén jól ismert erősen stacionárius idősorok esetére kiterjeszteni. Ehhez kezdetben vegyünk $(X_i)_{i \in \mathbb{Z}}$ erősen stacionárius idősort F stacionárius eloszlással, és $(\tilde{X}_i)_{i \in \mathbb{N}}$ iid mintát ugyanabból az F eloszlásból. Legyen $M_n = \max(X_1, \dots, X_n)$ és $\tilde{M}_n = \max(\tilde{X}_1, \dots, \tilde{X}_n)$ rendre az idősor és az iid minta maximuma.

A legtöbb $(X_i)_{i \in \mathbb{N}}$ folyamatra megmutatható, hogy létezik egy valós $\theta \in (0, 1]$ amivel

$$\lim_{n \rightarrow \infty} P \left\{ \left(\tilde{M}_n - d_n \right) / c_n \leq x \right\} = H(x)$$

valamilyen nem elfajuló $H(x)$ -re akkor és csak akkor, ha

$$\lim_{n \rightarrow \infty} P \left\{ (M_n - d_n) / c_n \leq x \right\} = H^\theta(x).$$

Ez a θ az idősorok elméletében *extremális index*ként ismert. Ezekre a folyamatokra teljesül, hogy az M_n maximumoknak létezik eloszlásbeli határértéke, amennyiben a hozzá tartozó \tilde{M}_n -nak létezik határértéke: tehát, ha $\exists \xi \in \mathbb{R}$, hogy $F \in MDA(H_\xi)$. Könnyen látható, hogy H_ξ^θ eltolás- és skálaparaméter erejéig megegyezik H_ξ -vel, ezért az erősen stacionárius idősorok maximumainak határeloszlása szintén GEV eloszlás, ugyanazzal a ξ paraméterrel, mint az idősorhoz tartozó iid mintáé; csak az eltolás- és skálaparaméter változik. Elég nagy n -re az előzőek alapján felírható:

$$P(M_n \leq u) \approx P^\theta(\tilde{M}_n \leq u) = F^{n\theta}(u),$$

tehát nagy u esetén egy θ extremális indexszel rendelkező idősorból származó n megfigyelés maximumának eloszlása jól közelíthető $n\theta < n$ darab független megfigyelés maximumának eloszlásával.

Bár nem minden erősen stacionárius folyamatra van ilyen θ , a számunkra jelenleg fontos, pénzügyi modellezésben gyakran használt folyamatok esetében θ létezik:

- Fehér zaj folyamatok (iid valószínűségi változók) esetén $\theta = 1$
- ARMA folyamatokra Gauss fehér zaj esetben $\theta = 1$, amennyiben az innovációs folyamat eloszlása $MDA(H_\xi)_{\xi > 0}$ -beli, akkor $\theta < 1$
- ARCH és GARCH folyamatok esetén $\theta < 1$.

Gyakorlatban a fenti eredmények azt jelentik, hogy erősen stacionárius idősorok vizsgálatakor ugyanolyan pontos becsléshez több (konkrétan $1/\theta$ -szor annyi) adatra lesz szükségünk, mint független esetben. Például, ARCH(1) folyamatra $\alpha_1 = 0,9$ esetén $\theta = 0.612$, tehát iid mintához képest 1.64-szer annyi adatra lesz szükségünk.

1.2. Küszöbmeghaladások, általánosított Pareto eloszlás

Egy másik, pénzügyi modellekben gyakrabban alkalmazott határeloszlás az általánosított Pareto eloszlás (Generalized Pareto Distribution, GPD).

1.2.1. Definíció (Általánosított Pareto eloszlás (GPD)). A GPD eloszlásfüggvénye:

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi} & \xi \neq 0, \beta > 0 \\ 1 - \exp(-x/\beta) & \xi = 0, \beta > 0 \end{cases}$$

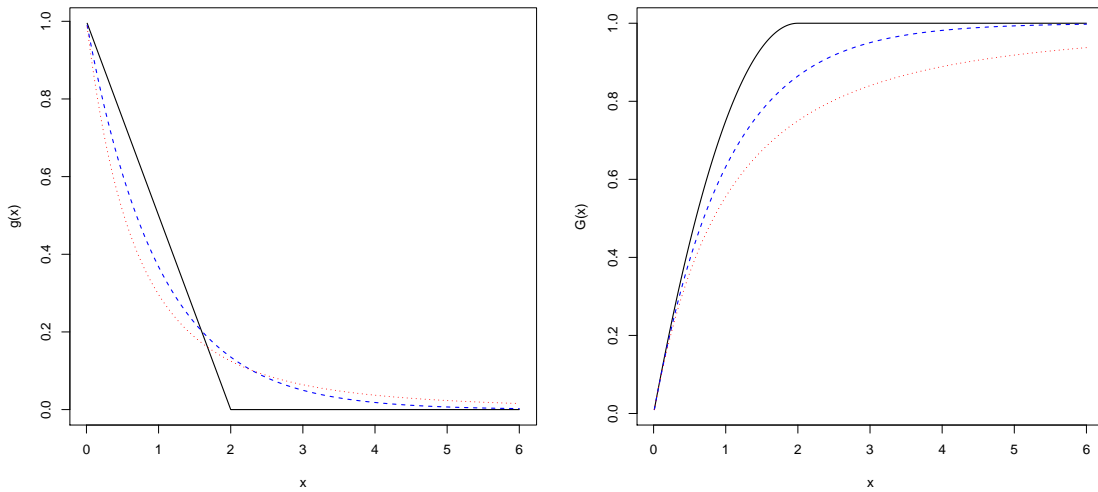
ahol $x \geq 0$ ha $\xi \geq 0$ és $0 \leq x \leq -\beta/\xi$ ha $\xi < 0$.

A GPD eloszlásoknál is 3 esetet különböztetünk meg (1.2 ábra):

- $\xi < 0$ esetén Pareto II, egy felülről korlátos tartójú eloszlás,
- $\xi = 0$ esetén a jól ismert exponenciális eloszlás,
- $\xi > 0$ esetén Pareto eloszlás, $\alpha = 1/\xi$ és $\kappa = \beta/\xi$ paraméterekkel.

Az extrémérték-eloszláshoz hasonlóan a GPD is ξ -ben folytonos, és k -adik momentuma $k \geq 1/\xi > 0$ esetén végtelen. Várható értéke $\xi < 1$ esetén $E(X) = \beta/(1 - \xi)$. A max-vonzási tartományoknál tárgyalt karakterizációs tételekből (1.1.5 és 1.1.7 tétel) következik, hogy $G_{\xi,\beta} \in MDA(H_\xi)$.

A küszöbmeghaladásos modellben fontos szerepet fog játszani a minta egy bizonyos értéket meghaladó elemeinek eloszlása.



1.2. ábra. Az általánosított Pareto-eloszlás eloszlás sűrűségfüggvényei (balra) és eloszlásfüggvényei (jobbra). Folytonos vonallal $\xi = -0.5$ (Pareto II), szaggatottal $\xi = 0$ (exponenciális), pontozottal $\xi = 0.5$ (Pareto) esetben. Mindhárom változatnál $\beta = 1$.

1.2.2. Definíció (a meghaladás feltételes eloszlása). Adott X valószínűségi változó F eloszlásfüggvénnyel. Ekkor X -nek az u küszöb meghaladása esetén a meghaladás feltételes eloszlása:

$$F_u(x) = P(X - u \leq x | X > u) = \frac{F(x + u) - F(u)}{1 - F(u)}$$

Ahol $0 \leq x < x_F - u$, és $x_F \leq \infty$ jelöli az eloszlás jobboldali végpontját.

1.2.3. Definíció (Várható meghaladás függvény). Egy véges várható értékű X valószínűségi változó várható meghaladási függvénye:

$$e(u) = E(X - u | X > u).$$

F_u -t gyakran maradék élet függvénynek is nevezik, mivel azt adja meg, hogy egy eddig u ideig működő komponens mekkora eséllyel hibásodik meg az $(u, u + x)$ időintervallumban. $e(u)$ pedig az u ideig működő komponens u időpont után várható élettartamát határozza meg. Általánosított Pareto eloszlás esetén ezek könnyen kiszámolhatók:

- $\xi = 0$ esetben az exponenciális eloszlás örökifjú tulajdonsága miatt $F_u(x) = F(x)$,

- Általában, legyen $F = G_{\xi, \beta}$, $\beta(u) = \beta + \xi u$. Ekkor

$$\begin{aligned}
F_u(x) &= \frac{F(x+u) - F(u)}{1 - F(u)} = \frac{-\left(1 + \frac{\xi x + \xi u}{\beta}\right)^{-1/\xi} + \left(1 + \frac{\xi u}{\beta}\right)^{-1/\xi}}{\left(1 + \frac{\xi u}{\beta}\right)^{-1/\xi}} \\
&= -\left(\frac{1 + \frac{\xi x + \xi u}{\beta}}{1 + \frac{\xi u}{\beta}}\right)^{-1/\xi} = 1 - \left(\frac{\frac{\xi x + \xi u + \beta}{\beta}}{\frac{\xi u + \beta}{\beta}}\right)^{-1/\xi} \\
&= 1 - \left(\frac{\frac{\xi x + \xi u + \beta}{\beta}}{\frac{\xi u + \beta}{\beta}}\right)^{-1/\xi} = 1 - \left(1 + \frac{\xi x}{\beta + \xi u}\right)^{-1/\xi} = G_{\xi, \beta(u)}
\end{aligned}$$

ahol $0 \leq u < \infty$ ha $\xi \geq 0$ és $0 \leq x \leq -(\beta/\xi) - u$ ha $\xi < 0$, és

$$e(u) = \frac{\beta(u)}{1 - \xi} = \frac{\beta + \xi u}{1 - \xi}.$$

Hasonlóan belátható, hogy $v > u$ esetén $F_v(x) = G_{\xi, \beta + \xi(v-u)}$ és $e(v) = \frac{\xi v}{1 - \xi} + \frac{\beta - \xi u}{1 - \xi}$. $e(v)$ linearitása a modellezésben alapvető jelentőségű karakterizációját adja a GPD-nek.

A következő tétel kapcsolatot teremt a korábban tárgyalt GEV eloszlás és a GPD között:

1.2.4. Tétel (Pickands, 1975; Balkema-de Haan, 1974). *Akkor és csak akkor létezik olyan mérhető $\beta(u)$ függvény, amire*

$$\lim_{u \rightarrow x_F} \sup_{0 \leq x < x_F - u} |F_u(x) - G_{\xi, \beta(u)}(x)| = 0,$$

ha $F \in MDA(H_\xi)$, valamilyen $\xi \in \mathbb{R}$ esetén.

A tétel szerint tehát az u küszöb növelésével minden $MDA(H_\xi)$ -beli eloszlás farokeloszlása GPD-hez konvergál, ráadásul a GPD ξ paramétere ugyanaz, mint a GEV esetben.

2. fejezet

Extrémérték - modellek

2.1. Blokk maximumok módszere

Tegyük fel, hogy adott egy ismeretlen F eloszlásból származó minta. Először a GEV eloszlás segítségével szeretnénk valamilyen előrejelzést adni a jövőben várható maximális érték(ek)re. Ilyen eset előfordulhat pl. folyamszabályozási terveknél (mekkora az a vízszint, aminél bizonyos valószínűséggel nem lesz nagyobb a következő időszakban). Feltesszük, hogy $F \in MDA(H_\xi)$, és hogy a mintánk vagy független, azonos eloszlású, vagy egy extrémális indexszel rendelkező folyamatból származik. Ekkor elég nagy n -re az n elemű minta maximuma M_n , és ennek határeloszlása egy háromparaméteres GEV eloszlás, $H_{\xi,\mu,\sigma}$. A gyakorlatban úgy tekintjük, hogy M_n eloszlása már pontosan GEV eloszlás.

Természetesen az eloszlás illesztéséhez több ilyen n méretű adatsorra van szükségünk, így az összes adatot felosztjuk m darab, n méretű blokkra. A gyakorlatban ezek a blokkok általában valamilyen jól kezelhető, természetes felosztást jelentenek, például napi adatok esetén negyedéves, féléves, vagy éves bontást használhatunk. A kapott m darab maximumra illesztjük a GEV eloszlást. A j -edik maximumot M_{nj} -vel jelölve, kapjuk az M_{n1}, \dots, M_{nm} mintát. Erre a GEV eloszlás illesztése pl. maximum likelihood, vagy momentum módszerrel történhet. A blokk méretét, n -et lehetőleg minél nagyobbra szeretnénk választani, egyrészt hogy a maximum megfigyelések függetlennek tekinthetők legyenek (maximum likelihood módszernél), másrészt hogy az (1.1) egyenlet alapján a határeloszlást jól lehessen becsülni. Másrészt viszont, a blokkok számát, m -et is megfelelően nagynak kell választani, hogy az eloszlás illesztésénél minél kisebb hibát kapjunk. Megjegyzendő még, hogy nem független adatoknál nagyobb n érték szükséges (azonos pontosságú becsléshez n/θ) mint iid esetben.

$h_{\xi,\mu,\sigma}$ -val jelölve a GEV eloszlás sűrűségfüggvényét, a log-likelihood függvény az alábbi módon írható fel:

$$\begin{aligned} l(\xi, \mu, \sigma, M_{n1}, \dots, M_{nm}) &= \sum_{i=1}^m \ln h_{\xi,\mu,\sigma}(M_{ni}) \\ &= -m \ln \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^m \ln \left(1 + \xi \frac{M_{ni} - \mu}{\sigma}\right) - \\ &\quad - \sum_{i=1}^m \left(1 + \xi \frac{M_{ni} - \mu}{\sigma}\right)^{-1/\xi}, \end{aligned}$$

ami a $\sigma > 0$, $1 + \xi(M_{ni} - \mu)/\sigma > 0 : i = 1, \dots, m$ feltételek mellett maximalizálendő. A kapott ML becslés létezik, ráadásul aszimptotikusan hatásos és konzisztens, ha $\xi > -\frac{1}{2}$, tehát teljesül a következő (Smith, 1985):

$$\sqrt{n} \begin{pmatrix} \hat{\xi}_n - \xi \\ \hat{\beta}_n - \beta \end{pmatrix} \xrightarrow{d} N \left[0, \begin{pmatrix} (1 + \xi)^2 & \beta(1 + \xi) \\ \beta(1 + \xi) & 2\beta^2(1 + \xi) \end{pmatrix} \right].$$

A modell használatával két jelentős kérdésre adhatunk választ:

1. Adott méretű maximális érték várhatóan milyen gyakran következik be (a visszatérési idő problémája)?

H -val jelölve a megfelelő GEV eloszlást, az $\{M_n > u\}$ esemény visszatérési ideje $k_{n,u} = 1/(1 - H(u))$.

Ez azt jelenti, hogy várhatóan $k_{n,u}$ darab, n méretű blokkból egy lesz, ahol a maximum meghaladja u -t. Becslése:

$$\hat{k}_{n,u} = 1 / \left(1 - H_{\hat{\xi}, \hat{\mu}, \hat{\sigma}}(u)\right).$$

2. Adott időintervallumban mekkora a legnagyobb érték, ami várhatóan bekövetkezik (a visszatérési szint problémája)?

H -val jelölve a megfelelő GEV eloszlást, a k idő alatt várhatóan bekövetkező legnagyobb érték $r_{n,k} = q_{1-1/k}(H)$, azaz a H eloszlás $(1 - 1/k)$ -kvantilise. Jelentése: az az érték, amelyet átlagosan minden k darab, n méretű blokk esetén egyszer lépünk túl. Becslése:

$$\hat{r}_{n,k} = H_{\hat{\xi}, \hat{\mu}, \hat{\sigma}}^{-1} \left(1 - \frac{1}{k} \right) = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}} \left(\left(-\ln \left(1 - \frac{1}{k} \right) \right)^{-\hat{\xi}} - 1 \right).$$

Vegyük észre, hogy a definíciók szerint $r_{n,k_{n,u}} = u$.

2.2. Küszöbmeghaladások

A GEV modell hátránya, hogy minden blokkból csak egy maximális értékkel dolgozik, így a rendelkezésre álló adatoknak csak nagyon kis részét használja. Egy másik modell, a küszöbmeghaladások modellje minden olyan adatot extrémálisnak tekint, és felhasznál az eloszlás-illesztés során, amelyek meghaladnak egy bizonyos küszöbértéket. Ez pl. viszontbiztosítási szerződések árazásakor lehet fontos. Ezért ez a modell a pénzügyi modellezés területén jobban használható - máshol az összefüggőségek és esetleges periódusok kiszűrése miatt a GEV modell alkalmazása is elterjedt.

2.2.1. GPD modell

Adott tehát X_1, \dots, X_n mintánk egy ismeretlen F eloszlásból. Az alábbi feltételezésekkel élünk:

- $F \in MDA(H_\xi)$ valamilyen $\xi \in \mathbb{R}$ -re
- Megfelelően nagy u küszöb választása esetén $F_u(x) = G_{\xi, \beta}(x)$ valamilyen $\beta > 0$ -val.

A mintánk u -t meghaladó értékeit jelöljük $\tilde{X}_1, \dots, \tilde{X}_{N_u}$ -val, ahol N_u a küszöbmeghaladások száma. Legyen $Y_j = \tilde{X}_j - u$ a küszöbmeghaladások értéke. A GPD modell illesztéséhez maximum likelihood módszert használunk, a log-likelihood függvény:

$$\begin{aligned} l(\xi, \beta, Y_1, \dots, Y_{N_u}) &= \sum_{j=1}^{N_u} \ln g_{\xi, \beta}(Y_j) \\ &= -N_u \ln \beta - \left(1 + \frac{1}{\xi} \right) \sum_{j=1}^{N_u} \ln \left(1 + \xi \frac{Y_j}{\beta} \right), \end{aligned}$$

ami a $\beta > 0$, $1 + \xi Y_j / \beta > 0$: $j = 1, \dots, N_u$ feltételek mellett maximalizálandó. A kapott ML becslés $G_{\hat{\xi}, \hat{\beta}}$.

A küszöb megfelelő megválasztásához a korábban tárgyalt várható meghaladás függvényt hívjuk segítségül. Láttuk, hogy $v > u$ esetén $e(v)$ v-ben lineáris. Ennek alapján, empirikus módszerrel határozzuk meg u értékét. Szükségünk lesz $e(v)$ becslésére:

$$e_n(v) = \frac{\sum_{i=1}^n (X_i - v) \chi_{\{X_i > v\}}}{\sum_{i=1}^n \chi_{\{X_i > v\}}}$$

Ennek segítségével ábrázolhatjuk az $\{(X_{i,n}, e_n(X_{i,n}))\}$ pontokat, ahol $X_{i,n}$ a rendezett minta i -edik eleme. Amennyiben az adatainkra illeszkedik a GPD modell, ez a függvény magasabb $X_{i,n}$ értékek után közel lineárisává válik. Növekedő trend $\xi > 0$, közel vízszintes függvény $\xi = 0$, a csökkenés pedig $\xi < 0$ értékeket jelez. Érdekes az utolsó néhány pontot figyelmen kívül hagyni, mivel ott már kis számú, nagy értékű adatok átlagos értékét vesszük, és gyakran ezek az értékek közel sem lineárisan helyezkednek el. A küszöböt a lineáris trend elejénél határozzuk meg. A paraméterek becslésére lehetséges módszer itt is a maximum likelihood becslés, ekkor a GPD sűrűségfüggvényét $g_{\xi, \beta}$ -val, a küszöbmeghaladás mértékét $Y_i = X_i - u$ -val jelölve, a log-likelihood függvényt:

$$\begin{aligned} \ln L(\xi, \beta, Y_1, \dots, Y_n) &= \sum_{j=1}^{N_u} \ln g_{\xi, \beta}(Y_j) \\ &= -N_u \ln \beta - \left(1 + \frac{1}{\xi}\right) \sum_{j=1}^{N_u} \ln \left(1 + \xi \frac{Y_j}{\beta}\right), \end{aligned}$$

a $\beta > 0$, $1 + \xi Y_j / \beta > 0$, $j = 1, \dots, N_u$ feltételek mellett maximalizálva kapjuk a paraméterek $\hat{\xi}$, $\hat{\beta}$ becslését.

2.2.2. Farokeloszlások és kockázati mértékek becslése

A GPD modell egyik fontos felhasználási területe a kockázatelemzésben a különböző kockázati mértékek (value at risk, expected shortfall) becslése. Ehhez először szükségünk lesz a kezdeti F eloszlásfüggvény farokeloszlásának meghatározására. $x \geq u$ esetén:

$$\begin{aligned}\bar{F}(x) &= 1 - F(x) = P(X > u) P(X > x | X > u) \\ &= \bar{F}(u) P(X - u > x - u | X > u) \\ &= \bar{F}(u) \left(1 + \xi \frac{x - u}{\beta}\right)^{-1/\xi}.\end{aligned}\tag{2.1}$$

Innen invertálva kaphatjuk az α -kvantilist, azaz VaR_α -t. $\alpha \geq F(u)$ esetén:

$$VaR_\alpha = q_\alpha(F) = u + \frac{\beta}{\xi} \left(\left(\frac{1 - \alpha}{\bar{F}(u)} \right)^{-\xi} - 1 \right).$$

$\xi < 1$ esetén ebből tovább számolható az ES:

$$ES_\alpha = \frac{1}{1 - \alpha} \int_\alpha^1 q_x(F) dx = \frac{VaR_\alpha}{1 - \xi} + \frac{\beta - \xi u}{1 - \xi}$$

továbbá a kettő hányadosának határértéke:

$$\lim_{\alpha \rightarrow 1} \frac{ES_\alpha}{VaR_\alpha} = \begin{cases} (1 - \xi)^{-1}, & 1 > \xi \geq 0 \\ 1, & \xi < 0 \end{cases}.$$

Látható, hogy a VaR és az ES becsléséhez szükségünk van egyrészt az illetett modell paramétereire, másrészt $\bar{F}(u)$ becslésre. Ez utóbbi legegyszerűbben a

tapasztalati N_u/n becslésből adódik, amennyiben elég sok adatunk van a küszöb felett. (2.1) alapján viszont egy jobb becslést remélünk $x \geq u$ esetben: (Smith, 1987)

$$\hat{F}(x) = \frac{N_u}{n} \left(1 + \hat{\xi} \frac{x - u}{\hat{\beta}} \right)^{-1/\hat{\xi}}$$

Ezután VaR_α és ES_α becslése az illesztett paraméterek, és a becsült farokeloszlás behelyettesítésével történik.

2.2.3. Viszontbiztosítási szerződés árazása

Ebben az esetben egy magas összegre szóló viszontbiztosítási réteget szeretnék árazni, a rendelkezésünkre álló, múltbeli biztosítási adatok alapján. A szerződés szerint, adott r és R alsó és felső határok, és az Y -al jelölt kifizetés a következőképpen alakul (X -el jelölve az összes kárkifizetést):

$$Y = \begin{cases} 0, & X < r \\ X - r, & r \leq X < R \\ R - r, & R \leq X \end{cases}$$

Két kérdés vetődik fel:

- Árazás: adott r és R határok, és múltbeli adatok mellett mi lesz az ára ennek a rétegnek?
- Határok megválasztása: hogyan határozzuk meg az alsó r határt, hogy a károk viszontbiztosítóhoz kerülésének valószínűsége legfeljebb p legyen?

Árazási probléma. A viszontbiztosító teljes kifizetése: $S = \sum_{i=1}^N Y_i$, ahol az Y_i kárnagyság és az N kárszám egy valószínűségi változó. A népszerű díjkalkulációs elvek általában Z momentumainak kiszámolásán, megbecslésén alapulnak. Például a szórásnégyzet elvénél $Pr = E(Z) + kD^2(Z)$. Könnyen látható, hogy $E(Z) = E(N)E(Y_1)$. N eloszlásának meghatározására léteznek egyéb kárszám modellek, Y_1 eloszlását, és ennek segítségével momentumait pedig a küszöbmeghaladásos modellel becsülhetjük.

Határok. Adott p esetén a feladat r megválasztása úgy, hogy $P(Z > 0) < p$ teljesüljön, azaz a biztosítási időszakban p -nél kisebb valószínűséggel történjen a viszontbiztosítóhoz kerülő káresemény. Ez az eredeti F_X eloszlás $(1 - p)$ -kvantilisének becslésére vezet.

2.3. A Hill-módszer

A ξ alakparaméter a becsléseink fő célpontja, mivel ez határozza meg az eloszlás farkvastagságát, az extrém értékek előfordulásának valószínűségét. A Hill-módszer egy alternatív lehetőség ξ becslésére.

2.3.1. Tétel (Karamata). *Legyen L lassú változású, $[x_0, \infty)$ -ben lokálisan korlátos valamilyen $x_0 \geq 0$ -ra. Ekkor*

1. $\kappa > -1$ esetén $\int_{x_0}^x t^\kappa L(t) dt \sim \frac{1}{\kappa+1} x^{\kappa+1} L(x)$, amennyiben $x \rightarrow \infty$
2. $\kappa < -1$ esetén $\int_x^\infty t^\kappa L(t) dt \sim -\frac{1}{\kappa+1} x^{\kappa+1} L(x)$, amennyiben $x \rightarrow \infty$.

Feltételezzük, hogy az eredeti eloszlás $F \in MDA(H_\xi)$ valamilyen $\xi > 0$ értékre, azaz az 1.1.5 tétel szerint $\bar{F}(x) = L(x)x^{-1/\xi}$, ahol L egy lassú változású függvény (1.1.4 definíció). Célunk $\alpha := 1/\xi$ értékének meghatározása, az adott X_1, \dots, X_n megfigyelés alapján. Amennyiben egy X valószínűségi változó eloszlása az 1.1.5 tétel szerinti, akkor felírhatjuk a várható meghaladás függvényét $\ln X$ -re:

$$\begin{aligned} e^*(\ln u) &= E(\ln X - \ln u \mid \ln X > \ln u) \\ &= \frac{1}{\bar{F}(u)} \int_u^\infty (\ln x - \ln u) dF(x) \\ &= \frac{1}{\bar{F}(u)} \int_u^\infty \frac{\bar{F}(x)}{x} dx \\ &= \frac{1}{\bar{F}(u)} \int_u^\infty L(x) x^{-(\alpha+1)} dx. \end{aligned}$$

Elég nagy u esetén $L(x)$ konstansként kiemelhető az integrálból ($x \geq u$). Karamata tétele alapján, amennyiben $u \rightarrow \infty$:

$$e^*(\ln u) \sim \frac{L(u)u^{-\alpha}\alpha^{-1}}{\bar{F}(u)} = \alpha^{-1},$$

így $\lim_{u \rightarrow \infty} \alpha e^*(\ln u) = 1$.

Ezt a tulajdonságot felhasználva, kiszámoljuk a sorba rendezett tapasztalati értékek logaritmusából a várható meghaladás ábráját, $e_n^*(\ln X_{k,n})$ - kat. Ekkor $\hat{\xi}^{-1} = \hat{\alpha}$ becslése:

$$\hat{\alpha}_{k,n}^{(H)} = \left(\frac{1}{k} \sum_{j=1}^k \ln X_{j,n} - \ln X_{k,n} \right)^{-1}, 2 \leq k \leq n.$$

Ennek k ellenében történő ábrázolásával kapjuk a Hill-ábrát, amin olyan stabil részt keresünk, ahol a becült paraméterek hasonlóak a különböző k értékek esetén.

3. fejezet

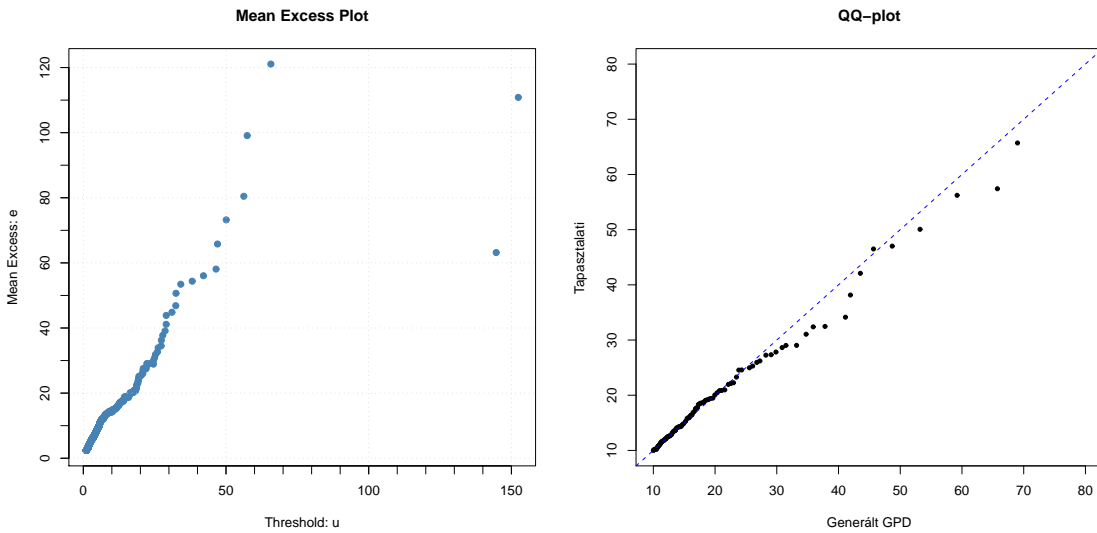
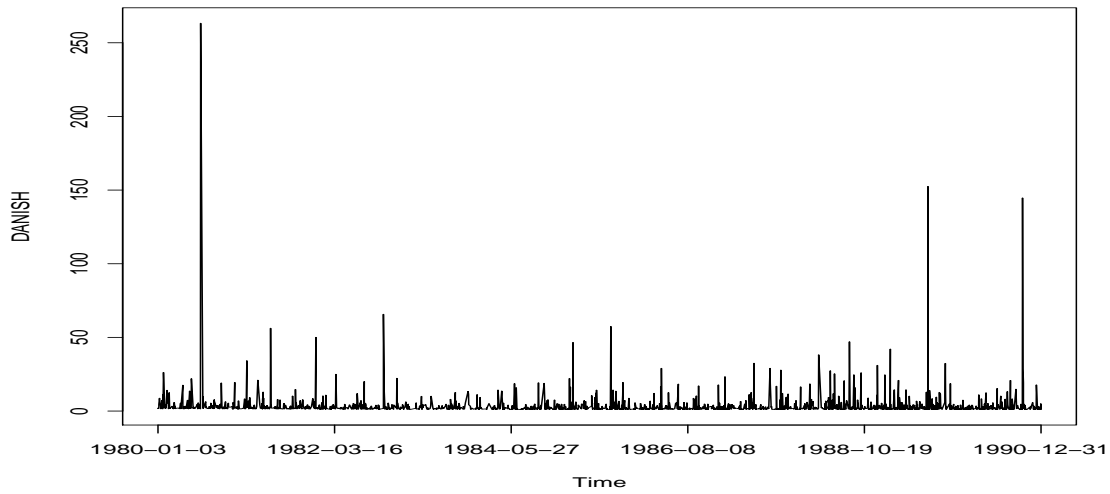
Adatelemzés

Ebben a fejezetben két adatsorral fogok foglalkozni. Az egyik a témában klasszikusnak számító dán tűzkárok idősora, a másik pedig egy 2000.-2004. közötti, magyarországi biztosítási kifizetéseket tartalmazó adatsor. A fent tárgyalt küszöbmeghaladásos GPD modell alapján próbálok majd következtetéseket levonni, valamint paramétereket becsülni. Ezek alapján a paraméterek alapján illesztett modellek később akár XL viszontbiztosítási szerződések árazásához, akár nagy károk tartalékának számításához, akár a Basel II követelményrendszerhez használható VaR és ES becsléséhez használhatók.

3.1. Dán tűzkárok

3.1.1. Küszöb-választás

Az adatsor 1980. január 1.-1990 december 31. között bekövetkezett 1 millió dán korona feletti összegű károkat tartalmazza, a teljes kárnagyságot 1985.-re indexálva, millió koronában. Összesen 2167 adat, minimális érték 1, maximális érték 263.25, átlag 3.38. A becsült $e_n(v)$ függvényt vizsgálva próbálunk megfelelő küszöb-értéket keresni. Ez itt az egész adatsoron közel lineárisnak tűnik, talán a 10-es küszöbnél van benne egy törés, ezért válasszuk u -nak ezt az értéket. Ezt a küszöböt az összes 5%-a, 109 adat lépi túl. A pozitív meredekség miatt $\xi > 0$ lesz a modellben, az ML becslések értékére $\hat{\xi} = 0.497$ és $\hat{\beta} = 6.975$ adódott, a standard hibák rendre 0.14 és 1.1. Az adatok idősora, $e_n(v)$, és az illesztett eloszlás ellenőrzésére szolgáló QQ-plot látható a 3.1 ábrán.



3.1. ábra. A dán tűzadatok idősora(fent), a mean excess plot $e(v)$ (lent, balra) és a QQ-plot az illesztett eloszlás ellenében ($u=10$).

Az illesztett eloszlás egy igen lassú lecsengésű, vastagfarkú, közel végtelen második momentummal rendelkező GPD. A várható meghaladás ábráját jobban megvizsgálva láthatjuk, hogy nem csak a 10 jöhet szóba, mint lehetséges küszöb, a 20-as értéknél is van egy kis „hullám”, valamint akár az egész ábrára ráfogható, hogy közel lineáris képet mutat. Az is észrevehető, hogy a QQ-plot a 20-as érték felett kissé mintha „lefelé hajló” tendenciát mutatna a referencia-egyeneshez képest. A különböző küszöbök esetén illesztett modellekből persze különböző paraméter-bebecsléseket kapunk, valamint egy modellen belül is kiszámolhatjuk megnövelt küszöb esetén az

értékeket. Az optimális küszöb megválasztásához több szempontot is figyelembe kell vennünk: a modellnek a küszöb emelésére való érzékenységét, a rendelkezésre álló adatok mennyiségét, valamint üzleti szempontokat.

Nézzük meg, mit kapunk $u = 1$ és $u = 20$ esetben (utolsó oszlopban a ξ -re vonatkozó 97,5%-os konfidencia-intervallum):

u	$\hat{P}(X > u)$	$\hat{\xi}$	$\hat{\beta}$	97.5%
1	1	0.186	2.578	(0.154, 0.219)
10	0.05	0.497	6.975	(0.230, 0.764)
20	0.017	0.684	9.635	(0.145, 1.223)

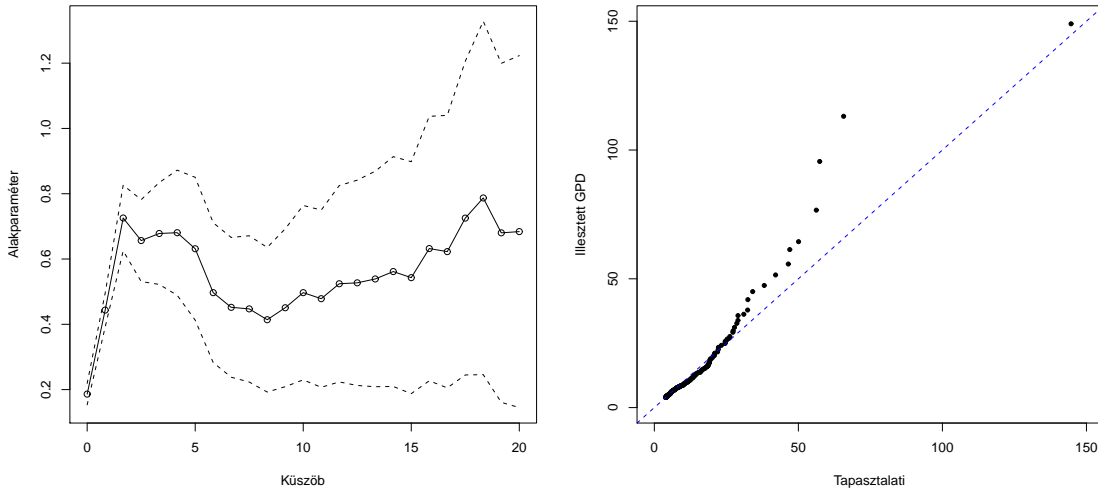
3.1. táblázat. *Különböző küszöb-értékek esetén illesztett modellek*

A harmadik változatnál már csak 36 adatra illesztettük a GPD-t, így viszonylag nagy standard hibát mutat a becslés (ez a konfidencia-intervallumokból jól látszik). Tudjuk, hogy $v > u$ esetén $F_v(x) = G_{\xi, \beta + \xi(v-u)}$, így olyan küszöböt szeretnénk választani, ami felett a ξ paraméter értéke közel állandó. A 3.2 ábrán látszik, hogy ez nagyjából a 7 és 15 közé eső értékeknél teljesül, tehát azt mondhatjuk, hogy a modell ebben az intervallumban a legmegbízhatóbb.

Természetesen az eloszlás jobb illeszkedése miatt szeretnénk minél alacsonyabb küszöböt választani. A 7-9 körüli értékek viszont a legkisebb ξ értékeket eredményezik, így biztosítási szempontból nem túl szerencsés a használatuk, mivel alulbecsülik a farokeloszlást, és ezzel a várható veszteséget. Ezért gyakorlati megfontolásból inkább egy kisebb küszöböt szeretnénk alkalmazni, így igaz, hogy a modell nem lesz olyan pontosan illeszkedő, de eleget tesz a konzervatív árazás követelményének, és nagyobb biztonságot ad a viszontbiztosítónak. Ilyen küszöb lehet pl. az $u = 4$ választás. Ekkor az ML-becslés eredményei:

$ \{X : X > u\} $	$\hat{P}(X > u)$	$\hat{\xi}$	$\hat{\beta}$	97.5%
362	0.167	0.72	2.63	(0.531, 0.91)

3.2. táblázat. *$u = 4$ küszöb-re illesztett modell*



3.2. ábra. Az alakparaméter(ξ) becsült értékei és a hozzá tartozó 97.5%-os konfidencia-intervallumok különböző küszöb-választások esetén (balra) és az $u = 4$ küszöb-re illesztett modell QQ-plotja (jobbra)

Innen, és a QQ-plot alapján is látszik, hogy ez a modell túlbecsüli a farokeloszlást, így a várható veszteséget is, és egy biztonságosabb díjszabást eredményez.

3.1.2. Kvantilisek, díjkalkuláció, érzékenység

A viszontbiztosítási réteg határainak meghatározásakor fontos szerepet játszik egy kár viszontbiztosítóhoz kerülésének valószínűsége, azaz az illesztett eloszlás valamilyen magas rendű kvantilise. Ez adja meg azt az alsó határt, amit várhatóan csak egy megadott kis valószínűséggel halad meg egy kár értéke. GPD esetén a p -kvantilis értéke: $u + \beta/\xi \left((1 - p)^{-\xi} - 1 \right)$.

A viszontbiztosítási termék árazásához alapvetően a két oldalról levágott biztosítási kifizetések várható értékét vizsgáljuk, mivel ez döntő szerepet játszik a különböző díjelvekben. Ekkor egy $[r, R]$ viszontbiztosítási réteg ára:

$$P = E(Y_1) = \int_r^R (x - r) f_{X_1}(x) dx + (R - r) (\bar{F}_{X_1}(R)),$$

ahol F_{X_1} és f_{X_1} az eredeti káreloszlás eloszlás- és sűrűségfüggvénye. A GPD modellt

egy megfelelően magas $u < r$ küszöbre illesztve $x > u$ esetén becsülhető \bar{F}_{X_1} :

$$\hat{F}_{X_1}(x) = \frac{N_u}{n} \left(1 + \hat{\xi} \frac{x - u}{\hat{\beta}} \right)^{-1/\hat{\xi}},$$

ahol $\hat{\xi}$ és $\hat{\beta}$ a ML-becsléssel kapott paraméterek, N_u/n pedig a minta küszöböt meghaladó elemeinek relatív gyakorisága, azaz a küszöbmeghaladás tapasztalati becslése. A fenti képletet deriválva kaphatjuk a sűrűségfüggvény becslését, így lehetővé válik az ár kiszámítása. Mivel az illesztett eloszlás farkának vastagsága, és így a nagyobb károk bekövetkezésének valószínűsége a $\hat{\xi}$ becsült paraméter értékétől függ, így természetesen ennek magasabb értékei magasabb árakat fognak eredményezni. Az alábbi táblázat összefoglalja különböző küszöb-értékek választása esetén az illeszkedő modell fontosabb tulajdonságait, $r = 50$, $R = 200$ esetén:

u	$\hat{P}(X > u)$	$\hat{\xi}$	$\hat{\beta}$	97.5%	.999	P
3	0.246	0.668	2.189	(0.524, 0.811)	129	0.21
4	0.167	0.720	2.632	(0.531, 0.910)	147	0.24
5	0.117	0.631	3.809	(0.413, 0.850)	122	0.19
10	0.050	0.497	6.975	(0.230, 0.764)	95	0.13
20	0.017	0.684	9.635	(0.145, 1.223)	103	0.15

3.3. táblázat. *Paraméter-becslések, a 0.999-kvantilis és a számított ár különböző küszöbök esetén.*

A 0.999-kvantilis számítása során az összehasonlíthatóság és gyakorlati szemlélet miatt a következőképpen jártunk el: a táblázatban szereplő érték az illesztett GPD $1 - 1/\left(1000 \left(\hat{P}(X > u)\right)\right)$ -kvantilise, így a teljes adatsorra adja meg a 0.001 valószínűséggel túllépendő alsó határt. Ennek tapasztalati értéke 144 - ez a harmadik legkisebb adat az összes 2167 közül. (A negyedik legnagyobb adat viszont már jóval kisebb, csak 65.)

Láthattuk, hogy a modell paraméterei, és így az azokból számolt adatok is igen érzékenyek a küszöb megválasztására. Egy másik ilyen tényező, amire figyelnünk

kell, az adatsor legnagyobb elemeinek viselkedése. amennyiben elhagyunk vagy hozzáadunk kiugró értékeket, a modell alapvetően megváltozhat. A küszöb-választásra nézve viszonylag stabil, $u = 10$ -es modellt 3 különböző sokk-szenárióban vizsgálva a következő eredményeket kapjuk:

<i>sokk</i>	$\hat{\xi}$	$\hat{\beta}$.999
alapeset	0.497	6.975	95
legnagyobb kár nélkül	0.390	7.230	77
3 legnagyobb kár nélkül	0.167	7.932	53
+1 nagy kár, 350 értékkel	0.597	6.783	117

3.4. táblázat. *Becsült paraméterek és a 0.999-quantilis különböző sokkok esetén, $u = 10$.*

Látható, hogy a legnagyobb károkkal manipulálva a modellben komoly változások történnek, akár már 1-2 érték megváltoztatása esetén is. Ezért érdemes folyamatosan nyomon követni a károk alakulását, és a használt modellt ennek megfelelően frissíteni.

A fenti eredmények birtokában, szeretnénk eldönteni, hogy hogyan válasszuk meg a modellt a viszontbiztosítási réteg megfelelő díjkalkulációjához. Az első ránézésre próbált 10-es küszöbről kiderült, hogy valójában gyakorlati szempontból nem túl szerencsés választás, mivel alulbecsüli a farokeloszlást, ezzel megnövelve a viszontbiztosítóra háruló kockázatot. 10-nél nagyobb küszöbértékek esetén a modell már igen kevés adatot használ, ezért instabillá válik, $u = 20$ -nál például a 97.5%-os konfidencia-intervallumba beletartozik a már-már exponenciálisra hasonlító $\xi = 0.15$ és a végtelen várható értékű $\xi = 1.2$ eset is, így itt az eloszlás illesztésének hibája okoz magas kockázatot. Az $u = 4$ esetet a magas ára miatt valószínűleg a piaci viszonyok zárják ki. $u = 3$ esetben már az adatok közel negyede a küszöb fölé esik, így az alaptételként használt határeloszlás közelítése megkérdőjelezhető. Mindent összevetve az $u = 5$ küszöbérték tűnik megfelelő kiindulási alapnak, esetleg a piaci viszonyok figyelembe vételével ettől negatív irányba eltérve nő, pozitív irányba eltérve csökken a $\hat{\xi}$ értéke, és ezzel a biztosítás díja. Ezt a modellt átskálázva árazhatók a magasabb értékű viszontbiztosítási szerződések. A jövőben modell folyamatos felülvizsgálata szükséges, különös tekintettel a legnagyobb beérkező károkra.

3.2. Magyarországi biztosítási adatok

3.2.1. Küszöb-választás

A nyers adatsor több mint 47,000 biztosítási kifizetést tartalmaz 2000. és 2004. között. Az adatokat először a KSH fogyasztóiár-index alakulása alapján 2000-es értékre indexáltuk, majd a könnyebb kezelhetőség érdekében csak az 1,000,000 Ft feletti értékeket tartottuk meg. Ez a levágás a modellt nem befolyásolja. Az így kapott 2528 adat legnagyobb értékeinek viselkedését próbáljuk jellemezni. A főbb jellemzők millió forintban kifejezve:

n	$\hat{E}(X)$	$\hat{D}^2(X)$	$\min(X_i)$	$\max(X_i)$
2528	3.126	61.91	1	191

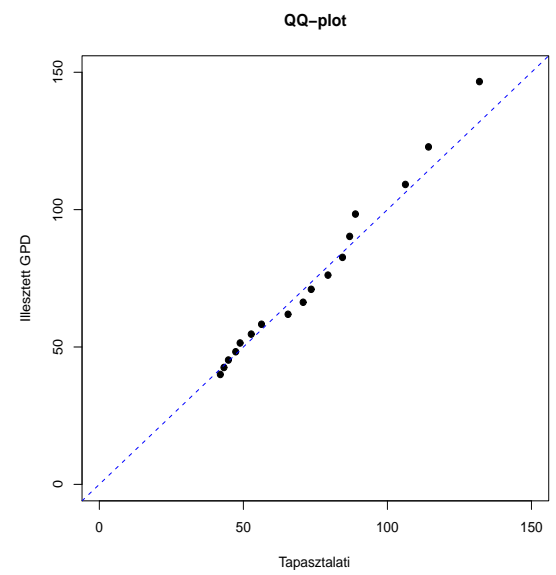
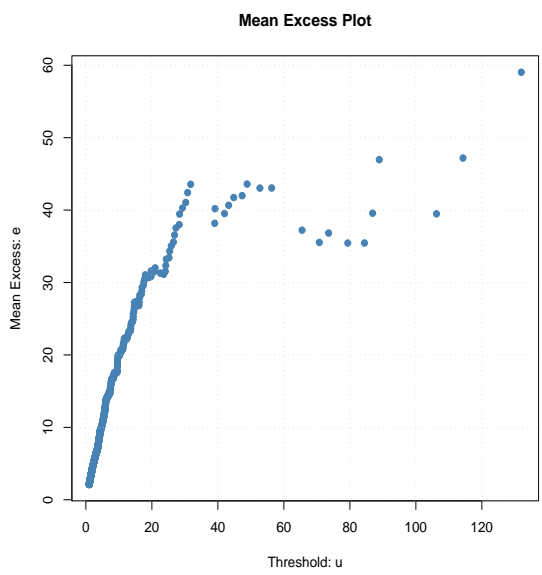
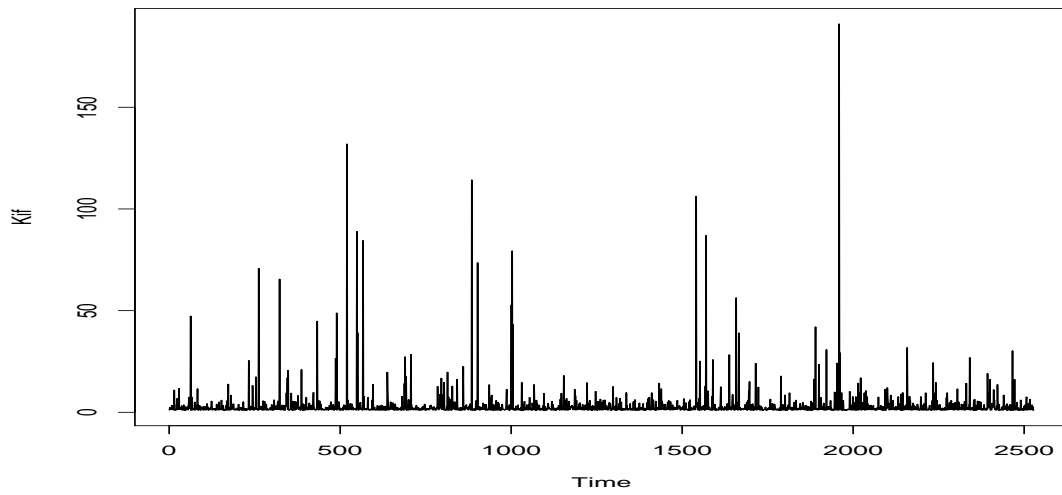
3.5. táblázat. *A magyarországi biztosítási adatok főbb jellemzői*

Biztosítási környezetben a káradatok függetlenségének feltételezése általában nem okoz problémát. Az autokovariancia-függvény alapján ez jelen esetben is így van. Kezdetben a fentiekhez hasonlóan járunk el, a várható meghaladás ábráján keresünk lineárishoz közelítő részt (3.3 ábra). Ennél az adatsornál az előzővel ellentétben nem a viszontbiztosító, hanem a direkt biztosító szemszögéből próbálunk hasznos dolgokat megfigyelni a károk viselkedésével kapcsolatban. Látható, hogy a várható meghaladás ábrája a 40-es érték felett erősen megváltozik, igen nagy szóródást mutat. Észrevehető viszont egy bizonyos vízszinteshez közeli szóródás. Az illesztett modell paraméterei láthatók a 3.6 táblázatban, az illeszkedést ellenőrző QQ-plot a 3.3 ábrán.

u	$\hat{P}(X > u)$	$\hat{\xi}$	$\hat{\beta}$	97.5%
40	0.0071	-0.075	42.223	(-0.571, 0.421)

3.6. táblázat. $u = 40$ küszöb-választás esetén illesztett modell paraméterei

Látható, hogy ez a modell erősen instabil, a levágott adatoknak csak a 7 ezrelékét, 18 db megfigyelést használ, és a konfidencia-intervallum is igen nagy. Ennyi adatból



3.3. ábra. A magyarországi biztosítási adatok idősora(fent), a mean excess plot $e(v)$ (lent, balra) és a QQ-plot az illesztett eloszlás ellenében ($u = 40$).

nem lehet biztonságosan becsülni a legnagyobb károkat. Az itt jelentkező kockázat kezelésére ekkor több lehetőség van:

- A kockázat teljes átadása, azaz egy megfelelő (legfeljebb 40 millió Ft) alsó határú XL viszontbiztosítási szerződés kötése. Ekkor az összes nagy értékű kifizetés egy fix értékre mérséklődik, cserébe pedig rendszeres díjat kell fizetnünk a viszontbiztosítónak.
- A kockázat túlbecslése, azaz ξ értékének magas megválasztása. Ekkor egy kevésbé jól illeszkedő modellt kapunk, ami nagy valószínűséggel túlbecsüli a legnagyobb károk mértékét és bekövetkezési valószínűségét, viszont ekkor kisebb eséllyel érik a biztosítót váratlanul nagy veszteségek.

A várható kárkifizetés mindkét esetben nagyobb lesz, mint az először illesztett modelnél - az első változatban ξ megnövelt értéke miatt nő a küszöb feletti károk várható értéke, a másodikban a viszontbiztosító díjkalkulációs politikájától függően kell megfizetnünk a kockázati felárat. Viszontbiztosítás vásárlása esetén a kárfolyamatunk $S' = \sum_{i=1}^N \min(X_i, r)$ -re változik, ahol r a viszontbiztosítási szerződés határa. Ez felírható továbbá

$$\begin{aligned} S' &= \sum_{i=1}^N \min(X_i, r) \\ &= \sum_{i=1}^{N_1} X_i \chi_{\{X_i < r\}} + \sum_{j=1}^{N_2} r \\ &= \sum_{k=1}^{N_1} Y_k + N_2 r \end{aligned}$$

alakban, ahol N_1 és N_2 kárszámok eloszlása az eredeti N kárszám eloszlásának $p = P(X_1 < r)$ -el illetve $1 - p$ -vel ritkított változatai.

Nézzük most a második esetet. Sajnos a várható meghaladás ábrája most nem olyan segítőkész, mint a dán tűzadatok esetén - egy lehetséges megközelítés $e(u)$ legnagyobb értékeinek figyelmen kívül hagyása. Ekkor a következő szóba jövő küszöbérték $u = 24$ lehet az ábra alapján, így érdemes ezt a lehetőséget is megvizsgálni:

u	$\hat{P}(X > u)$	$\hat{\xi}$	$\hat{\beta}$	97.5%
24	0.0138	0.437	19.134	(-0.251, 1.125)

3.7. táblázat. $u = 24$ küszöb-választás esetén illesztett modell paraméterei

$\hat{\xi}$ becsült értéke itt már valamelyest hihetőbb képet mutat: az illesztett eloszlás véges várható értékű és szórású. A felhasznált adatok kis mennyisége (38 db) és a várható meghaladás utolsó értékeinek figyelmen kívül hagyása miatt a konfidencia-intervallum itt is igen nagy lett. Tovább haladva a dán tűzadatoknál bejárt úton, nézzünk meg néhány további küszöb-értéket:

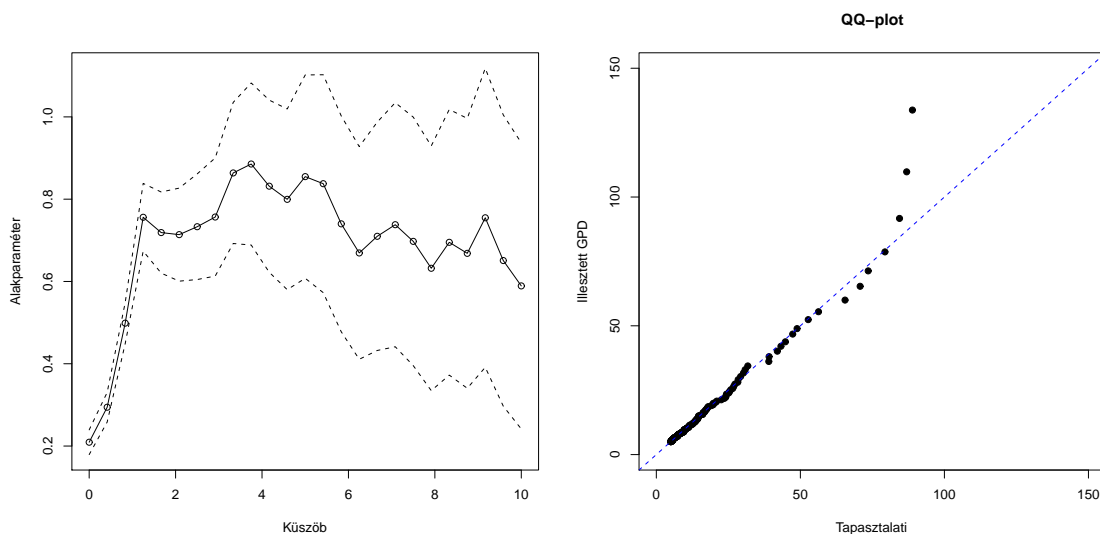
u	$\hat{P}(X > u)$	$\hat{\xi}$	$\hat{\beta}$	97.5%	0.999
1	1	0.752	0.633	(0.683, 0.820)	151.93
5	0.096	0.855	3.241	(0.607, 1.102)	188.95
10	0.036	0.589	9.754	(0.241, 0.937)	130.13
20	0.0154	0.234	24.266	(-0.236, 0.714)	112.93
24	0.0138	0.437	19.134	(-0.251, 1.125)	118.08
30	0.0091	-0.057	41.698	(-0,499, 0.385)	116.52
40	0.0071	-0.075	42.223	(-0.571, 0.421)	116.97

3.8. táblázat. A modell paraméterei további küszöb-értékek esetén

Megfigyelhető, hogy a 20-as határt az adataink nagyon kis része lépi túl: a levágott adatsor alig 1.5%-a, az eredeti adatok nagyjából 8 tízezreléke. Így további vizsgálatok tárgyául csak a 10 és annál kisebb küszöb-értékeket tartjuk meg.

3.2.2. Érzékenység

A dán adatsorhoz hasonlóan itt is szeretnénk megnézni, mennyire stabil a modell, hogyan reagál az újonnan beérkező adatokra. A küszöb választására nézve a ξ



3.4. ábra. Az alakparaméter(ξ) becsült értékei és a hozzá tartozó 97.5%-os konfidencia-intervallumok különböző küszöb-választások esetén (balra) és az $u = 5$ küszöb-re illesztett modell QQ-plotja (jobbra)

alakparaméter értéke igen változó, többnyire 0.6 és 0.8 között van (3.4 ábra). Nagyjából 1-5 között látható egy viszonylag egyenes érték, így a határ választására nézve ebben az intervallumban a legstabilabb a modell. Látható továbbá, hogy ezek a küszöb-értékek adják a legmagasabb ξ -értékeket is. $u = 5$ esetén a különböző sokk-szenáriókban becsült értékeket tartalmazza a 3.9 táblázat:

<i>sokk</i>	$\hat{\xi}$	$\hat{\beta}$.999
alapeset	0.855	3.241	188.95
legnagyobb kár nélkül	0.819	3.268	167.83
3 legnagyobb kár nélkül	0.760	3.300	139.38
+1 nagy kár, 250 értékkel	0.894	3.209	212.64

3.9. táblázat. Becsült paraméterek és a 0.999-quantilis különböző sokkok esetén, $u = 5$

Ennél az adatsornál az előzővel ellentétben viszont az látszik, hogy a modell kevésbé érzékeny a legnagyobb értékek változásaira, $\hat{\xi}$ értéke egytizednyi határon belül marad a sokkok hatására. Azt látjuk tehát, hogy a kevésbé jó illeszkedés,

nagyobb konfidencia-intervallumokért cserébe egy jóval időt állóbb modellt kaptunk, ami stabilabban viselkedik a megfigyelt farokeloszlás változásaira nézve. Ráadásul itt a konzervativizmus jegyében ajánlott, magasabbra paraméter-értékeket produkáló u küszöbök egybeesnek a több szempontból is legstabilabb modellekhez vezetőekkel.

A QQ-plotok alapján az is észrevehető, hogy a legnagyobb károk jól becsülhetők a magas, $u = 40$ küszöb használatával illesztett modellel, viszont itt igen kevés adatunk van, így gyakorlatilag minden újonnan beérkező, 40 millió Ft feletti kár esetén újra kell vizsgálni ezt a modellt. $u = 5$ esetben a modell jól illeszkedik a legtöbb adatra, a legnagyobbak kivételével - azokat erősen túlbecsüli -, így amennyiben ezeket XL viszontbiztosítás alá helyeztük, akkor ajánlott az alacsonyabb küszöb-választás.

3.3. Összegzés

Látható, hogy a GPD modell illesztése közel sem olyan egyértelmű, mint amilyenek először tűnik. Az igazán nagy megfigyelések kis száma és a modell legnagyobb értékekre való érzékenysége miatt megfelelő körültekintéssel, gyakorlati szempontokat is figyelembe véve kell eljárni a küszöb választása esetén. A modellek állandó felülvizsgálata szükséges, mivel akár 1-2 évnyi plusz megfigyelés is komolyan változtathat az illeszkedő paramétereken. Hasznos lehet továbbá a különböző modellek összehasonlítása (a fenti két adatsorra vonatkozó Hill-ábra a Függelék 3.ábrán látható). Amennyiben a függetlenségre vonatkozó feltevésünk nem megalapozott, érdemes az összefüggő adatokat jobban kezelő GEV modellt is megvizsgálni (az elemzett adatsorokra vonatkozó - függetlenségi feltevésünket alátámasztó - autokovariancia-ábrák és az extrémális indexek becslései a Függelék 1. és 2. ábrán láthatóak).

Irodalomjegyzék

- [1] P. Embrechts A. J. McNeil, R. Frey. *Quantitative Risk Management*. Princeton University, 2005.
- [2] L. de Haan A.A. Balkema. Residual life time at great age. *Annals of Probability*, 2:792–804, 1974.
- [3] B.V. Gnedenko. Limit theorems for the maximal term of a variational series. *Comptes Rendus (Doklady) de L'Académie des Sciences de l'URSS*, 32:7–9, 1975.
- [4] B.M. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3:1163–1174, 1975.
- [5] A. J. McNeil. Estimating the tails of loss severity distributions using extreme value theory. 1999.
- [6] G. Samorodnitsky P. Embrechts, S. I. Resnick. Extreme value theory as a risk management tool. 1996.
- [7] T. Mikosch P. Embrechts, C. Klüppelberg. *Modelling Extremal Events for Insurance and Finance*. Springer, 1997.
- [8] J. Pickands. Statistical inference using extreme order statistics. *Annals of Statistics*, 3:119–131, 1975.
- [9] L.H.C. Tippett R.A. Fischer. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24:180–190, 1928.
- [10] S.I. Resnick. *Extreme Values, Regular Variation and point processes*. Springer, 1987.
- [11] R.L. Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72:67–92, 1985.
- [12] Wikipedia. Extreme value theory.

Függelék

Konfidencia-intervallumok

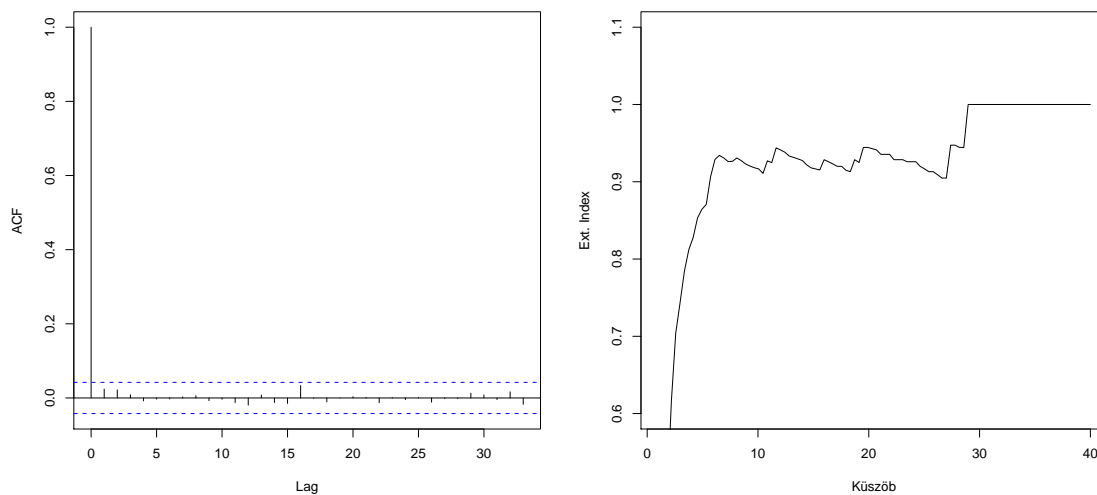
A bemutatott számításokban szereplő konfidencia-intervallumok számolására egy jól használható módszer a likelihood-hányados próba. $\boldsymbol{\theta}$ paraméter becslése esetén a nullhipotézis $H_0 : \boldsymbol{\theta} \in \Theta_0$, az ellenhipotézis pedig $H_1 : \boldsymbol{\theta} \in \Theta \setminus \Theta_0$, $\Theta_0 \subset \Theta$. Ekkor a likelihood-hányados statisztika:

$$\lambda(\mathbf{X}) = \frac{\sup_{\boldsymbol{\theta} \in \Theta_0} L(\boldsymbol{\theta}; \mathbf{X})}{\sup_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}; \mathbf{X})}$$

feltételezve, hogy $\mathbf{X}_1, \dots, \mathbf{X}_n$ iid minta, és teljesülnek a regularitási feltételek. Ekkor a nullhipotézis mellett $-2 \ln \lambda(\mathbf{X}) \sim \chi^2_\nu$, $\nu = p - q$, ahol p a Θ által meghatározott szabad paraméterek száma, q pedig a nullhipotézisben meghatározott szabad paraméterek száma. Esetünkben (az alakparaméter ML-becslése) $p = 2$: (β, ξ) és $q = 1$: (β) . Az α szintű konfidencia-intervallumokat ezután $\{\xi : -2 \ln \lambda(\mathbf{X}) \leq c_{1-\alpha}\}$ adja meg, ahol $c_{1-\alpha}$ a χ^2 -eloszlás $(1 - \alpha)$ -kvantilise. Az így számított intervallum nem mindig szimmetrikus.

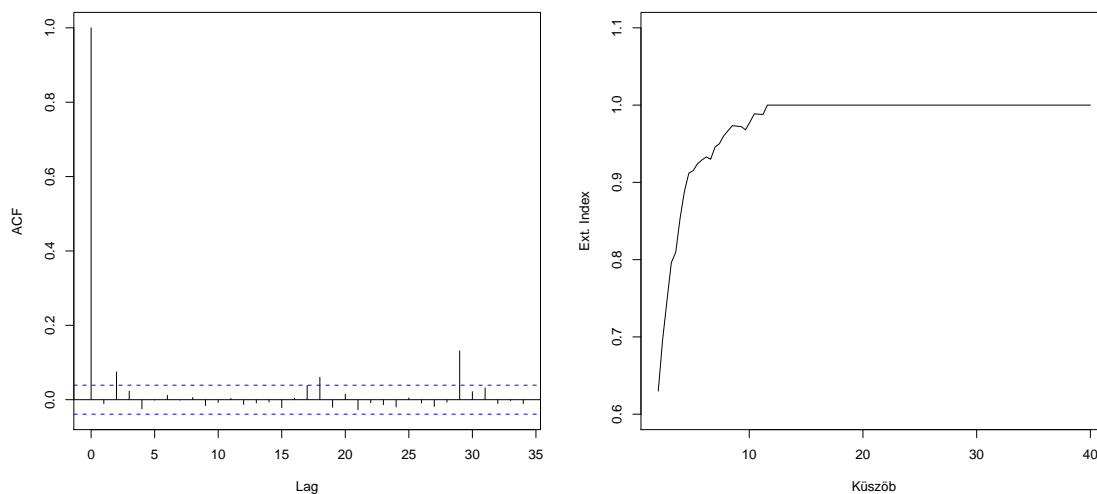
Extremális indexek és autokovariancia-ábrák

Dán tűzkár-adatok



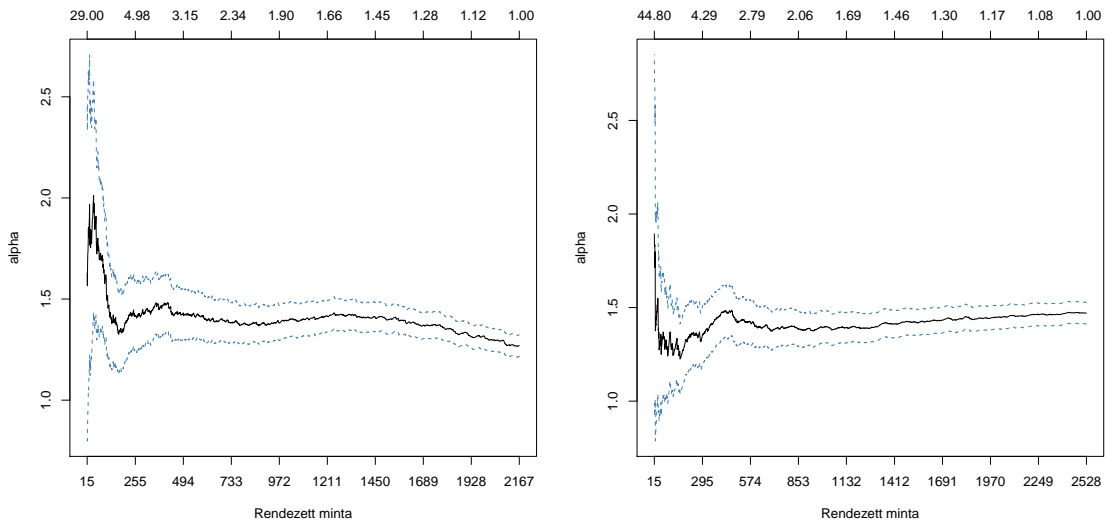
3.5. ábra. A dán tűzkár-adatokra vonatkozó függetlenségi feltevést alátámasztó autokovariancia-ábra (balra), és az extrémális index becslése (jobbra).

Magyar biztosítási adatok



3.6. ábra. A magyar biztosítási adatokra vonatkozó függetlenségi feltevést alátámasztó autokovariancia-ábra (balra), és az extrémális index becslése (jobbra).

Hill-ábrák



3.7. ábra. *Egy alternatív módszer az alakparaméter becslésére: balra a dán tűzadatokra, jobbra a magyar biztosítási adatokra vonatkozó Hill-ábra. A felül lévő beosztás a választott küszöböt, az alul lévő a küszöböt meghaladó megfigyelések számát mutatja*

Felhasznált szoftverek

A szakdolgozatban szereplő számítások és ábrák az R programban készültek, az *fExtremes* (Diethelm Wuertz és sokan mások, 2009) és az *evd* (Alec Stephenson, Chris Ferro, 2008) csomagok felhasználásával.

A szakdolgozat L^AT_EX (ver.1.6.9)-ben készült.

Köszönetnyilvánítás

Ezúton szeretnék köszönetet mondani témavezetőmnek, Zempléni Andrásnak türelméért, tanácsaiért és a kölcsön adott szakirodalomért. Az utolsó napokban többször is átnézte a dolgozatot, felhívta a figyelmem a legapróbb matematikai és nyelvtani hibákra is, amiért külön hálás vagyok.