

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



Použitie teórie extrémnych hodnôt vo finančníctve

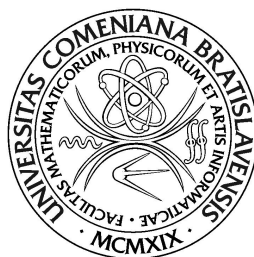
DIPLOMOVÁ PRÁCA

Bratislava 2008

Enikő Kováčsová

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

Katedra aplikovanej matematiky a štatistiky



Použitie teórie extrémnych hodnôt vo finančníctve

DIPLOMOVÁ PRÁCA

Špecializácia: Ekonomická a finančná matematika

Diplomant: Enikő Kováčsová

Vedúca diplomovej práce: Mgr. Alexandra Urbánová Csajková, PhD.

Bratislava 2008

Čestne prehlasujem, že som diplomovú prácu vypracovala samostatne
s využitím teoretických vedomostí a s použitím uvedenej literatúry.

V Bratislave 2008

.....

Podakovanie

Touto cestou by som sa chcela poďakovať svojej vedúcej diplomovej práce Alexandre Urbánovej Csajkovej za jej odborné vedenie, cenné rady a za množstvo času, ktoré mi venovala pri vypracovávaní diplomovej práce.

Abstrakt

Teória extrémnych hodnôt modeluje zriedkavé udalosti mimo rozsahu prípustných pozorovaní s veľkým dopadom. Táto metóda sa stala v posledných rokoch veľmi využívaným aparátom na odhadovanie rizika. Signifikantné zmeny na finančných trhoch viedli k vylepšeniu riadenia rizík a hľadaniu nových metód.

Cieľom diplomovej práce je vysvetliť teóriu extrémnych hodnôt a aplikovať na reálne dáta. Budeme sa zaoberať s dvoma metódami modelovania extrémov: metódou blokového maxima a metódou prekročenia ponad prah. Na modelovanie príslušných rozdelení sa používa zovšeobecnené rozdelenie extrémnych hodnôt (GEV) a zovšeobecnené Pareto rozdelenie (GPD). V záverečnej časti využijeme teoretické poznatky na empirickú analýzu vybraných dát.

Kľúčové slová: Teória extrémnych hodnôt (EVT), rozdelenie GEV, rozdelenie GPD, blokové maximá, prekročenia ponad prah, model s prahovou hodnotou (POT)

Obsah

Abstrakt	1
Zoznam skratiek	4
Úvod	5
1 Klasická teória extrémnych hodnôt (EVT)	6
1.1 Formulácia modelu	6
1.2 Zovšeobecnené rozdelenie extrémnych hodnôt (GEV)	7
1.3 Maximálne vierohodný odhad (MLE)	10
1.4 Profil vierohodnosti	11
1.5 Overovanie modelu	11
2 Model s prahovou hodnotou (POT)	14
2.1 Formulácia modelu	14
2.2 Zovšeobecnené Pareto rozdelenie (GPD)	14
2.3 Metódy na stanovenie prahovej hodnoty	16
2.3.1 Výskumná metóda	17
2.3.2 Hodnotenie stability odhadov parametrov	18
2.4 ML odhad parametrov GPD rozdelenia	18
2.5 Overovanie modelu	19
3 Extrémy závislých časových radov	21
3.1 Extrémy stacionárnych postupností	21
3.1.1 Maximá stacionárnych postupností	21
3.1.2 Modelovanie stacionárnych radov	23
3.2 Extrémy nestacionárnych postupností	24
3.2.1 Odhad parametrov	24
4 Ciele práce	26
5 Praktická časť	27
5.1 Analýza dát	27
5.2 Modely blokového maxima, resp. minima	29
5.2.1 Kalibrácia modelov na intervale 9 rokov	30
5.2.2 Kalibrácia modelov na intervale 4 rokov	35
5.3 Modely POT	41
5.3.1 Kalibrácia modelov na intervale 9 rokov	41
5.3.2 Kalibrácia modelov na intervale 4 rokov	48
5.4 Interpretácia výsledkov	54

5.5	Stresové testovanie	56
5.5.1	Trhové riziko	56
5.5.2	Operačné riziko	57
6	Záver	58
	Príloha	60

Zoznam skratiek

EVT - Teória extrémnych hodnôt (Extreme Value Theory)

GEV - Zovšeobecnené rozdelenie extrémnych hodnôt (Generalized Extreme Value distribution)

MLE - Maximálne vierohodný odhad (Maximum Likelihood Estimation)

ML - Maximálna vierohodnosť (Maximum Likelihood)

GPD - Zovšeobecnené Pareto rozdelenie (Generalized Pareto Distribution)

POT - Peaks Over Threshold

MRL - Mean Residual Life plot

VaR - Miera rizika (Value-at-Risk)

NLLH - Negative log-likelihood

SE - Štandardná chyba

Úvod

Teória extrémnych hodnôt (EVT) je nástroj, ktorý sa používa pri skúmaní pravdepodobnosti extrémnych a zriedkavých udalostí s veľkým dopadom. Teória skúma konce rozdelení a preto jej implementácia čelí viacerým problémom. Na rozdiel od centrálnej limitnej vety, ktorá sa vzťahuje k limitnému rozdeleniu priemerov, EVT dáva limitné rozdelenie maxim nezávislých náhodných premenných. EVT sa používa napríklad na modelovanie kríz na devízovom a akciovom trhu, na modelovanie veľkých kreditných zlyhaní a na stresové testovanie. Najčastejšie sa používa v oblasti poisťovníctva, bankovníctva, operačného rizika, trhového rizika a kreditného rizika.

Diplomová práca pozostáva zo šiestich kapitol. Prvá kapitola vysvetľuje klasickú teóriu extrémnych hodnôt. Sú dva základné typy modelov: modely blokového maxima a modely prekročenia ponad prah (POT).

V prvej kapitole formulujeme model blokového maxima, ktorý modeluje maximálne pozorovania z veľkého množstva identicky rozdelených pozorovaní. Základná veta o extrémnych hodnotách nám hovorí o tom, ako vyzerá limitné rozdelenie extrémnych hodnôt. Zavedieme triedu rozdelení zovšeobecnených extrémnych hodnôt (GEV) a jej základné charakteristiky. Parametre rozdelenia sa odhadujú pomocou metódy maximálnej vierohodnosti. Na overovanie modelu používame grafickú analýzu dát.

V druhej kapitole formulujeme model prekročenia ponad prah, ktorý modeluje všetky pozorovania, ktoré prekročia nejakú vopred stanovenú prahovú hodnotu. Zdefinujeme triedu rozdelení zovšeobecneného Pareto rozdelenia. Uvedieme dva metódy na stanovenie prahovej hodnoty, a opíšeme základné vlastnosti týchto modelov.

Tretia kapitola popisuje extrémny závislých časových radov, a zaoberá sa modelovaním stacionárnych a nestacionárnych radov, ktoré je v praxi použiteľnejšie.

V štvrtej kapitole sú uvedené ciele práce.

Piatá kapitola sa venuje samotnej empirickej analýze, a analýze dosiahnutých výsledkov. Táto kapitola zahŕňa aj grafické znázornenie výsledkov. Prvá časť je venovaná metóde blokového maxima, resp. minima. Otestujeme modely pre rôzne dĺžky blokov. V druhej časti otestujeme model prekročenia ponad prah pre rôzne prahové hodnoty. Na záver interpretujeme dosiahnuté výsledky, a spomenieme ako môžeme teóriu extrémnych hodnôt prepojiť na prax.

Posledná, šiestá kapitola obsahuje záver diplomovej práce.

1 Klasická teória extrémnych hodnôt (EVT)

1.1 Formulácia modelu

Model je zameraný na štatistické správanie sa postupnosti

$$M_n = \max\{X_1, \dots, X_n\},$$

kde X_1, \dots, X_n je postupnosť nezávislých náhodných premenných s distribučnou funkciou F .

X_i – reprezentujú hodnoty procesu namerané v pravidelných časových intervaloch.

M_n – reprezentuje maximum procesu cez nejaké vopred stanovené obdobie pozorovania (napr. dĺžky n).

Rozdelenie M_n sa potom určí ako:

$$\begin{aligned} P\{M_n \leq z\} &= P\{X_1 \leq z, \dots, X_n \leq z\} = P\{X_1 \leq z\} \dots P\{X_n \leq z\} \\ &= \{F(z)\}^n \end{aligned}$$

Toto vyjadrenie nám často v praxi nepomôže, keďže F je neznáme rozdelenie. Jedna z možných riešení je používanie štandardných štatistických metód na odhadovanie F na základe pozorovaných dát. Môže sa ale stať, že menšie chyby spôsobia veľké odchýlky v F^n . Z toho dôvodu sa odhaduje F^n priamo pomocou extrémnych dát. Hladáme F^n , pre $n \rightarrow \infty$. Zaveďme normalizáciu postupnosti M_n takú, ktorá stabilizuje jej polohu a škálu pri rastúcom n :

$$M_n^* = \frac{M_n - b_n}{a_n}, \quad (1)$$

kde $\{a_n > 0\}$ a $\{b_n\}$ sú postupnosti konštant.

Veta 1.1. (Coles [1]) Ak existujú postupnosti konštant $\{a_n > 0\}$ a $\{b_n\}$ také, že :

$$P\{(M_n - b_n)/a_n \leq z\} \rightarrow G(z) \quad \text{pre } n \rightarrow \infty,$$

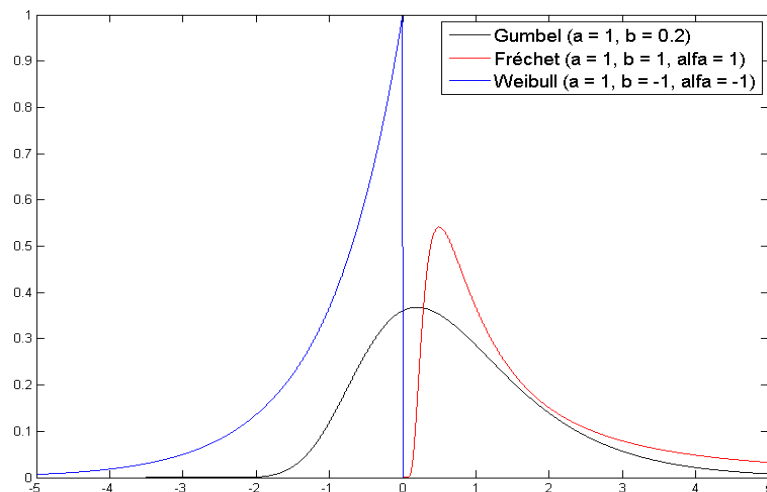
kde G je nedegenerovaná distribučná funkcia, tak G patrí do jednej z nasledujúcich tried:

$$\begin{aligned} 1. G_{EVT}(z) &= \exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\} && -\infty < z < \infty \\ 2. G_{EVT}(z) &= \begin{cases} 0 & z \leq b \\ \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\} & z > b \end{cases} \\ 3. G_{EVT}(z) &= \begin{cases} \exp\left\{-\left[\left(\frac{z-b}{a}\right)^\alpha\right]\right\} & z < b \\ 1 & z \geq b \end{cases} \end{aligned}$$

To znamená, že preškálované maximum konverguje v distribučnej funkcii k jednej z vyššie spomínaných troch tried. Prvé je **Gumbelovo** rozdelenie, druhé je **Fréchetovo** a tretie je **Weibullovo** rozdelenie. Všetky tri sa všeobecne nazývajú ako rozdelenie extrémnych hodnôt. Majú dva resp. tri parametre: $a > 0$ je parameter rozsahu (scale parameter), b je parameter polohy (location parameter), $\alpha > 0$ je parameter tvaru (shape parameter).¹

1.2 Zovšeobecnené rozdelenie extrémnych hodnôt (GEV)

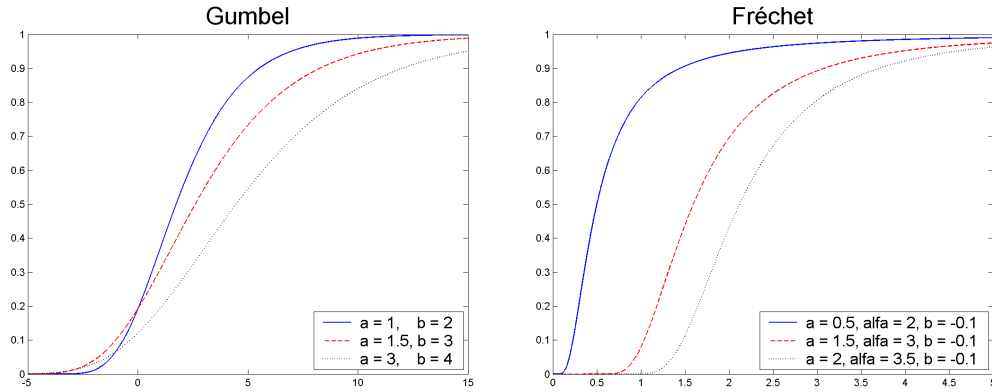
Limitné rozdelenia vo Vete 1.1. majú rôzne tvary koncov. Pre Gumbelovo a Fréchetovo rozdelenie z_+ (z_+ je koncový bod G_{EVT}) je nekonečný zatiaľ čo pre Weibullovo je konečný, t.j. pre Gumbelovo rozdelenie hustota G_{EVT} klesá exponenciálne, pre Fréchetovo polynomiálne, ako to je znázornené na Obrázku 1.



Obrázok 1: *Hustoty extrémnych hodnôt Gumbelovho, Fréchetovho a Weibullovo rozdelenia*

Pre nás najzaujímavejšie je Fréchetovo rozdelenie, lebo má tzv. ťažké konce a klesá polynomiálne. Na Obrázku 2 sú znázornené distribučné funkcie Gumbelovho a Fréchetovho rozdelenia pri zmene parametrov a, b resp. a, b, α .

¹Podrobnejšie charakteristiky rozdelení a limitné pravidlá sú odvodené v publikácii Fisher a Tippett [3], dôkaz Vety 1.1 formalizoval Gnedenko [4].



Obrázok 2: Distribučné funkcie Gumbelovho a Fréchetovho rozdelenia pre rôzne parametre

Tvrdenie 1.1. (GEV rozdelenie) (Coles [1])

Trieda rozdelení zovšeobecnených extrémnych hodnôt má tvar:

$$G_{GEV}(z) = \exp \left\{ - \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}. \quad (2)$$

Funkcia je definovaná na množine $\{z : 1 + \frac{\xi(z-\mu)}{\sigma} > 0\}$, kde parametre spĺňajú $-\infty < \mu < \infty$, $\sigma > 0$ a $-\infty < \xi < \infty$.

Poznámka 1.1: Tento model má tri parametre: μ -parameter polohy (location), σ -parameter rozsahu (scale) a ξ -parameter tvaru (shape).

- $\xi = 0 \rightarrow$ Gumbelovo rozdelenie
- $\xi > 0 \rightarrow$ Fréchetovo rozdelenie
- $\xi < 0 \rightarrow$ Weibullovo rozdelenie

Tvar rozdelenia G_{GEV} je ovplyvnený parametrom ξ . Pomocou parametra ξ dáta predurčujú triedu rozdelenia, t.j. nie je potrebné zvoliť vopred triedu z ktorej dáta pochádzajú. Na základe GEV rozdelenia sa dá preformulovať Veta 1.1.

Veta 1.2. (Coles [1]) Nech existujú postupnosti konštánt $\{a_n > 0\}$ a $\{b_n\}$ také, že

$$P\{(M_n - b_n)/a_n \leq z\} \rightarrow G(z) \quad \text{pre } n \rightarrow \infty$$

kde G je nedegenerovaná distribučná funkcia, potom G je člen GEV triedy:

$$G_{GEV}(z) = \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

na množine $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$, kde $-\infty < \mu < \infty$, $\sigma > 0$ a $-\infty < \xi < \infty$.²

Normalizované konštanty síce nie sú známe, ale zdanlivý problém sa dá obísť nasledovným spôsobom:

$$\begin{aligned} P\{(M_n - b_n)/a_n \leq z\} &\approx G_{GEV}(z) \\ P\{M_n \leq z\} &\approx G_{GEV}\{(z - b_n)/a_n\} = G_{GEV}^*(z) \end{aligned}$$

kde G_{GEV}^* je iný člen GEV triedy. Inak povedané, ak sa na základe Vety 1.2. M_n^* dá aproximovať rozdelením z GEV triedy, potom M_n sa dá aproximovať iným členom z tej istej triedy. Treba podotknúť, že parametre rozdelenia G_{GEV}^* sú rôzne od G_{GEV} .

Majme nezávislé dáta X_1, X_2, \dots . Ak určíme nejaký blok dĺžky n (n -dosťatočne veľké), tak postupnosť maxim týchto blokov bude Z_1, \dots, Z_m . Tie sa dajú potom aproximovať GEV rozdelením.

Odhad kvantilov pre rozdelenie je nasledovné:

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}[1 - \{-\log(1-p)\}^{-\xi}], & \text{pre } \xi \neq 0 \\ \mu - \sigma \log\{-\log(1-p)\}, & \text{pre } \xi = 0 \end{cases} \quad (3)$$

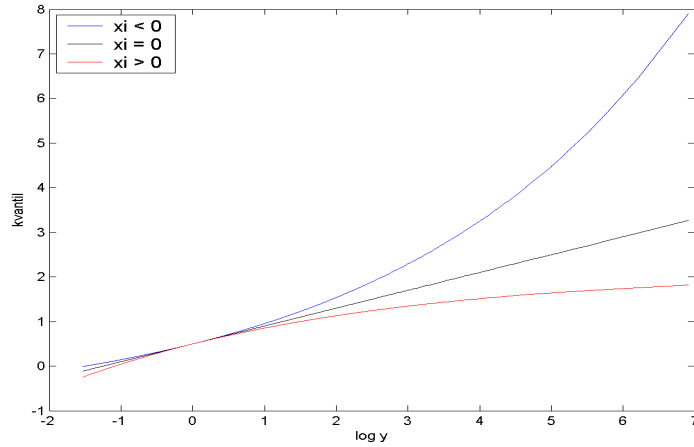
kde $G_{GEV}(z_p) = (1-p)$, z_p je miera návratnosti (return level), $1/p$ je perióda návratnosti. Hodnota z_p sa prekročí podľa očakávaní v priemere jeden krát každých $1/p$ rokov. Presnejšie, z_p sa prekročí ročným maximom v každom roku s pravdepodobnosťou p .

Ak definujeme $y_p = -\log(1-p)$, potom:

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}[1 - y_p^{-\xi}], & \text{pre } \xi \neq 0 \\ \mu - \sigma \log y_p, & \text{pre } \xi = 0 \end{cases}$$

Ak hodnoty z_p znázorníme oproti y_p v logaritmickej miere, dostaneme tzv. **'return level plot'**. Tento graf je lineárny ak $\xi = 0$, konvexný ak $\xi < 0$ a konkávny ak $\xi > 0$ (viď Obrázok 3). Používa sa predovšetkým na overovanie modelu.

²Podrobný dôkaz Vety 1.2. (Leadbetter et al.) [5]



Obrázok 3: *Return level plot GEV rozdelenia*

Ako sme už uvádzali, z dát sa vytvárajú bloky dĺžky n . Prirodzene vzniká otázka: aká je správna dĺžka bloku? Je to kompromis medzi varianciou a vychýlenosťou odhadov:

- ak sú bloky príliš malé, tak aproximácia je menej kvalitná,
- ak sú bloky veľké, potom variancia odhadu je veľká.

Poznámka 1.2: Skonstruovaný model v časti 1.1 pochopiteľne môžeme aplikovať aj na modely minimál blokov, len analogické výrazy treba cielene preformulovať.

$$\min(X_1, \dots, X_n) = -\max(-X_1, \dots, -X_n).$$

Existujú viaceré metódy na odhad parametrov GEV rozdelenia, napr. metódy založené na pravdepodobnostných grafoch, na momentoch, prípadne na vierohodnostných funkciách.

1.3 Maximálne vierohodný odhad (MLE)

Tvrdenie 1.2. (Coles [1]) Ak máme Z_1, \dots, Z_m nezávislé premenné s rozdelením GEV, potom 'log-likelihood' funkcia má nasledovný tvar: pre $\xi = 0$:

$$l(\mu, \sigma) = -m \log \sigma - \sum_{i=1}^m k_i - \sum_{i=1}^m \exp(-k_i) \quad (4)$$

pre $\xi \neq 0$:

$$l(\mu, \sigma, \xi) = -m \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^m \log(1 + \xi k_i) - \sum_{i=1}^m (1 + \xi k_i)^{-1/\xi} \quad (5)$$

pričom $k_i = \frac{z_i - \mu}{\sigma}$, za predpokladu, že $1 + \xi k_i > 0$ pre $i = 1, \dots, m$.

Maximalizáciou (4) alebo (5) dostaneme maximálne vierohodný odhad parametrov rozdelenia z GEV triedy. Analytické riešenie neexistuje, ale používajú sa štandardné numerické optimalizačné metódy. Maximálne vierohodný odhad pre z_p , ak $0 < p < 1$ je vyjadrený nasledovným spôsobom:

$$\hat{z}_p = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} [1 - y_p^{-\hat{\xi}}], & \text{pre } \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma} \log y_p, & \text{pre } \hat{\xi} = 0 \end{cases} \quad (6)$$

Poznámka 1.3: Pri používaní maximálne vierohodných odhadov sa môže vyskytnúť problém, že GEV model nespĺňa podmienky regularity. Smith [7] skúmal túto problematiku podrobnejšie, a získal nasledujúce výsledky:

- ak $\xi > -0.5$, potom odhad má asymptotické vlastnosti.
- ak $-1 < \xi < -0.5$, existuje odhad, ale nemá asymptotické vlastnosti.
- ak $\xi < -1$, odhad často neexistuje.

1.4 Profil vierohodnosti

Štandardná aproximácia rozdelenia ML odhadu môže byť menej kvalitná. Lepšiu aproximáciu dostávame pomocou funkcie profilu vierohodnosti. Ak chceme určiť profil vierohodnosti napríklad pre ξ , tak zafixujeme $\xi = \xi_0$, potom maximalizáciou funkcie (5) vypočítame ostatné parametre μ a σ metódou maximálnej vierohodnosti. Tento postup viackrát opakujeme pre rôzne zafixované ξ . Zodpovedajúce maximalizované hodnoty tvoria profil 'log-likelihood' pre ξ .

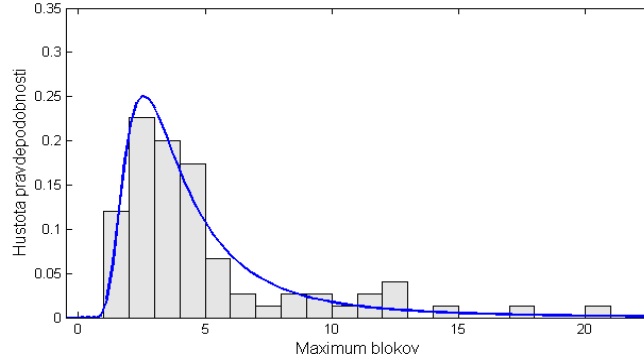
Poznámka 1.4: Samozrejme môžeme analogickým spôsobom určiť profil vierohodnosti pre ľubovoľnú kombináciu parametrov.

1.5 Overovanie modelu

Na overovanie modelu môžeme používať grafickú analýzu dát, ktorá zahŕňa napr. zostrojenie pravdepodobnostného grafu (P-P plot), kvantilového

grafu (Q-Q plot), grafu hustoty a 'return level plot'.

- **Histogram, graf hustoty:** histogram je nástroj vhodný na vizualizáciu frekvencie dát. Predstavuje typ stĺpcového grafu určený na znázornenie rozdelenia pravdepodobnosti (viď Obrázok 4).



Obrázok 4: Príklad histogramu GEV rozdelenia

- **P-P plot:** pravdepodobnostný graf je porovnanie empirického a odhadnutého pravdepodobnostného rozdelenia. Nech $z_1 \leq z_2 \leq \dots \leq z_m$ sú zoradené bloky maxím, potom empirická distribučná funkcia pre z_i je daná ako:

$$\tilde{G}_{GEV}(z_i) = \frac{i}{m+1}.$$

Do rovnice (2) dosadíme odhady parametrov $\hat{\xi}$, $\hat{\mu}$, $\hat{\sigma}$ tak, že:

$$\hat{G}_{GEV}(z_i) = \exp\left\{-\left[1 + \hat{\xi}\left(\frac{z_i - \hat{\mu}}{\hat{\sigma}}\right)\right]^{-1/\hat{\xi}}\right\},$$

potom P-P graf pozostáva z bodov:

$$\left\{\left(\tilde{G}_{GEV}(z_i), \hat{G}_{GEV}(z_i)\right), i = 1, \dots, m\right\}. \quad (7)$$

Poznámka: Model je dobrý ak $\hat{G}_{GEV}(z_i) \approx \tilde{G}_{GEV}(z_i)$ (t.j. graf je blízky diagonále). Tento graf nám poskytuje najmenej informácií v oblasti o ktorú máme najväčší záujem, t.j. chvostu rozdelenia.

- **Q-Q plot:** kvantilový graf je porovnanie kvantilu empirického a odhadnutého rozdelenia. Graf pozostáva z bodov:

$$\left\{\left(\hat{G}_{GEV}^{-1}\left(\frac{i}{m+1}\right), z_i\right), i = 1, \dots, m\right\}, \quad (8)$$

kde

$$\hat{G}_{GEV}^{-1}\left(\frac{i}{m+1}\right) = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} \left[1 - \left\{ -\log\left(\frac{i}{m+1}\right) \right\}^{-\hat{\xi}} \right].$$

Odlíšnosti od linearity indikujú zlyhanie modelu.

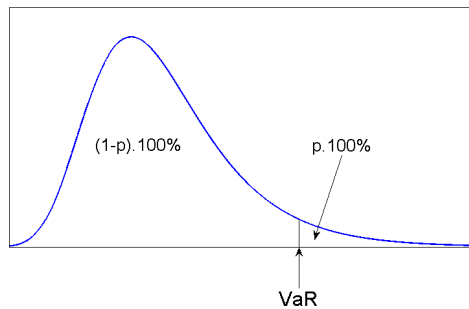
- **Return level plot** pozostáva z bodov:

$$\{(\log y_p, \hat{z}_p) : 0 < p < 1\}, \quad (9)$$

kde $y_p = -\log(1-p)$, a \hat{z}_p je maximálne vierohodný odhad z_p . Tento graf sa dá rozšíriť pomocou intervalov spoľahlivosti a taktiež pridaním empirického 'return level plot'-u.

Poznámka 1.5:

VaR predstavuje štandardný nástroj na meranie rôznych rizík, najmä trhového rizika alebo operačného rizika, v súčasnosti jeden z najčastejšie používaných v bankovom sektore. VaR je vlastne definovaná ako hodnota najväčšej možnej straty pri špecifikovanom stupni spoľahlivosti počas určitej periódy. Stupeň spoľahlivosti sa zvyčajne používa v intervale 90 až 99 percent (viď Obrázok 5). Táto miera rizika v podstate predstavuje opisnú štatistickú charakteristiku (kvantil) odhadovaného rozdelenia možných strát resp. ziskov.



Obrázok 5: Grafické znázornenie VaR na hladine významnosti $(1-p).100\%$

V našom prípade kvantil z_p vyjadrený pomocou výrazu (3) sa dá chápať ako '**extreme VaR**' napr. na hladine významnosti 99.5% alebo dokonca 99.9%.

2 Model s prahovou hodnotou (POT)

2.1 Formulácia modelu

Majme postupnosť X_1, X_2, \dots nezávislých, rovnako rozdelených náhodných premenných s distribučnou funkciou F . Je rozumné považovať za extrémne udalosti tie X_i , ktoré prekročia nejakú vopred zvolenú prahovú hodnotu u . Označíme ľubovoľný člen postupnosti X_i ako X , potom stochastické správanie sa extrémnych udalostí je dané podmienenou pravdepodobnosťou:

$$P\{X > u + y | X > u\} = \frac{1 - F(u + y)}{1 - F(u)}, \quad y > 0. \quad (10)$$

Ak poznáme rozdelenie F , tak poznáme aj rozdelenie prekročení prahovej hodnoty. V praxi ale F je neznáme rozdelenie.

2.2 Zovšeobecnené Pareto rozdelenie (GPD)

Veta 2.1. ³(Coles [1]) *Nech X_1, X_2, \dots je postupnosť nezávislých náhodných premenných so známou distribučnou funkciou F , a nech $M_n = \max\{X_1, \dots, X_n\}$. Predpokladajme, že F spĺňa Vetu 1.2., tak že pre dostatočne veľké u platí:*

$$P\{M_n \leq z\} \approx G_{GEV}(z),$$

kde

$$G_{GEV}(z) = \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

pre $\mu, \sigma > 0$ a ξ . Potom pre dostatočne veľké u distribučná funkcia výrazu $(X - u)$ za podmienky, že $X > u$ sa dá aproximovať pomocou:

$$H_{GPD}(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi}. \quad (11)$$

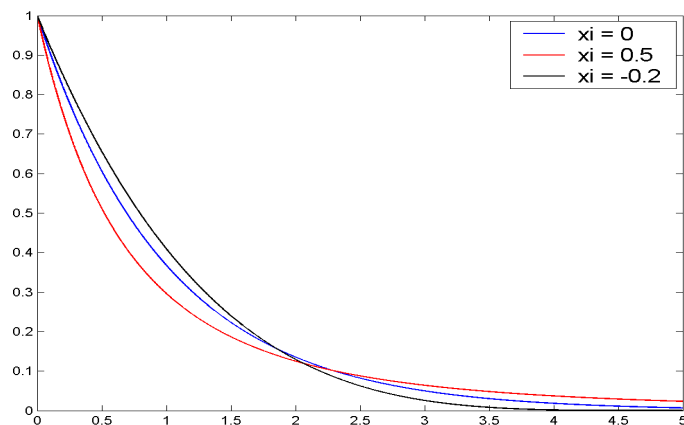
Funkcia je definovaná na $\{y : y > 0, (1 + \xi y/\tilde{\sigma}) > 0\}$, kde $\tilde{\sigma} = \sigma + \xi(u - \mu)$.⁴

Trieda rozdelenia definovaná výrazom (11) je trieda **zovšeobecneného Pareto rozdelenia** (GPD). Veta 2.1. tvrdí, že keď bloky maxím majú rozdelenie G_{GEV} , potom prekročenia hranice majú rozdelenie z Pareto triedy. Navyše aj parametre zovšeobecneného Pareto rozdelenia sú jednoznačne určené tými, ktoré zodpovedajú príslušnému GEV rozdeleniu.

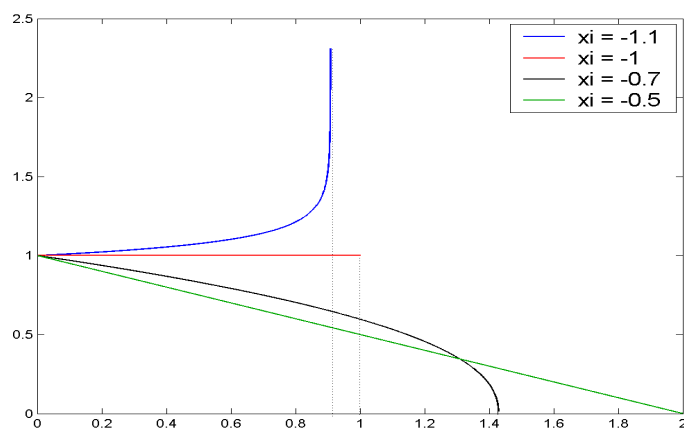
³podrobný dôkaz Vety 2.1. (Leadbetter et al.) [5]

⁴parametre μ, σ, ξ sú označené rovnako pre POT a GEV z toho dôvodu, že ich charakter je podobný (location, scale, shape), ale treba si uvedomiť, že sú to iné parametre.

Pre zaujímavosť na Obrázku 6 a 7 sú znázornené hustoty GPD pre rôzne parametre ξ a $\tilde{\sigma} = 1$. Je vidno ako sa mení tvar chvostu, ak sa mení ξ .



Obrázok 6: *Hustota GPD pre rôzne parametre ξ a $\tilde{\sigma} = 1$*



Obrázok 7: *Hustota GPD pre rôzne parametre $\xi < 0$ a $\tilde{\sigma} = 1$, na intervale $[0, -1/\xi]$*

Poznámka 2.1: Ak zmeníme dĺžku blokov, tak sa zmenia parametre GEV triedy (dôvodom je tvar ML odhadov), ale parametre Pareto triedy zostávajú nemenné vzhľadom na n . To znamená, že ξ nezávisí na n . ξ je dôležitým parametrom.

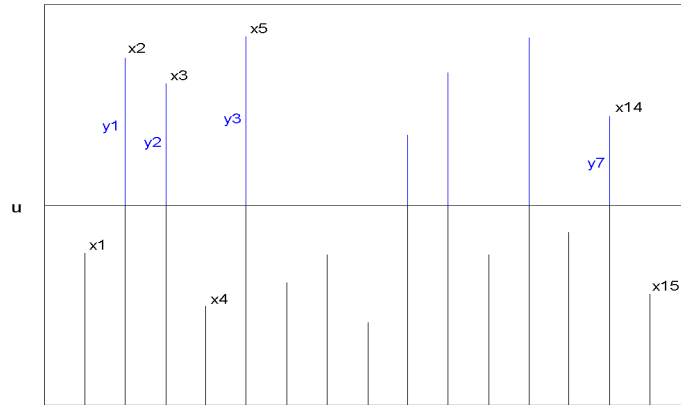
- ak $\xi < 0 \rightarrow$ existuje horné ohraničenie rozdelenia prekročení, $(u - \tilde{\sigma}/\xi)$
- ak $\xi > 0 \rightarrow$ neexistuje horné ohraničenie,
- ak $\xi = 0 \rightarrow$ rozdelenie je neohraničené.

$$\lim_{\xi \rightarrow 0} \left[1 - \left(1 + \frac{\xi y}{\tilde{\sigma}} \right)^{-1/\xi} \right] = 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right) = H_{GPD}(y), \quad y > 0,$$

čo zodpovedá exponenciálnemu rozdeleniu s parametrom $1/\tilde{\sigma}$.

2.3 Metódy na stanovenie prahovej hodnoty

Majme x_1, \dots, x_n nezávislé rovnako rozdelené náhodné premenné. Definujme prahovú hodnotu u a vyberme tie x_i , ktoré sú väčšie ako u t.j. $\{x_i : x_i > u\}$. Nech prekročenia sú $y_i = x_j - u$, $j = 1, \dots, k$ (ako napr. na Obrázku 8). Podľa Vety 2.1. hodnoty y_i by mali byť nezávislé realizácie náhodnej premennej, ich rozdelenie môžeme aproximovať pomocou triedy zovšeobecneného Pareto rozdelenia.



Obrázok 8: Dáta x_1, \dots, x_{15} a zodpovedajúce prekročenia y_1, \dots, y_7 ponad prah u .

Správna voľba hranice u je podobný problém ako stanovenie veľkosti bloku n v modeloch blokového maxima. Opäť problém rovnováhy medzi variáciou a vychýlenosťou odhadov:

- ak hranica u je nízka, tak aproximácia je menej kvalitná,
- ak hranica u je vysoká, potom variácia odhadu je veľká.

Používame dve metódy na stanovenie hranice u :

1. výskumná metóda (pred odhadovaním),
2. hodnotenie stability odhadov parametrov (fitovanie modelu cez rôzne u), ktoré sú podrobnejšie popísané nižšie.

Poznámka 2.2: Existujú aj iné metódy, napríklad 'bootstrapping MSE' metóda alebo špeciálna metóda založená na 'hill plot'-e, ktorá stanovuje hranicu u na základe nesubjektívneho posúdenia (viď článok Zhou [9]).

2.3.1 Výskumná metóda

Výskumná metóda je založená na strednej hodnote zovšeobecneného Pareto rozdelenia. Ak Y má GPD rozdelenie s parametrami σ a ξ , potom strednú hodnotu môžeme vyjadriť nasledovne:

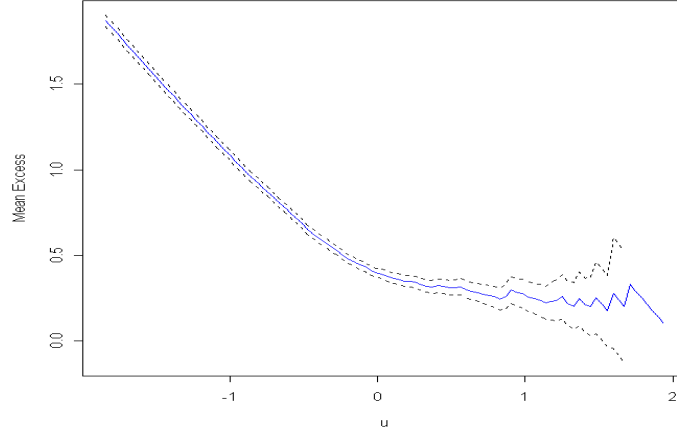
$$E(Y) = \begin{cases} \frac{\sigma}{1-\xi} & \text{pre } \xi < 1 \\ \infty & \text{pre } \xi \geq 1 \end{cases}$$

$$E(X - u | X > u) = \frac{\sigma_u}{1-\xi}, \quad \xi < 1. \quad (12)$$

Hranica u sa stanoví na základe grafu **MRL**, ktorý pozostáva z bodov:

$$\left\{ \left(u, \underbrace{\frac{1}{n_u} \sum_{i=1}^{n_u} (x_i - u)}_{\Delta} \right) : u < x_{max} \right\} \quad (13)$$

Δ = empirický odhad strednej hodnoty prekročení, n_u – sú pozorovania ktoré prekročia hranicu u a x_{max} – je najväčšia hodnota x_i . Tento graf nám stanovuje u na základe jej linearity. Hranica u sa stanoví na úrovni, kde táto vlastnosť nie je splnená. Ak už máme hranicu u , potom parametre GPD sa dajú odhadnúť ML metódou. Napríklad na Obrázku 9 vidíme, že najvhodnejšia voľba hodnoty u je okolo 1, lebo po tejto hodnote sa vyskytujú výraznejšie výkyvy.



Obrázok 9: Ukážka grafu MRL

2.3.2 Hodnotenie stability odhadov parametrov

Druhá metóda je založená na skúmaní stability odhadov parametrov. Fitujeme GPD pre rôzne hranice a skúmame stability odhadov parametrov.

Ak GPD je rozumný model pre prekročenia ponad prah u_0 , potom prekročenia nad $u > u_0$ tiež môžeme modelovať pomocou GPD. Parameter ξ je identický v oboch prípadoch a pre σ platí:

$$\sigma_u = \sigma_{u_0} + \xi(u - u_0), \quad (14)$$

σ_u – je σ pre prah u , σ_{u_0} – pre prah u_0 .

Reparametrizáciou $\sigma^* = \sigma_u - \xi u$ dosiahneme, že σ^* je konštantné na základe (14). Odhady σ^* a ξ budú konštantné nad u_0 v prípade, že je správne nastavená prahová hodnota vzhľadom na to, že prekročenia budú sledovať zovšeobecnené Pareto rozdelenie. V praxi nebudú parametre konštantné ale budú stabilné. Hodnoty $\hat{\sigma}^*$ a $\hat{\xi}$ znázorníme oproti u a pridáme interval spoľahlivosti. Zvolíme takú hodnotu u_0 , teda najnižšiu hodnotu u , pre ktoré sú takmer konštantné tieto parametre.

2.4 ML odhad parametrov GPD rozdelenia

Tvrdenie 2.1. (Coles [1]) Nech y_1, \dots, y_k sú prekročenia s rozdelením GPD, potom 'log-likelihood' má nasledovný tvar:

$\xi = 0$:

$$l(\sigma) = -k \log \sigma - \sigma^{-1} \sum_{i=1}^k y_i, \quad (15)$$

$\xi \neq 0$:

$$l(\sigma, \xi) = -k \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^k \log\left(1 + \frac{\xi y_i}{\sigma}\right) \quad (16)$$

za predpokladu, že $\left(\frac{1 + \xi y_i}{\sigma}\right) > 0$ pre $i = 1, \dots, k$.

Maximalizáciou (15) alebo (16) dostaneme ML odhad parametrov GPD rozdelenia. Opäť používame numerické optimalizačné metódy. Interpretácia je praktickejšia v tvare kvantilov alebo 'return level'.

Ak vieme že pre $X > u$:

$$P(X > x | X > u) = \left[1 + \xi \left(\frac{x - u}{\sigma}\right)\right]^{-1/\xi}$$

$$P(X > x) = \zeta_u \left[1 + \xi \left(\frac{x - u}{\sigma}\right)\right]^{-1/\xi},$$

kde $\zeta_u = P\{X > u\}$, potom úroveň x_m , ktorá je prekročená v priemere jedenkrát každých m pozorovaní (m je dostatočne veľké) je:

pre $\xi = 0$:

$$x_m = u + \sigma \log(m\zeta_u) \quad (17)$$

pre $\xi \neq 0$:

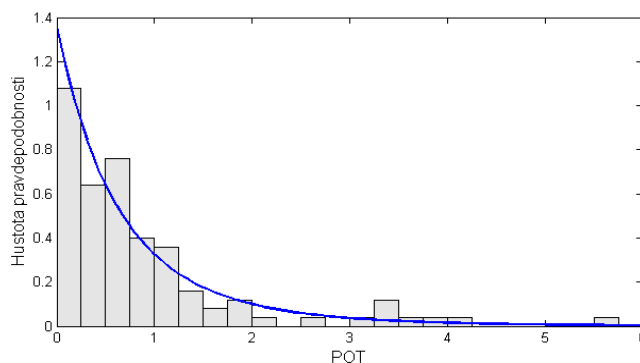
$$x_m = u + \frac{\sigma}{\xi} [(m\zeta_u)^\xi - 1]. \quad (18)$$

Ak hodnoty x_m znázorníme oproti m v logaritmickej miere, dostaneme '**return level plot**'. Graf je lineárny ak $\xi = 0$, konvexný ak $\xi < 0$ a konkávny ak $\xi > 0$. Samozrejme, kvantil x_m definovaný vo výrazoch (17) a (18) tiež môžeme chápať ako VaR (viď Poznámku 1.5).

2.5 Overovanie modelu

Na overovanie opäť môžeme používať grafickú analýzu dát. Jednotlivé grafy sa odvádzajú podobným spôsobom ako v časti 1.5, aj závery sa dajú podobným spôsobom interpretovať.

- **Histogram, graf hustoty:**



Obrázok 10: Príklad histogramu GPD rozdelenia

- **P-P plot:**

Nech prahová hodnota je u , prekročenia sú $y_1 \leq y_2 \leq \dots \leq y_k$ a odhadovaný model je \hat{H}_{GPD} , potom P-P graf pozostáva z bodov:

$$\xi \neq 0 : \left\{ \left(\frac{i}{k+1}, \hat{H}_{GPD}(y_i) \right); i = 1, \dots, k \right\} \quad (19)$$

kde

$$\hat{H}_{GPD}(y) = 1 - \left(1 + \frac{\hat{\xi}y}{\hat{\sigma}} \right)^{-1/\hat{\xi}}.$$

- **Q-Q plot:**

pozostáva z bodov:

$$\xi \neq 0 : \left\{ \left(\hat{H}_{GPD}^{-1} \left(\frac{i}{k+1} \right), y_i \right); i = 1, \dots, k \right\} \quad (20)$$

kde

$$\hat{H}_{GPD}^{-1}(y) = u + \frac{\hat{\sigma}}{\hat{\xi}} [y^{-\hat{\xi}} - 1].$$

- **Return level plot:**

pozostáva z bodov (pre dostatočne veľké m):

$$\{(m, \hat{x}_m)\} \quad (21)$$

kde \hat{x}_m je odhadovaná úroveň, ktorá je prekročená v priemere jeden krát každých m pozorovaní ('m-observation return level'):

$$\hat{x}_m = u + \frac{\hat{\sigma}}{\hat{\xi}} [(m\hat{\zeta}_u)^{\hat{\xi}} - 1].$$

3 Extrémy závislých časových radov

Modely extrémnych hodnôt v predchádzajúcich kapitolách boli odvodené pre matematický proces, ktorý pozostáva z radu nezávislých náhodných premenných. Nezávislosť je ale dosť nereálny predpoklad v praxi. Najprirodzenejšie zovšeobecnenie radu je stacionárny rad, ktorého premenné možno vzájomne závisia, ale ich stochastické vlastnosti sú homogénne vzhľadom na čas. Napríklad, ak X_1, X_2, \dots je stacionárny rad, potom X_1 musí mať rovnaké rozdelenie ako X_{101} , a združené rozdelenie (X_1, X_2) je rovnaké ako (X_{101}, X_{102}) , hoci X_1 nemusí byť závislí od X_2 a X_{102} .

3.1 Extrémy stacionárnych postupností

3.1.1 Maximá stacionárnych postupností

Stacionárny časový rad musí spĺňať predpoklad $D(u_n)$, ktorý hovorí o dostatočnej "časovej" vzdialenosti extrémnych udalostí, aby limitné pravidlá zostali zachované.

Definícia 3.1. (Podmienka $D(u_n)$) (Coles [1])

Hovoríme, že stacionárny rad X_1, X_2, \dots spĺňa podmienku $D(u_n)$ ak pre každé $i_1 < \dots < i_p < j_1 < \dots < j_q$ pričom $j_1 - i_p > l$ platí:

$$\begin{aligned} & |P\{X_{i_1} \leq u_n, \dots, X_{i_p} \leq u_n, X_{j_1} \leq u_n, \dots, X_{j_q} \leq u_n\} \\ & - P\{X_{i_1} \leq u_n, \dots, X_{i_p} \leq u_n\}P\{X_{j_1} \leq u_n, \dots, X_{j_q} \leq u_n\}| \leq \alpha(n, l), \end{aligned}$$

kde $\alpha(n, l_n)$ konverguje k nule pre ľubovoľnú postupnosť l_n takú, že: $\lim_{n \rightarrow \infty} l_n/n = 0$.

Poznámka 3.1: Pre postupnosti nezávislých premenných rozdiel pravdepodobností definovaný v Definičie 3.1. je nula pre každú postupnosť u_n . Podrobnejšia charakteristika podmienky $D(u_n)$ sa nachádza v knihe Leadbetter et al. [5] alebo Leadbetter a Rootzén [6].

Ak je splnená podmienka $D(u_n)$, potom platí nasledujúca veta:

Veta 3.1. (Coles [1]) Nech X_1, X_2, \dots je stacionárny proces a definujme $M_n = \max\{X_1, \dots, X_n\}$. Ak existujú postupnosti konštant $\{a_n > 0\}$ a $\{b_n\}$ také, že:

$$P\{(M_n - b_n)/a_n \leq z\} \rightarrow G_{GEV}(z),$$

kde G_{GEV} je nedegenerovaná distribučná funkcia, a $u_n = a_n z + b_n$ spĺňa podmienku $D(u_n)$ pre každé $z \in \mathbb{R}$, potom G_{GEV} je člen GEV triedy.⁵

⁵Veta 3.1. sa dokazuje analogickým spôsobom ako Veta 1.1.

Nech X_1, X_2, \dots je stacionárny rad s marginálnou distribučnou funkciou F , a definujme príslušný rad X_1^*, X_2^*, \dots nezávislých premenných, pričom každé X_i^* má distribučnú funkciu F . Ak porovnáme limitné rozdelenie postupnosti $M_n = \max\{X_1, \dots, X_n\}$ a $M_n^* = \max\{X_1^*, \dots, X_n^*\}$ pre $n \rightarrow \infty$, potom marginálne rozdelenia radov X_i a X_i^* sú rovnaké.

Za istých konvergenčných vlastností sa dá nájsť prepojenie medzi rozdeleniami nezávislých a stacionárnych radov s rovnakou marginálnou distribučnou funkciou.

Veta 3.2. (Coles [1]) *Nech X_1, X_2, \dots je stacionárny proces a X_1^*, X_2^*, \dots je postupnosť nezávislých premenných s rovnakým marginálnym rozdelením. Definujme $M_n = \max\{X_1, \dots, X_n\}$ a $M_n^* = \max\{X_1^*, \dots, X_n^*\}$. Pod podmienkou regularity platí, že:*

$$P\{(M_n^* - b_n)/a_n \leq z\} \rightarrow G_1(z)$$

ak $n \rightarrow \infty$ pre normalizované postupnosti $\{a_n > 0\}$ a $\{b_n\}$, kde G_1 je nede-generovaná distribučná funkcia, práve vtedy keď

$$P\{(M_n - b_n)/a_n \leq z\} \rightarrow G_2(z)$$

kde

$$G_2(z) = G_1^\theta(z) \tag{22}$$

pre konštantu $\theta : 0 < \theta \leq 1^1$.

Z Vety 3.2. vyplýva, že ak maximá stacionárneho radu konvergujú (čo platí podľa vety 3.1. za predpokladu, že je splnená podmienka $D(u_n)$) limitné rozdelenie je spojené s limitným rozdelením nezávislého radu, pričom platí rovnosť (22).

Ak G_1 zodpovedá GEV rozdeleniu s parametrami (μ, σ, ξ) , a $\xi \neq 0$, potom

$$\begin{aligned} G_1^\theta(z) &= \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}\right\}^\theta \\ &= \exp\left\{-\theta\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}\right\} \\ &= \exp\left\{-\left[1 + \xi\left(\frac{z - \mu^*}{\sigma^*}\right)\right]^{-1/\xi}\right\} \end{aligned}$$

kde
$$\mu^* = \mu - \frac{\sigma}{\xi}(1 - \theta^{-\xi}) \quad \text{a} \quad \sigma^* = \sigma\theta^\xi.$$

¹Dôkaz Vety 3.2. Leadbetter et al. [5]

Poznámka 3.2: Teda ak M_n^* má rozdelenie GEV s parametrami (μ, σ, ξ) , tak M_n má rozdelenie GEV s parametrami (μ^*, σ^*, ξ) . Parameter ξ je v oboch prípadoch rovnaký. Napríklad, ak G_1 je Gumbelovo rozdelenie s parametrami (μ, σ) , potom G_2 je Gumbelovo rozdelenie s parametrami $\mu^* = \mu + \sigma \log \theta$ a $\sigma^* = \sigma$. Veličina θ definovaná výrazom (22) sa nazýva '**extremal index**'⁶. Parameter θ môžeme interpretovať aj nasledovným spôsobom:

$$\theta = (\text{limitná}^7 \text{ stredná miera veľkosti 'cluster'-ov})^{-1}.$$

3.1.2 Modelovanie stacionárnych radov

Vety 3.1. a 3.2. sú ekvivalentné s asymptotickými vlastnosťami pre maximá získané v predchádzajúcich kapitolách. Ale aký to má efekt v modelovaní extrémnych hodnôt v skutočnosti?

Modely blokového maxima

Pokiaľ dlhodobá závislosť v extrémnych hodnotách je slabá, dáta môžeme považovať za realizáciu procesu spĺňajúcu podmienku $D(u_n)$, potom rozdelenie blokových maxím patrí do tej istej triedy rozdelení, ako pri nezávislých postupnostiach. Jeho parametre sú odlišné od tých, ktoré by sme získali v prípade nezávislého radu, ale z nášho pohľadu je to nepodstatné, lebo parametre sú odhadované rovnakým spôsobom. Z toho vyplýva, že môžeme ignorovať závislosť v daných dátach, a používať model blokového maxima.

Modely POT

Rozdelenie prekročení prahovej hodnoty v stacionárnych radoch sa dá aproximovať pomocou GPD. V stacionárnych radoch extrémny môžu mať tendenciu zoskupiť sa (clustering). Asymptotické vlastnosti naznačujú, že marginálne rozdelenie prekročení sleduje GPD, ale nevedú k špecifikácii združeného rozdelenia susedných prekročení. Teda rozdelenie hociktorého prekročenia sa dá modelovať pomocou GPD, ale zoskupenie sa extrémov indikuje závislosť v pozorovaniach. V takých prípadoch sa používa metóda '**declustering**'⁸, čo je vlastne filtrovanie závislých pozorovaní na získanie nezávislých prekročení prahovej hodnoty. Metóda funguje v hrubých rysoch nasledovným spôsobom: – používa sa empirické pravidlo na definovanie 'cluster'-ov prekročení

⁶Presnejšia interpretácia a ďalšie podmienky regularity sú uvedené v knihe Leadbetter et al. [5]

⁷limitná v zmysle extrémne vysokej prahovej hodnoty

⁸Základná idea a ďalšie detaily v článku Davison a Smith [8]

- identifikujú sa maximálne prekročenia vo vnútri každého 'cluster'-a
- predpokladá sa, že tieto maximá sú nezávislé a rozdelenie prekročení je dané pomocou GPD
- fituje sa GPD pre maximá 'cluster'-ov.

3.2 Extrémy nestacionárnych postupností

Všeobecne platný model pre nestacionárne rady neexistuje. Existujú modely pre špeciálne typy nestacionarity, ale v praxi sú nepoužiteľné. Namiesto toho používame štandardné modely extrémnych hodnôt, ktoré sa rozširujú o štatistické modelovanie (je to len reprezentatívny príklad takýchto metód). Maximálne hodnoty Z_t modelujeme pomocou GEV rozdelenia s parametrami μ , σ a ξ , označme to nasledovne:

$$Z_t \sim GEV(\mu(t), \sigma, \xi). \quad (23)$$

Zmeny parametra polohy $\mu(t)$ môžeme modelovať napríklad ako lineárny trend s parametrami β_0 a β_1 :

$$\mu(t) = \beta_0 + \beta_1 t. \quad (24)$$

Pri zložitejších zmenách v parametre μ používame napríklad kvadratický model:

$$\mu(t) = \beta_0 + \beta_1 t + \beta_2 t^2, \quad (25)$$

alebo exponenciálny model:

$$\mu(t) = \exp(\beta_0 + \beta_1(t)) \quad (26)$$

Nestacionarita sa môže prejavíť takisto aj v ostatných parametroch.

3.2.1 Odhad parametrov

Výhoda metódy maximálnej vierohodnosti oproti inými metódami odhadovania parametrov je prispôsobivosť k zmene v štruktúre modelu. Nestacionárny GEV model charakterizujúci rozdelenie blokových maxím Z_t pre $t = 1, \dots, m$:

$$Z_t \sim GEV(\mu(t), \sigma(t), \xi(t)), \quad (27)$$

kde $\mu(t)$, $\sigma(t)$ a $\xi(t)$ majú rozdelenie definované strednou hodnotou a varianciou. Označme vektor parametrov ako β , potom vierohodnostná funkcia je nasledovná:

$$L(\beta) = \prod_{t=1}^m g(z_t; \mu(t), \sigma(t), \xi(t)),$$

kde $g(z_t; \mu(t), \sigma(t), \xi(t))$ je funkcia hustoty GEV rozdelenia s parametrami $\mu(t), \sigma(t)$ a $\xi(t)$. Analogicky ako v časti 1.3. 'log-likelihood' funkcia má tvar:

$$l(\mu, \sigma, \xi) = - \sum_{t=1}^m \left\{ \log \sigma(t) + (1 + 1/\xi(t)) \log \left[1 + \xi(t) \left(\frac{z_t - \mu(t)}{\sigma(t)} \right) \right] + \left[1 + \xi(t) \left(\frac{z_t - \mu(t)}{\sigma(t)} \right) \right]^{-1/\xi(t)} \right\} \quad (28)$$

za predpokladu, že $1 + \xi(t) \left(\frac{z_t - \mu(t)}{\sigma(t)} \right) > 0$ pre $t = 1, \dots, m$.

Na maximalizáciu (28) sa používajú numerické metódy.

4 Ciele práce

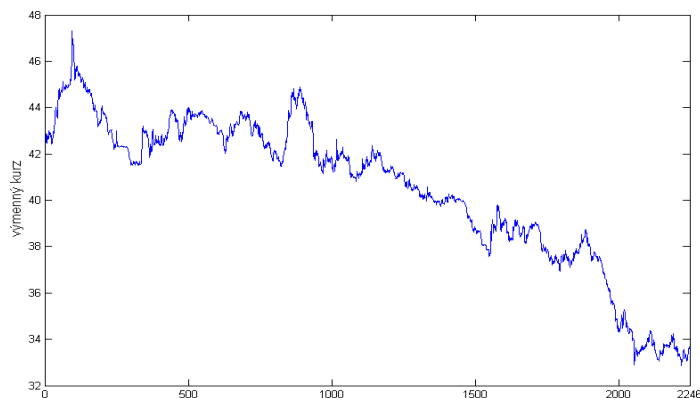
V predchádzajúcich kapitolách sme vysvetlili klasickú teóriu extrémnych hodnôt. Analyzovali sme dva základné modely (model blokového maxima a model s prahovou hodnotou), zaviedli sme zovšeobecné rozdelenie extrémnych hodnôt (2) a zovšeobecnené Pareto rozdelenie (11). Uvedené rozdelenia sú vhodné pri registrácii extrémov, a zjednodušujú štatistickú analýzu extrémnych údajov.

Diplomová práca sa zaoberá vysvetlením a krátkym zhrnutím EVT cieľom aplikácie tejto teórie na slovenské dáta a následnej interpretácie dosiahnutých výsledkov. Máme k dispozícii historické dáta výmenného kurzu SKK/EUR od 1.1.1999 do 31.12.2007, na ktoré sa pokúsime aplikovať modely blokového maxima a modely prekročení ponad prah. Po navrhnutí typu rozdelenia extrémov odhadujeme základné parametre vyššie spomínaných modelov μ , σ a ξ , ktoré charakterizujú polohu, rozsah a tvar rozdelenia. Na základe grafických a iných základných metód otestujeme správnosť modelov (P-P graf, Q-Q graf, 'return level plot', graf hustoty). Zároveň sa pokúsime navrhnúť podľa dosiahnutých výsledkov stresový scenár výmenného kurzu.

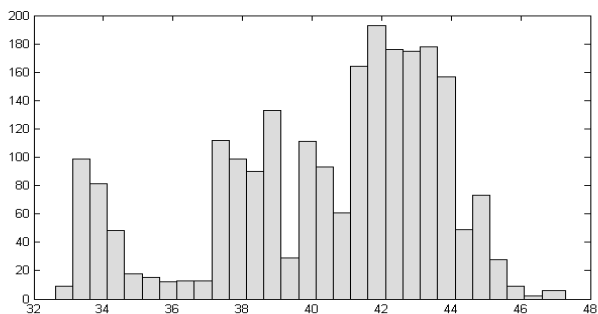
5 Praktická časť

5.1 Analýza dát

V tejto kapitole aplikujeme spomínané modely na reálne dáta. Naším cieľom je modelovať výkyvy denného výmenného kurzu slovenskej koruny voči euru. Máme k dispozícii denné historické dáta z obdobia 1.1.1999 až 31.12.2007⁹, pozostávajúci z 2246 údajov. Na týchto údajoch otestujeme modely zadané v kapitolách 3 a 4. Prvým krokom je grafické znázornenie údajov. Na Obrázku 11 môžeme odsledovať historický vývoj výmenného kurzu za periódu deviatich rokov, a na Obrázku 12 je znázornená frekvencia dát pomocou histogramu.



Obrázok 11: *Historický vývoj výmenného kurzu SKK/EUR v rokoch 99-07*



Obrázok 12: *Histogram dát*

⁹Dáta sme získali z oficiálnej webovej stránky Národnej banky Slovenska [11]

Opisné štatistické charakteristiky časového radu sú nasledovné:

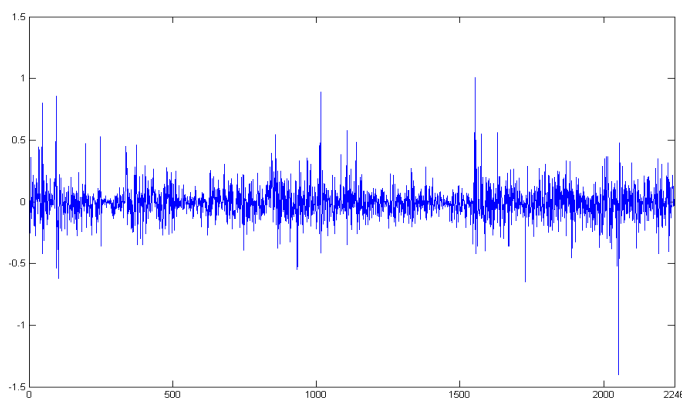
	Časový rad
Priemer	40.437
Variancia	10.579
Štandardná odchýlka	3.2525
Medián	41.388
Minimum	32.861
Maximum	47.299
1.kvantil	38.39
3.kvantil	42.9455

Tabuľka 1: Základné charakteristiky

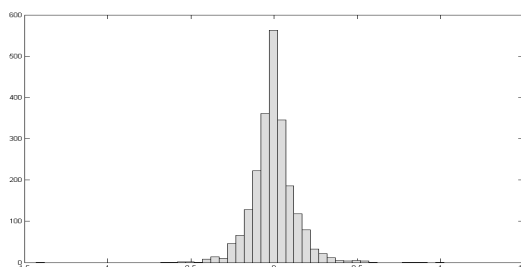
Grafický záznam dát ukazuje, že časový rad obsahuje klesajúci trend, ktorý eliminujeme transformáciou dát nasledovným spôsobom:

$$\tilde{x}_i = \log\left(\frac{x_i}{x_{i-1}}\right)100.$$

Pomocou transformácie sme dostali časový rad ktorý je vizualizované na Obrázku 13, a priložíme k tomu aj histogram, viď Obrázok 14. Je vidno, že veľké zmeny majú relatívne malú, hoci kladnú pravdepodobnosť vzniku. Budeme sa zaoberať analýzou týchto denných zmien nachádzajúcich sa na chvoste rozdelenia.



Obrázok 13: Denné zmeny hodnoty výmenného kurzu SKK/EUR v rokoch 99-07



Obrázok 14: *Histogram dát*

Z toho histogramu očividne vyplýva, že rozdelenie pravdepodobnosti má ťažké konce, teda dáta splňajú naše očakávania a môžeme aplikovať na modelovanie extrémov teóriu extrémnych hodnôt.

Opisné štatistické charakteristiky časového radu uvedené v Tabuľke 2 nám poskytujú obraz o súbore dát:

	Časový rad
Priemer	-0.00481162
Variancia	0.01824101
Štandardná odchýlka	0.13505929
Medián	-0.005585855
Minimum	-1.40495018
Maximum	1.00840655
1.kvantil	-0.06942574
3.kvantil	0.05613432

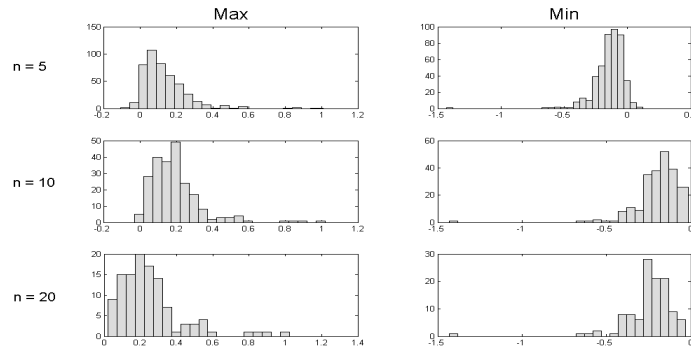
Tabuľka 2: Základné charakteristiky

5.2 Modely blokového maxima, resp. minima

V teórii sme formulovali modely blokového maxima (minima), ktoré chceme v tejto časti používať na opis a modelovanie extrémov našich dát. Zo zaujímavosti otestujeme tieto modely najprv na celom 9 ročnom intervale (t.j. od 1.1.99 do 31.12.07), potom zmeníme dĺžku časového radu na 4 roky (napríklad od 1.1.04 do 31.12.07), a na záver porovnáme kvalitu dosiahnutých výsledkov.

V prvom kroku vytvárame z dát celého 9 ročného intervalu bloky dĺžky n , a v rámci každého jedného bloku vyberieme maximum (minimum) a tieto extrémny modelujeme. Tento postup zopakujeme pre rôzne dĺžky blokov ($n = 5$, $n = 10$, $n = 20$) z toho dôvodu, aby sme zistili, ktorý spôsob

nám dáva pomerne rozumné výsledky. Teraz znázorníme rozdelenia extrémov jednotlivých blokov na zdôraznenie existencie tzv. ťažkých koncov (viď Obrázok 15).



Obrázok 15: *Histogramy modelov blokového maxima a blokového minima, pre rôzne veľkosti blokov.*

5.2.1 Kalibrácia modelov na intervale 9 rokov

Praktickú analýzu výsledkov uskutočníme využitím funkcií knižnice IS-MEV softvérovej aplikácie R 2.6.2 voľne dostupnej na internetovej stránke [13]. Po aplikácii modelov blokového maxima, resp. minima sme dostali výsledky, ktoré sú zhrnuté v Tabuľkách 3 a 4.

99-07	MAX		
dĺžka bloku	$n = 5$	$n = 10$	$n = 20$
počet dát	450	225	113
NLLH	-394.4002	-174.7725	-67.9387
MLE($\hat{\mu}$)	0.07562881	0.12576767	0.16725231
MLE($\hat{\sigma}$)	0.08055019	0.08699775	0.09985667
MLE($\hat{\xi}$)	0.11605080	0.15531538	0.20936871
SE($\hat{\mu}$)	0.004205897	0.006545053	0.010871509
SE($\hat{\sigma}$)	0.003160910	0.005051299	0.008694646
SE($\hat{\xi}$)	0.031239979	0.051228364	0.084435822

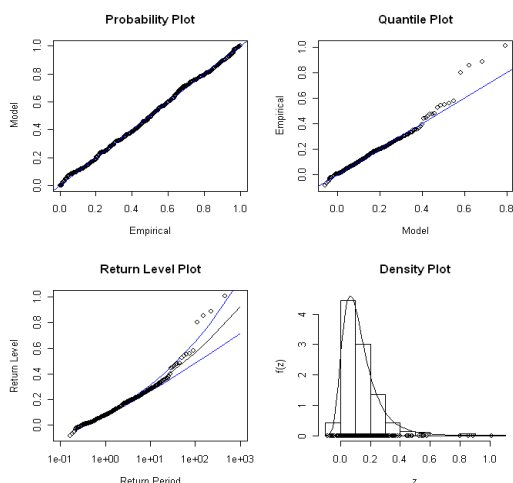
Tabuľka 3:

NLLH - 'negative log-likelihood',

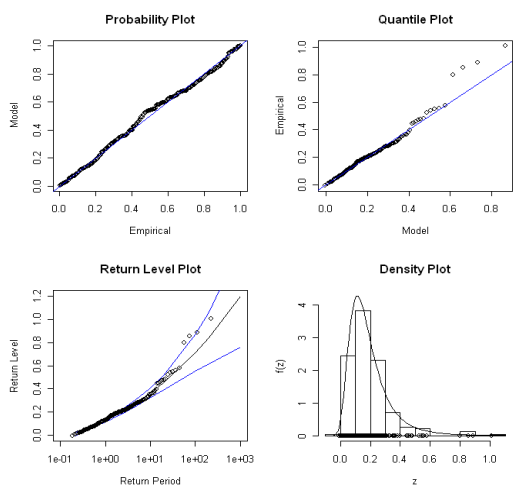
MLE - maximálne vierohodný odhad jednotlivých parametrov,

SE - štandardná chyba odhadu jednotlivých parametrov.

Ako sme už spomínali v časti 1.5, na overovanie týchto modelov používame grafickú analýzu, ktorá zahŕňa zostrojenie P-P grafu, Q-Q grafu, grafu hustoty a 'return level plot'. Na Obrázku 16, 17 a 18 môžeme odsledovať priebeh spomínaných grafov.

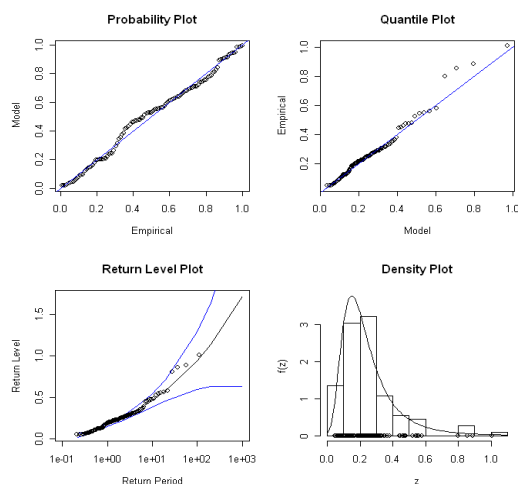


Obrázok 16: Model blokového maxima, $n = 5$



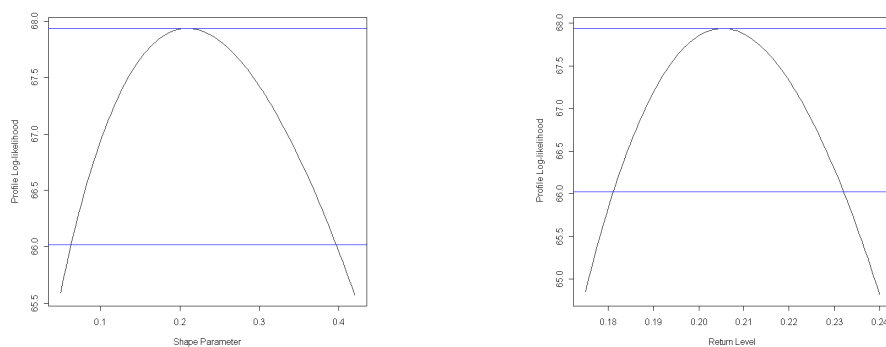
Obrázok 17: Model blokového maxima, $n = 10$

Ako už vieme z teórie, P-P graf porovnáva empirické a odhadnuté rozdelenie pravdepodobnosti, Q-Q graf kvantil empirického a odhadnutého rozdelenia. Model je dobre sformulovaný, keď body týchto grafov ležia približne pozdĺž diagonálnej priamky, ako to vidno aj v našom prípade. Keď



Obrázok 18: *Model blokového maxima, $n = 20$*

porovnáme 'return level plot'-y vidíme, že niektoré hodnoty na Obrázku 16 a 17 sa nachádzajú mimo intervalu spoľahlivosti, zatiaľ čo na obrázku 18 interval obsahuje všetky modelované hodnoty, čo aj spĺňa naše očakávania, t.j. ak bloky sú malé, tak aproximácia je menej kvalitná, naopak ak sú bloky veľké, variancia odhadu je veľká. Toto sa odzrkadľuje aj vo vyššie uvedených tabuľkách. Z toho nám vyplýva, že treba nájsť vhodnú rovnováhu medzi vychýlenosťou a varianciou odhadov. Na základe týchto poznatkov najlepší výber z testovaných dĺžok bloku je $n = 20$. Ukážku grafov profilu vierohodnosti pre ξ a 2-ročný 'return level' v prípade nášho zvoleného modelu môžeme vidieť na Obrázku 19. Je vidno, že intervaly spoľahlivosti nie sú symetrické.



Obrázok 19: *Profil vierohodnosti pre ξ (vľavo); profil vierohodnosti pre 2-ročný 'return level' (vpravo)*

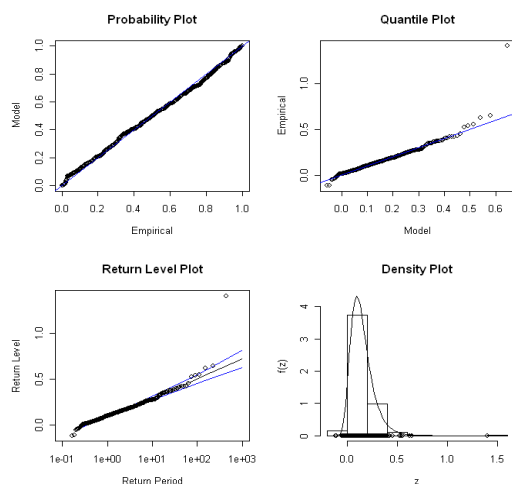
Profil vierohodnosti spomínaný v časti 1.4 nám dáva lepšiu aproximáciu rozdelenia ML odhadov parametrov. Napríklad pre parameter ξ interval spoľahlivosti vypočítaný pomocou normálneho rozdelenia je $[0.0439, 0.3749]$, ktorý je symetrický. Ako vidíme na Obrázku 19, profil vierohodnosti určí presnejší interval spoľahlivosti, ako normálne rozdelenie. Postup na výpočet intervalu poskytuje kniha [1], str. 34-35. Na základe profilu vierohodnosti môžeme určiť konfidenčné intervaly aj pre 2-ročný 'return level' ('return level' pre 2-ročný časový horizont), teda pre odhad kvantilu z_p , čo je zobrazené na Obrázku 19 (vpravo).

Pri modelovaní blokových miním postup je rovnaký len s tou výnimkou, že prístupné dáta vynásobíme -1 a potom pre tieto hodnoty fitujeme GEV rozdelenie pre maximá: $\min\{X_1, \dots, X_n\} = -\max\{-X_1, \dots, -X_n\}$. Ale v tomto prípade musíme brať do úvahy protikladné hodnoty odhadu parametra $\hat{\mu}$, lebo práve tento parameter nám vyjadruje polohu rozdelenia. Ostatné parametre zostanú nemenné. V Tabuľke 4 sú uvedené výsledky modelovania v prípade blokových miním.

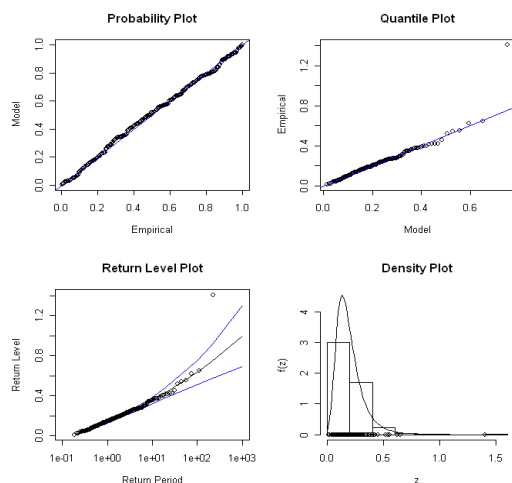
99-07	MIN		
dĺžka blokov	$n = 5$	$n = 10$	$n = 20$
počet dát	450	225	113
NLLH	-400.5228	-194.8525	-82.89461
MLE($\hat{\mu}$)	-0.09560398	-0.14076437	-0.18468728
MLE($\hat{\sigma}$)	0.08511040	0.08139155	0.09425815
MLE($\hat{\xi}$)	0.01831857	0.11419705	0.09414792
SE($\hat{\mu}$)	0.004331858	0.006078511	0.00979840
SE($\hat{\sigma}$)	0.003029563	0.004586885	0.00720672
SE($\hat{\xi}$)	0.020504813	0.046436863	0.05539810

Tabuľka 4:

Teraz sa pozrime na grafickú analýzu dát. Všimnime si, že nám vždy vyskočí jedna veľmi nízka hodnota (viď 'return level plot' a Q-Q plot), čo je dôsledkom toho, že 20.3.07 sa SKK posilnila voči EUR veľmi razantne v porovnaní s hodnotou predchádzajúceho dňa. Túto výraznú zmenu môžeme postrehnúť aj na Obrázku 13. Táto hodnota sa ťažko modeluje, lebo ako to vidíme na 'return level plot', hocikaký blok zvolíme, táto hodnota sa vždy nachádza mimo predikčného intervalu. Keby sa to modelovalo napríklad s jednoduchým normálnym rozdelením, tak by sa tento údaj z ďaleka nedalo modelovať, keď to nie je v intervale spoľahlivosti ani pri teórii extrémnych hodnôt.

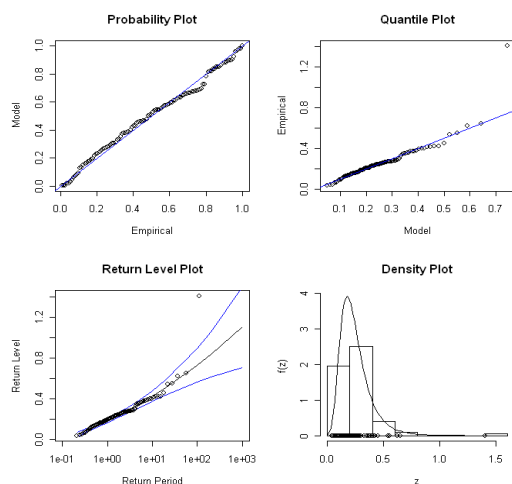


Obrázok 20: *Model blokového minima, $n = 5$*

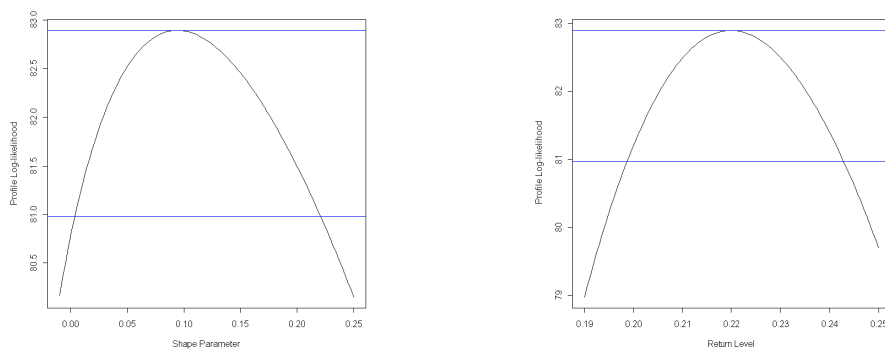


Obrázok 21: *Model blokového minima, $n = 10$*

Je vidno, že grafy P-P a Q-Q sú takmer lineárne, a graf hustoty zodpovedá rozdeleniu GEV. Pri modelovaní blokových miním tiež môžeme vyhlásiť, že najlepšia z testovaných dĺžok bloku je $n = 20$. Pre tento model profil vierohodnosti pre ξ a 2-ročný 'return level' je na Obrázku 23.



Obrázok 22: Model blokového minima, $n = 20$



Obrázok 23: Profil vierohodnosti pre ξ (vľavo); profil vierohodnosti pre 2-ročný 'return level' (vpravo)

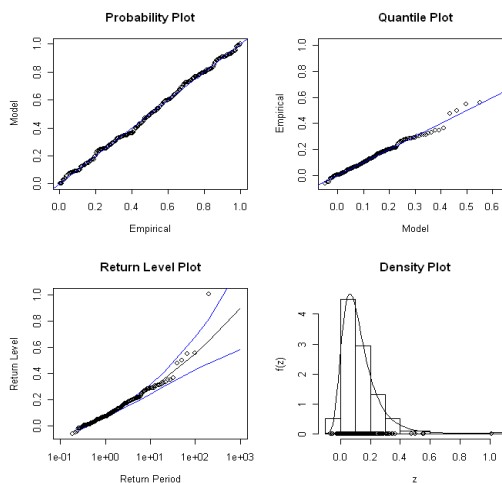
5.2.2 Kalibrácia modelov na intervale 4 rokov

V tejto časti overíme, ako vplýva výber dĺžky historického obdobia využitého pri modelovaní na výkonnosť modelu. 9 ročný interval skrátime na 4 roky, teda časový rad analyzovaného výmenného kurzu bude zahŕňať obdobie od 1.1.04 do 31.12.07, t.j. obsahuje 998 dát. Prakticky staršie dáta môžu byť menej relevantné pri analýze extrémnych zmien, keďže sa časový rad aj ekonomika stále mení. Z ekonomického hľadiska je lepšie uvažovať "čerstvé" dáta. Najprv modelujeme blokové maximá rovnakým spôsobom ako v predchádzajúcej časti. Dosiahnuté výsledky sú uvedené v Tabuľke 5.

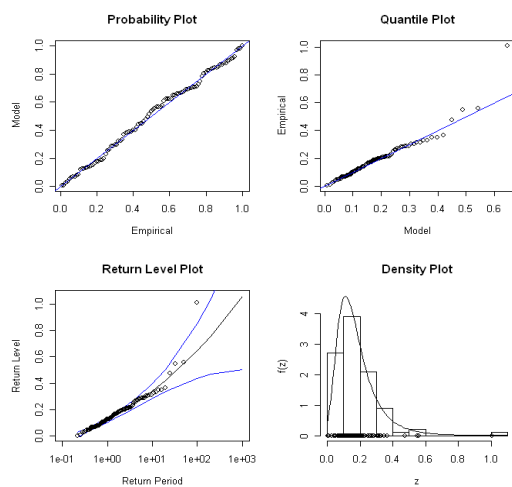
04-07	MAX		
dĺžka blokov	$n = 5$	$n = 10$	$n = 15$
počet dát	200	100	67
NLLH	-179.1489	-85.59715	-51.20026
MLE($\hat{\mu}$)	0.07268576	0.12175385	0.15364647
MLE($\hat{\sigma}$)	0.07903108	0.08124893	0.09175501
MLE($\hat{\xi}$)	0.11440240	0.13806700	0.09350040
SE($\hat{\mu}$)	0.006245538	0.009160395	0.012390569
SE($\hat{\sigma}$)	0.004717423	0.007021070	0.009141145
SE($\hat{\xi}$)	0.050247731	0.075424202	0.076209440

Tabuľka 5:

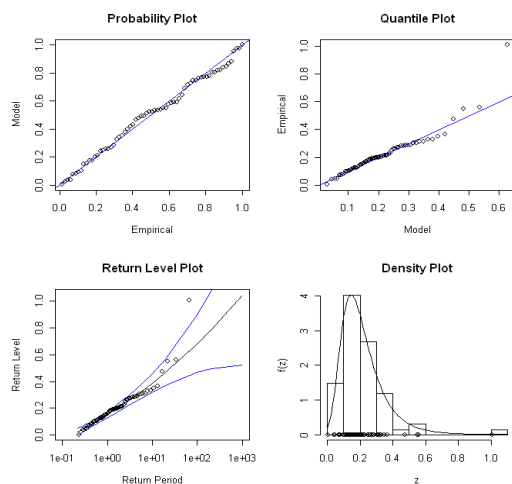
Pomocou nasledujúcich grafov môžeme overovať kalibrované modely. Aj v tomto prípade sú splnené predpoklady, že P-P, a Q-Q sú takmer lineárne a graf hustoty zodpovedá GEV rozdeleniu. Znovu sa objaví jedna extrémna hodnota (16.3.05), ktorá sa vždy nachádza mimo intervalu spoľahlivosti, a preto sa ťažko modeluje. Teraz porovnajme výsledky v prípade $n = 10$ pre 9 a 4 ročný interval. Z výsledkov môžeme vyvodit' záver, že chyba odhadu v prípade dlhšieho časového radu je menšia, ale aproximácia je menej kvalitná, čo sme aj očakávali. Po analýze môžeme konštatovať, že najlepšia voľba dĺžky bloku je $n = 15$.



Obrázok 24: Model blokového maxima, $n = 5$

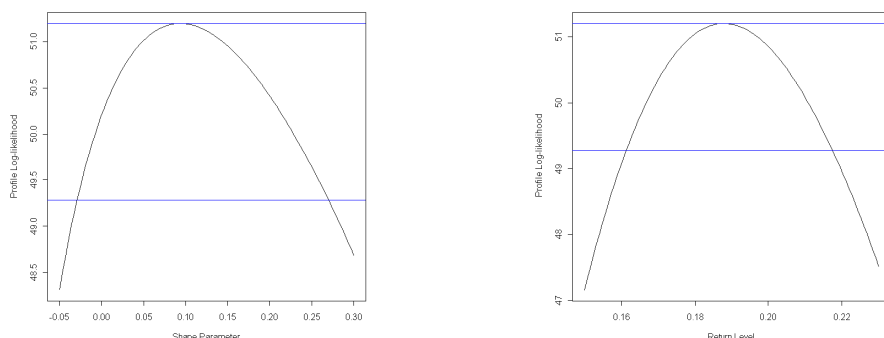


Obrázok 25: Model blokového maxima, $n = 10$



Obrázok 26: Model blokového maxima, $n = 15$

Nasledujúce grafy znázorňujú profil vierohodnosti pre ξ a 2-ročný 'return level', pre náš zvolený model ($n = 15$).

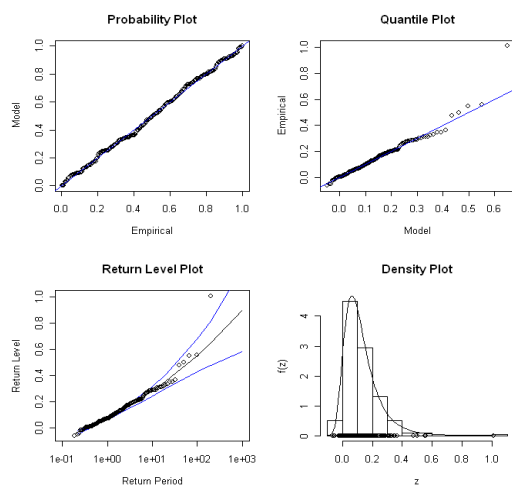


Obrázok 27: Profil vierohodnosti pre ξ (vľavo); profil vierohodnosti pre 2-ročný 'return level' (vpravo)

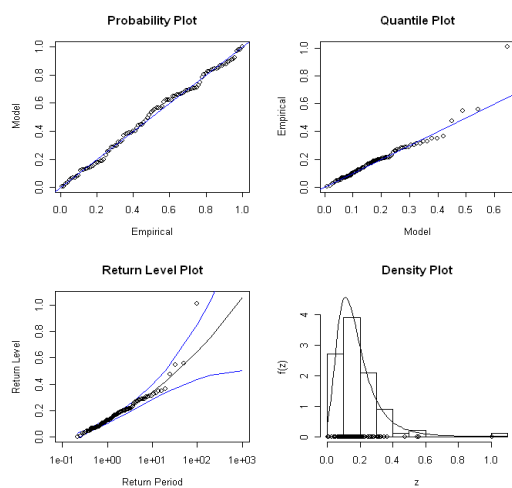
A teraz modelujeme blokové minimá na 4 ročnom časovom horizonte pre rôzne dĺžky blokov. Postup je taký istý, ako v časti 5.2.1. Opäť používame protikladné hodnoty, a fitujeme GEV rozdelenie pre maximá. Tento postup nám dáva výsledky zhrnuté v Tabuľke 6. Aj v tomto prípade je najlepšia voľba dĺžky bloku $n = 15$. Pridáme k tomu aj grafy na overovanie modelov (viď Obrázok 28, 29, 30), profil vierohodnosti pre ξ a 2-ročný 'return level' v prípade $n = 15$ (Obrázok 31).

04-07	MIN		
dĺžka blokov	$n = 5$	$n = 10$	$n = 15$
počet dát	200	100	67
NLLH	-192.0059	-84.49632	-54.11884
MLE($\hat{\mu}$)	-0.10115345	-0.14836363	-0.17307351
MLE($\hat{\sigma}$)	0.07290595	0.07935695	0.08488946
MLE($\hat{\xi}$)	0.14737372	0.19837104	0.15330380
SE($\hat{\mu}$)	0.005736601	0.008924237	0.011309389
SE($\hat{\sigma}$)	0.004383969	0.007037449	0.008503271
SE($\hat{\xi}$)	0.048471250	0.076035771	0.069149131

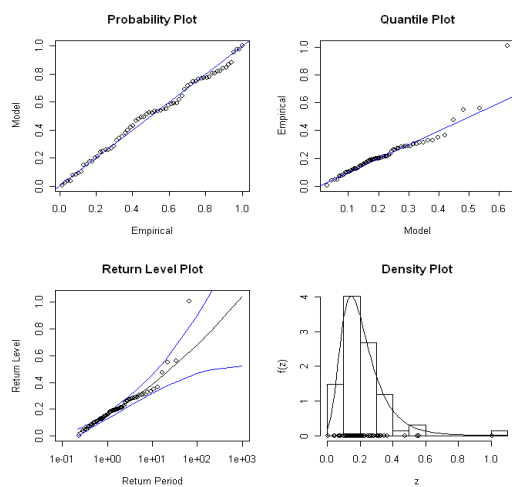
Tabuľka 6:



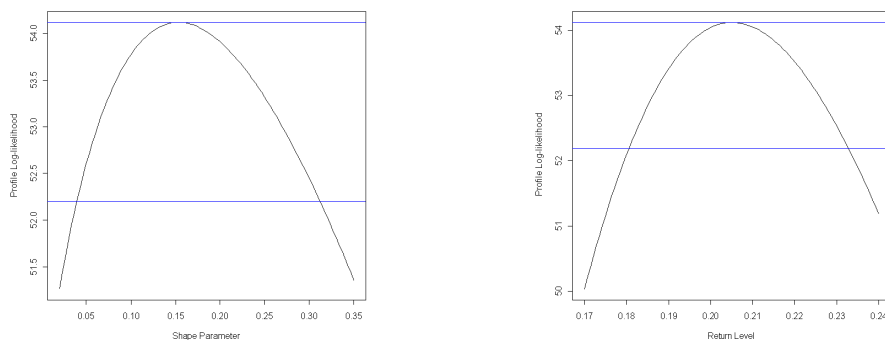
Obrázok 28: Model blokového minima, $n = 5$



Obrázok 29: Model blokového minima, $n = 10$



Obrázok 30: Model blokového minima, $n = 15$

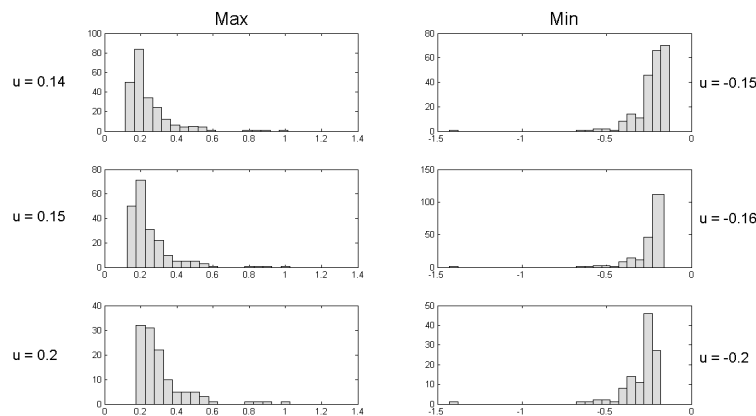


Obrázok 31: Profil vierohodnosti pre ξ (vľavo); profil vierohodnosti pre 2-ročný 'return level' (vpravo)

5.3 Modely POT

Druhý postup, ktorý nám teória extrémnych hodnôt poskytuje na modelovanie extrémov, je založený na opise správania sa tých pozorovaní, ktoré prekročia nejakú vopred stanovenú prahovú hodnotu. Metóda má tú výhodu, že upriamuje pozornosť nielen na najvyššie hodnoty v určitých časových intervaloch (ako v prípade metódy blokového maxima), ale na všetky hodnoty, ktoré presahujú vysoký prah. V nasledujúcej časti sa budeme zaoberať analyzovaním a modelovaním extrémnych hodnôt ponad prah.

Otestujeme modely najprv na celom 9 ročnom, potom na skrátenom 4 ročnom časovom horizonte. Rozdelenie prekročení prahovej hodnoty aproximujeme pomocou GPD rozdelenia (11). Cieľom je porovnať dosiahnuté výsledky pri rôznych prahových úrovniach. V praktických aplikáciách je veľmi dôležitá správna voľba prahu, na určenie optimálnej hodnoty používame metódy spomínané v časti 2.3, t.j. výskumnú metódu a hodnotenie stability odhadov parametrov. Na Obrázku 32 sú znázornené histogramy pre rôzne prahové hodnoty, pomocou ktorých môžeme overiť existenciu ťažkých koncov rozdelení.



Obrázok 32: Histogramy modelov pre rôzne prahové hodnoty

5.3.1 Kalibrácia modelov na intervale 9 rokov

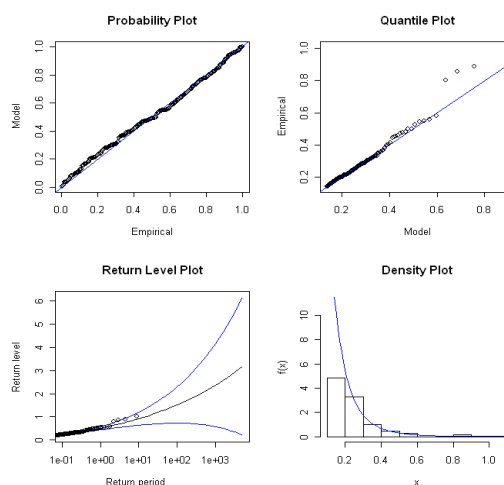
Cieľovo si zvolíme 3 rôzne prahové hodnoty pre maximá, a modelujeme extrémny reprezentované pozorovaniami, ktoré prekročia tieto hodnoty. Použitím spomínaných softvérových aplikácií dostaneme odhad jednotlivých parametrov a štandardné chyby týchto odhadov, ktoré sú zhrnuté v Tabuľke 7.

99-07	MAX		
prahová hodnota	$u = 0.14$	$u = 0.15$	$u = 0.2$
počet prekročení	228	207	118
NLLH	-292.422	-266.4049	-142.7113
MLE($\hat{\sigma}$)	0.08682783	0.08447302	0.08434427
MLE($\hat{\xi}$)	0.16113398	0.18441160	0.26334503
SE($\hat{\sigma}$)	0.008494383	0.008858939	0.01295048
SE($\hat{\xi}$)	0.072838657	0.079610955	0.12530436

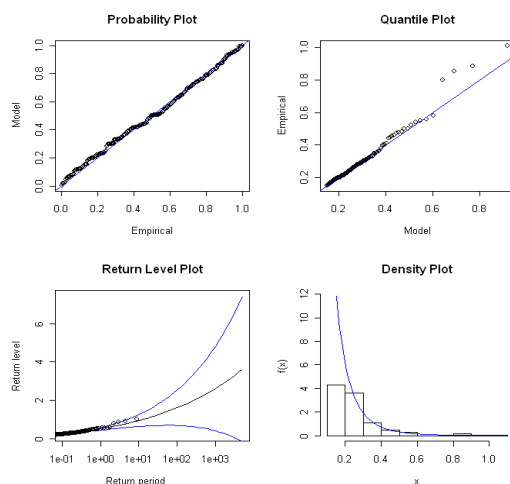
Tabuľka 7:

NLLH - 'negative log-likelihood',
MLE - maximálne vierohodný odhad jednotlivých parametrov,
SE - štandardná chyba odhadu jednotlivých parametrov.

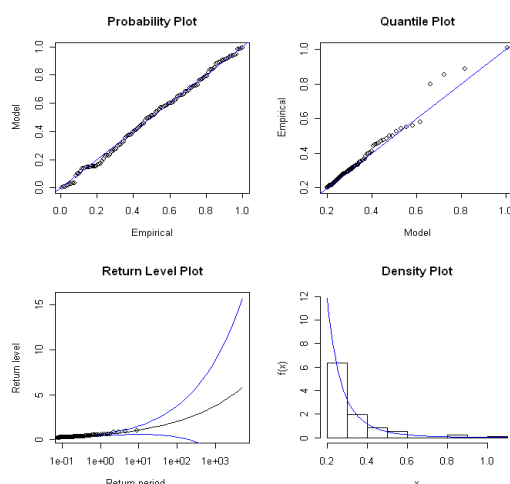
Aj v prípade metódy POT sa na overovanie kvality modelu používa grafická analýza dát, t.j. P-P plot, Q-Q plot, 'return level' plot a graf hustoty. Tieto grafy nám naznačujú, že zvolené modely sú vhodnými nástrojmi na kalibráciu extrémov. P-P a Q-Q sú primerane lineárne, a graf hustoty zodpovedá GPD rozdeleniu. Z porovnania 'return level plot'-ov vyplýva, že interval spoľahlivosti najlepšie mapuje extrémne hodnoty pri prahovej hodnote $u = 0.2$. Vychádzajúc z grafickej analýzy, predpokladáme, že vhodným modelom je GPD rozdelenie s prahovou hodnotou $u = 0.2$.



Obrázok 33: Model POT, $u = 0.14$

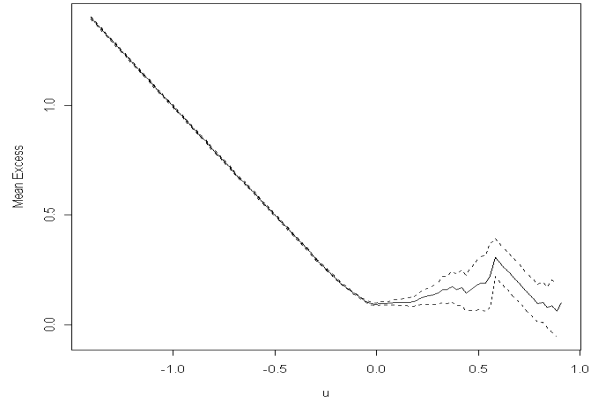


Obrázok 34: *Model POT*, $u = 0.15$



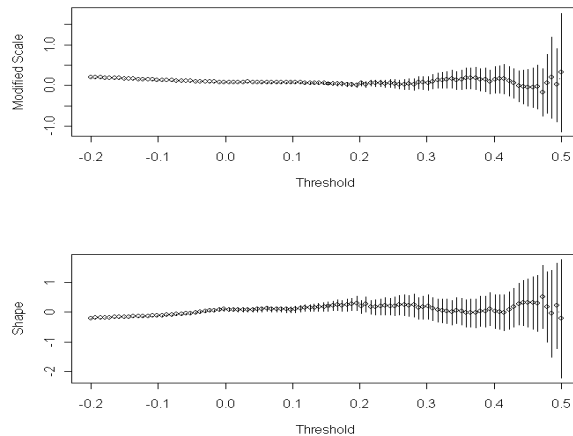
Obrázok 35: *Model POT*, $u = 0.2$

Teória nám hovorí, že prah u by mal byť dostatočne vysoký, avšak čím je vyššia táto hranica, tým nám zostáva menej údajov na modelovanie. Opäť sa vyskytuje problém medzi varianciou a vychýlenosťou odhadov. Na správnu voľbu hranice u používame metódy spomínané v časti 2.3. Pri výskumnej metóde hranica u sa stanoví na základe grafu 'mean residual life' (MRL). Na Obrázku 36 môžeme vidieť priebeh tohto grafu pre naše údaje.



Obrázok 36: MRL plot pre výmenný kurz SKK/EUR (99-07), 9 ročný časový rad

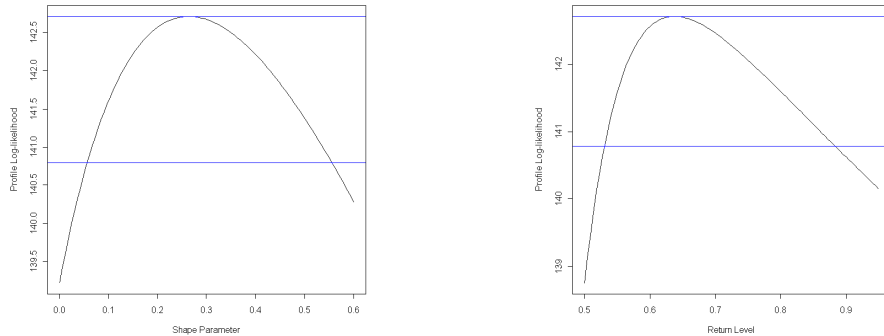
MRL graf nám stanovuje u na základe jej linearity. Hranica u sa stanoví na úrovni, kde táto vlastnosť nie je splnená. V našom prípade graf je približne lineárny do hodnoty $u = 0.2$. Efektívnosť prahovej hodnoty môžeme posúdiť detailnejšie pomocou hodnotenia stability odhadov parametrov. Výsledok je zobrazený na Obrázku 37, ktorý znázorňuje maximálne vierohodné odhady a konfidénčné intervaly pre zovšeobecnený Pareto model na škále rôznych prahových hodnôt (t.j. fitujeme model cez rôzne u).



Obrázok 37: ML odhady ako funkcie prahu

Na základe vyššie uvedených grafov sa zdá byť prijateľné stanoviť prah na úrovni $u = 0.2$. Podľa prvého grafu Obrázku 37 hodnoty na intervale $[0.2, 0.3]$ sú dostatočne stabilné, zatiaľ čo na druhom grafe interval stabil-

ných hodnôt je $[0.1, 0.2]$, a práve prienik týchto intervalov nám určí hľadaný prah. Profil vierohodnosti pre ξ a 2-ročný 'return level' pre vybraný model zobrazuje Obrázok 38.



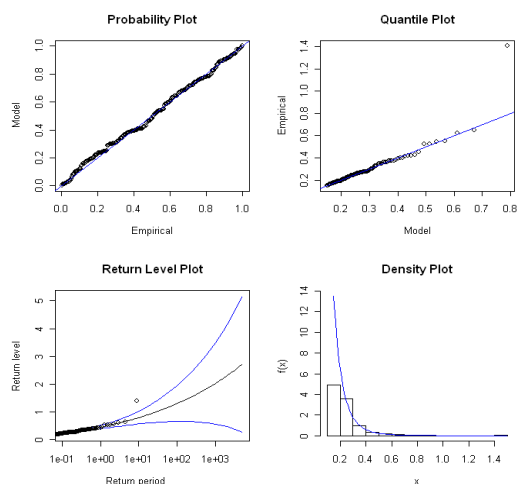
Obrázok 38: Profil vierohodnosti pre ξ (vľavo); profil vierohodnosti pre 2-ročný 'return level' (vpravo)

Pri modelovaní miním môžeme postupovať podobným spôsobom, tak že prístupné dáta vynásobíme -1 , a potom pre tieto hodnoty fitujeme GPD rozdelenie pre maximá. Dosiahnuté výsledky sú zhrnuté v Tabuľke 8.

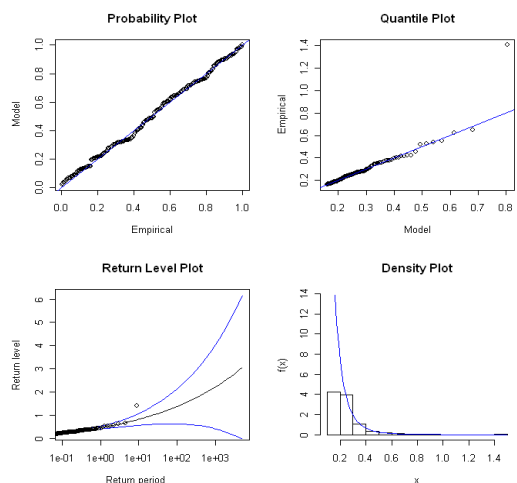
99-07	MIN		
prahová hodnota	$u = 0.15$	$u = 0.16$	$u = 0.2$
počet prekročení	223	199	114
NLLH	-321.7456	-287.7788	-150.2838
MLE($\hat{\sigma}$)	0.07400185	0.07213727	0.08331326
MLE($\hat{\xi}$)	0.16074768	0.18273687	0.16677038
SE($\hat{\sigma}$)	0.00714885	0.007552297	0.01099540
SE($\hat{\xi}$)	0.07050804	0.078211827	0.09433254

Tabuľka 8:

Po analýze nasledujúcich grafov opäť môžeme konštatovať, že minimálna hodnota, ktorú sme už spomínali aj pri modelovaní blokových miním (20.3.2007), sa pri každej prahovej úrovni nachádza mimo predikčného intervalu. Odhliadnuc od toho pozorovania, grafy spĺňajú predpoklady, ktoré sa očakávajú od nášho modelu.



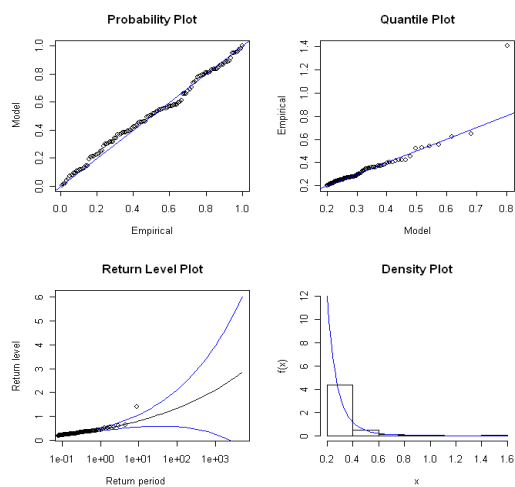
Obrázok 39: Model POT, $u = 0.15$



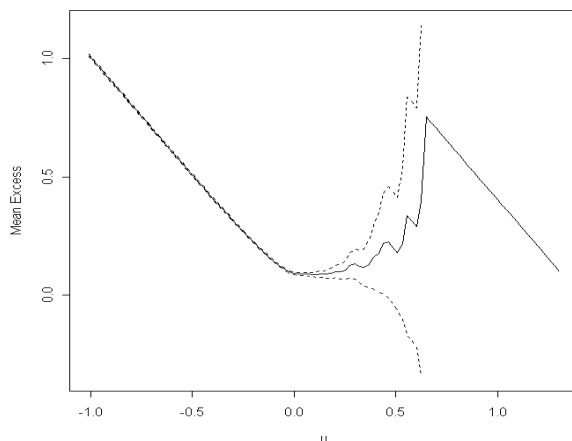
Obrázok 40: Model POT, $u = 0.16$

Vychádzajúc z grafickej analýzy, hodnotu $u = 0.2$ môžeme považovať za vhodnú voľbu prahu, lebo táto hodnota nie je až natoľko nízka, aby sme ohrozili kvalitu odhadov, a na druhej strane nie je ani vysoká, teda rozptyl odhadov parametrov neprevýši akceptovateľnú hranicu¹⁰. Túto voľbu môžeme overiť pomocou grafu MRL (viď Obrázok 42).

¹⁰Akceptovateľnú hranicu definujeme ako 10-15%

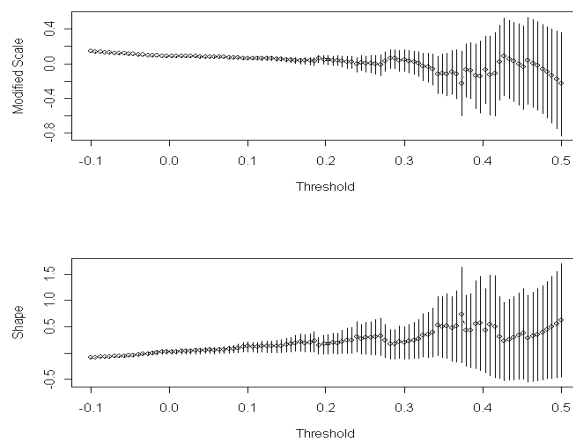


Obrázok 41: Model POT, $u = 0.2$

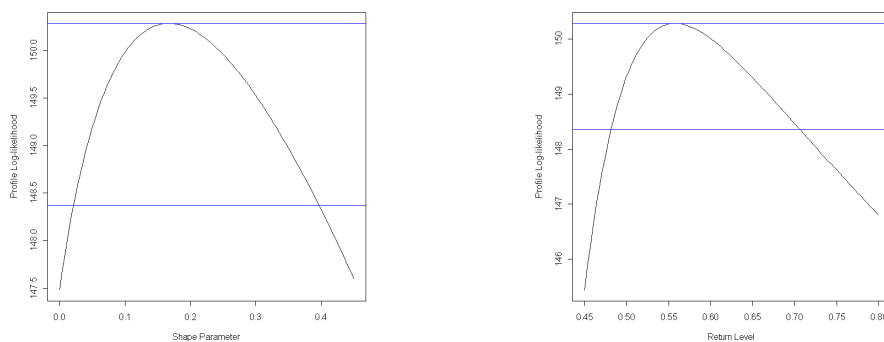


Obrázok 42: MRL plot

Na základe grafu MRL môžeme vyhlásiť $u = 0.2$ za vhodný výber prahovej hodnoty, ktorý je relatívne stabilný, čo sa aj odzrkadľuje na Obrázku 43. Interval stabilných prahových hodnôt určený pomocou prvého grafu je približne $[0.2, 0.25]$, druhého grafu je $[0.1, 0.2]$, teda prienikom je tá spomínaná hodnota $u = 0.2$. Profil vierohodnosti pre ξ a 2-ročný return level pre zvolený model je na Obrázku 44.



Obrázok 43: *ML odhady ako funkcie prahu*



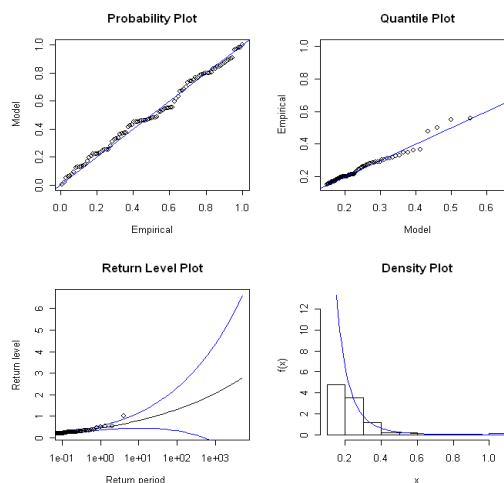
Obrázok 44: *Profil vierohodnosti pre ξ (vľavo); profil vierohodnosti pre 2-ročný 'return level' (vpravo)*

5.3.2 Kalibrácia modelov na intervale 4 rokov

Po aplikácii modelov prekročení ponad prah pre maximá dostávame výsledky zhrnuté v Tabuľke 9. Keď porovnáme tieto výsledky s výsledkami dosiahnuté pomocou analýzy 9 ročného časového radu, môžeme vyvodiť záver, že chyba odhadu v prípade dlhšieho časového radu je menšia, ale aproximácia je menej kvalitná. Prahová hodnota $u = 0.15$ poskytuje rozumný kompromis medzi varianciou a vychýlenosťou odhadov. Z toho dôvodu môžeme danú úroveň považovať za najlepšiu možnú voľbu, čo nám potvrdia aj Obrázky 46 a 47. Na Obrázku 48 môžeme odsledovať profil vierohodnosti pre ξ a 2-ročný return level.

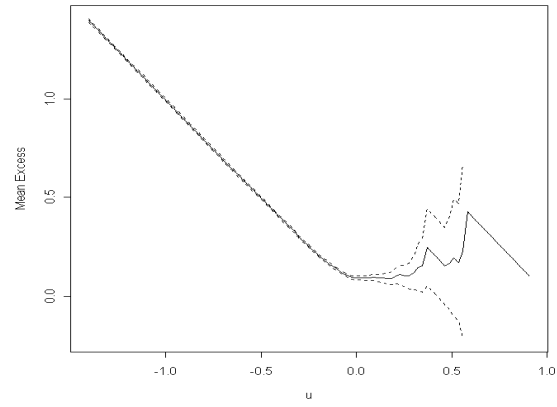
04-07	MAX		
prahová hodnota	$u = 0.12$	$u = 0.14$	$u = 0.15$
počet prekročení	130	101	94
NLLH	-184.2983	-139.5293	-134.1163
MLE($\hat{\sigma}$)	0.07875383	0.08218258	0.07502036
MLE($\hat{\xi}$)	0.12367570	0.11726022	0.16328222
SE($\hat{\sigma}$)	0.009736576	0.01120342	0.01113471
SE($\hat{\xi}$)	0.087872823	0.09397312	0.10815486

Tabuľka 9:

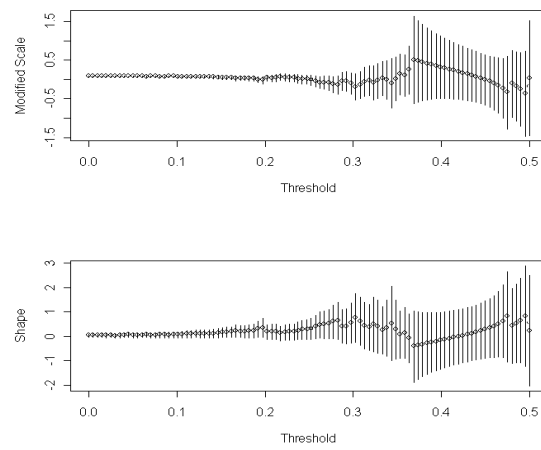


Obrázok 45: Model POT, $u = 0.15$

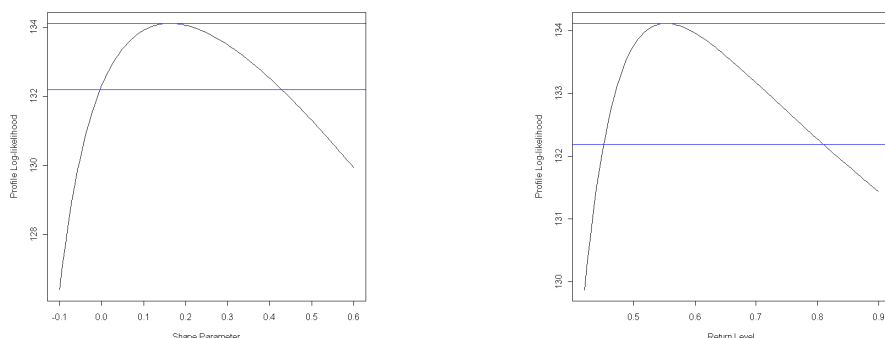
Poznámka: Určili sme prahovú hodnotu na úrovni $u = 0.15$, ktorému zodpovedá 94 prekročení. V knihe [10] Loretan a Phillips sa zaoberali s modelom prekročení ponad prah a ukázali, že počet prekročení by nemal prevýšiť 10% pozorovaných dát. Táto podmienka je splnená aj v prípade nášho zvoleného modelu.



Obrázok 46: *MRL plot*



Obrázok 47: *ML odhady ako funkcie prahu*



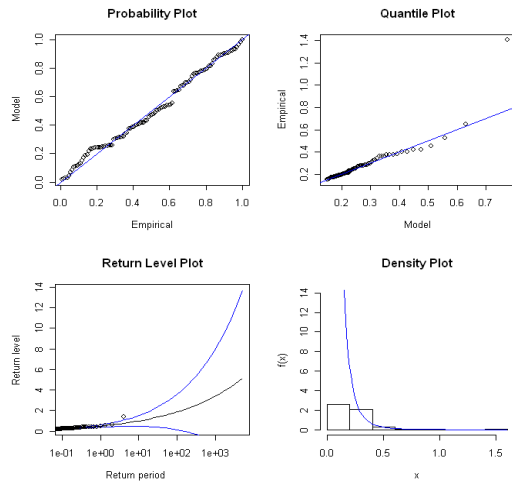
Obrázok 48: Profil vierohodnosti pre ξ (vľavo); profil vierohodnosti pre 2-ročný 'return level' (vpravo)

A teraz modelujeme minimá na 4 ročnom časovom horizonte. Postup je taký istý, ako v časti 5.3.1. V Tabuľke 10 sú uvedené výsledky modelovania.

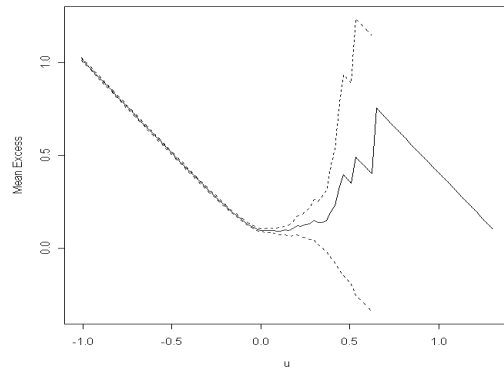
04-07	MIN		
prahová hodnota	$u = 0.1$	$u = 0.14$	$u = 0.15$
počet prekročení	196	120	109
NLLH	-281.6987	-166.838	-153.8265
MLE($\hat{\sigma}$)	0.0738131	0.07461595	0.0700261
MLE($\hat{\xi}$)	0.1689140	0.20502592	0.2475100
SE($\hat{\sigma}$)	0.007587461	0.00989383	0.01020632
SE($\hat{\xi}$)	0.074953344	0.09804131	0.11247512

Tabuľka 10:

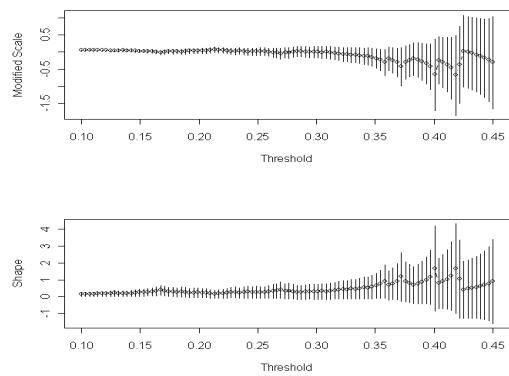
Vychádzajúc z grafickej analýzy pomocou grafov na stanovenie vhodnej úrovne prahu (viď Obrázok 49 a 50) hodnotu $u = 0.15$ môžeme považovať za relatívne stabilnú, a z toho dôvodu zvolíme prah na tejto úrovni. Obrázok 51 zobrazuje profil vierohodnosti pre ξ a 2-ročný return level v prípade vybraného modelu.



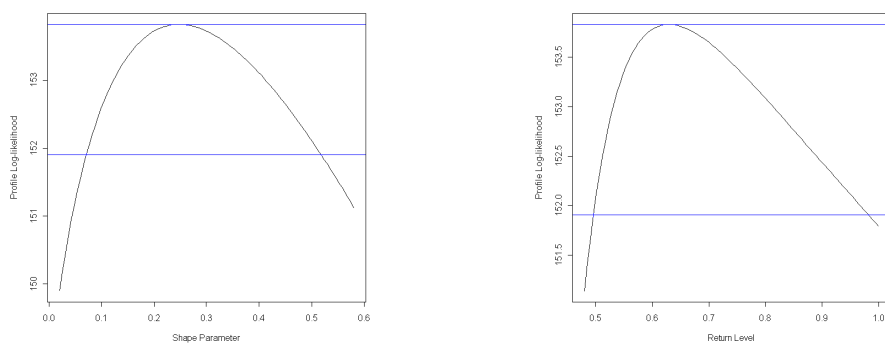
Obrázok 49: *Model POT*, $u = 0.15$



Obrázok 50: *MRL plot*



Obrázok 51: *ML odhady ako funkcie prahu*



Obrázok 52: *Profil vierohodnosti pre ξ (vľavo); profil vierohodnosti pre 2-ročný 'return level' (vpravo)*

5.4 Interpretácia výsledkov

Modely blokového maxima, resp. minima na 9 ročnom intervale:

99-07	dĺžka blokov	$\hat{\xi}$	$SE(\hat{\xi})$	\hat{z}_p
	$n = 5$	0.11605080	0.031239979	0.1058
max	$n = 10$	0.15531538	0.051228364	0.1586
	$n = 20$	0.20936871	0.084435822	0.2053
	$n = 5$	0.01831857	0.020504813	0.1269
min	$n = 10$	0.11419705	0.046436863	0.1712
	$n = 20$	0.09414792	0.055398100	0.2198

Tabuľka 11:

$\hat{\xi}$ – odhad parametra tvaru (ξ),

$SE(\hat{\xi})$ – štandardná chyba odhadu,

\hat{z}_p – hodnota, ktorá sa prekročí ročným maximom v každom roku s pravdepodobnosťou 5%.

Modely blokového maxima, resp. minima na 4 ročnom intervale:

04-07	dĺžka blokov	$\hat{\xi}$	$SE(\hat{\xi})$	\hat{z}_p
	$n = 5$	0.11440240	0.050247731	0.1023
max	$n = 10$	0.13806700	0.075424202	0.1523
	$n = 15$	0.09350040	0.076209440	0.1879
	$n = 5$	0.14737372	0.048471250	0.1286
min	$n = 10$	0.19837104	0.076035771	0.1785
	$n = 15$	0.15330380	0.069149131	0.2051

Tabuľka 12:

Parameter ξ je dôležitým parametrom, lebo určí tvar rozdelenia a potvrdí existenciu tzv. ťažkých koncov. Pri každom modelovaní farebne sme vyznačili nami zvolený najlepší model.

Modely prekročení ponad prah na 9 ročnom intervale:

99-07	prahová hodnota	$\hat{\xi}$	$SE(\hat{\xi})$	\hat{x}_m
	$u = 0.14$	0.16113398	0.072838657	0.6157
max	$u = 0.15$	0.18441160	0.079610955	0.6203
	$u = 0.2$	0.26334503	0.125304360	0.6371
	$u = 0.15$	0.16074768	0.070508040	0.5520
min	$u = 0.16$	0.18273687	0.078211827	0.5545
	$u = 0.2$	0.16677038	0.094332540	0.5571

Tabuľka 13:

$\hat{\xi}$ – odhad parametra tvaru (ξ),
 $SE(\hat{\xi})$ – štandardná chyba odhadu,
 \hat{x}_m – úroveň, ktorá je prekročená v priemere jeden krát každých m pozorovaní, v našom prípade $m = 500$.

Modely prekročení ponad prah na 4 ročnom intervale:

04-07	prahová hodnota	$\hat{\xi}$	$SE(\hat{\xi})$	\hat{x}_m
	$u = 0.12$	0.12367570	0.087872823	0.5506
max	$u = 0.14$	0.11726022	0.093973120	0.5495
	$u = 0.15$	0.16328222	0.108154860	0.5523
	$u = 0.1$	0.16891400	0.074953344	0.6113
min	$u = 0.14$	0.20502592	0.098041310	0.6190
	$u = 0.15$	0.24751000	0.112475120	0.6286

Tabuľka 14:

Metódy prekročení ponad prah sú efektívnejšie, lebo upriamujú pozornosť na všetky hodnoty, ktoré presahujú vysoký prah, zatiaľ čo modely blokového maxima využívajú iba jeden bod z každého bloku.

5.5 Stresové testovanie

Stresové testovanie je nástroj pre finančné inštitúcie, ktorý im umožní pripraviť sa na extrémne zmeny ktoré ich môžu negatívne ovplyvniť. Často sa to využíva pri meraní rôznych rizík, ako aj trhového a operačného rizika. Hodnotí sa vlastne zmena kapitálu pri zmene nejakých podmienok. Stresové testovanie je zamerané na všetky typy trhových rizík, ktorým sú finančné inštitúcie vystavené. Zahŕňa identifikovanie možných udalostí alebo budúcich zmien trhových faktorov, ktoré by mohli mať nepriaznivé účinky (napríklad zmena výmenného kurzu, pohyb úrokových mier, atď.) Ďalej špecifikuje, či kapitál inštitúcie je dostatočný na krytie rizika zo zmeny trhových faktorov.

Národná banka Slovenska vykonáva stresové testovanie s cieľom sledovať schopnosť a pripravenosť jednotlivých bánk a celého bankového sektora zvládnuť extrémne situácie na finančnom trhu. Predmetom stresového testovania sú hlavné riziká, ktorým sú banky vystavené, t.j. kreditné riziko, trhové riziko a riziko likvidity.

5.5.1 Trhové riziko

Na základe Opatrenia 4/2007 NBS [12] banky a pobočky zahraničných bánk v prípade, že chcú používať vlastný model na meranie trhového rizika musia postupovať podľa §184 – 201, v ktorých okrem iných vecí sa spomína aj stresové testovanie trhových faktorov (napr. výmenný kurz). Špeciálny typ stresového testovania je test senzitivity na tieto trhové faktory.

V našom prípade tento trhový faktor je výmenný kurz SKK/EUR, na základe dosiahnutých výsledkov z modelov blokového maxima a POT môžeme navrhnúť stresový scenár podľa nasledujúcej tabuľky.

Model (99-07)	dĺžka blokov	\hat{z}_p	$var(\hat{z}_p)$
Blok(max)	$n = 20$	0.2053	0.007
Blok(min)	$n = 20$	0.2198	0.003
	prahová hodnota	\hat{x}_m	$var(\hat{x}_m)$
POT(max)	$u = 0.2$	0.6371	0.0094
POT(min)	$u = 0.2$	0.5571	0.0035

Tabuľka 15:

\hat{z}_p – hodnota, ktorá sa prekročí ročným maximom v každom roku s pravdepodobnosťou 5%,

$var(\hat{z}_p)$ – hodnota v riziku prislúchajúca extrémnemu kvantilu \hat{z}_p ,

\hat{x}_m – úroveň, ktorá je prekročená v priemere jeden krát každých m pozorovaní, v našom prípade $m = 500$,

$var(\hat{x}_m)$ – hodnota v riziku prislúchajúca extrémnemu kvantilu \hat{x}_m .

Model (04-07)	dĺžka blokov	\hat{z}_p	$var(\hat{z}_p)$
Blok(max)	$n = 15$	0.1879	0.0058
Blok(min)	$n = 15$	0.2051	0.0048
	prahová hodnota	\hat{x}_m	$var(\hat{x}_m)$
POT(max)	$u = 0.15$	0.5523	0.0062
POT(min)	$u = 0.15$	0.6286	0.0074

Tabuľka 16:

5.5.2 Operačné riziko

V prípade, že banky majú záujem merať operačné riziko pomocou pokročilého prístupu podľa §208 – 215, tak sa musí modelovať aj neočakávaná strata a presnosť modelu sa očakáva na intervale spoľahlivosti 99.9%, pričom sa zohľadňujú interné, externé dáta, scenáre a faktory odrážajúce obchodné prostredie a systémy vnútornej kontroly. Je to pomerne náročná úloha, ktorá sa klasickými nástrojmi štatistiky alebo modelovania je takmer nemožné uskutočniť a preto sa často hľadajú nové prístupy, ktoré by mohli byť schopné modelovať extrémne udalosti operačného rizika ako napríklad EVT. Keďže nemáme k dispozícii dáta, ktoré by nám umožnili modelovať operačné riziko bánk, tak len poskytujeme štatistický nástroj, ktorý sa dá použiť aj v takomto prípade.

6 Záver

V predloženej diplomovej práci sme zhrnuli prakticky najvyužívanejšie metódy teórie extrémnych hodnôt na modelovanie mimoriadnych udalostí. V súlade s cieľom práce bolo hlavné ťažisko sústredené na dve základné metódy - metódy blokového maxima, metódy prekročení ponad prah. V praktickej časti práce sme demonštrovali použitie uvedených metód pri analýze extrémnych zmien výmenného kurzu SKK/EUR. Na analýzu sme ako prvú využili metódu blokového maxima. Parametre modelov sme nakalibrovali pomocou GEV rozdelenia na 9 a 4 ročnom časovom horizonte. V ďalšej časti sme pri analýze používali modely POT na odhad úrovne extrémnej zmeny, ktorá bude prekročená v nasledujúcom období s vopred danou pravdepodobnosťou. Na kalibráciu sme používali GPD rozdelenie. Jednotlivé prístupy poskytujú výsledky, ktoré sa väčšinou systematicky výrazne nelíšia. Prístupy využívajúce historické údaje na dlhšie časové periódy produkujú menej variabilné výsledky ako tie, ktoré využívajú kratšie časové periódy, ale z praktického hľadiska kratší časový rad je adekvátnejší. Z ekonomického hľadiska je lepšie uvažovať "čerstvé" dáta, keďže sa ekonomika stále mení. Na základe dosiahnutých výsledkov sme sa pokúsili navrhnúť stresový scenár výmenného kurzu.

Diplomová práca poskytuje nielen analýzu výsledkov vybraného časového radu, ale slúži ako podklad k ľubovoľnej analýze extrémnych udalostí, keďže sa postup a teória zhoduje pre akýkoľvek časový rad.

Referencie

- [1] Stuart Coles: *An introduction to statistical modeling of extreme values*, Springer (2001)
- [2] P.Embrechts, C.Klüppelberg, T. Mikosch: *Modelling Extremal Events for Insurance and Finance*, Springer (2003)
- [3] R.A.Fisher, L.H.C.Tippett: *On the estimation of the frequency distributions of the largest or smallest member of a sample*, Proceedings of the Cambridge Philosophical Society (1928)
- [4] B.V.Gnedenko: *Sur la distribution limite du terme maximum d'une série aléatoire*, Annals of mathematics (1943)
- [5] M.R.Leadbetter, G.Lindgren, H. Rootzén: *Extremes and Related Properties of Random Sequences and Series*, Springer Verlag (1983)
- [6] M.R.Leadbetter, H.Rootzen: *Extremal theory for stochastic processes* Annals of Probability (1988)
- [7] R.L.Smith: *Maximum likelihood estimation in a class of non-regular cases*, (1985)
- [8] A.C.Davison, R.L.Smith: *Models for exceedances over high thresholds*, Journal of the Royal Statistical Society (1990)
- [9] Ch.Zhou, Ch.Wu, H.Liu, F.Liu: *A new method to choose the threshold in the POT model*, (2000)
<http://ssrn.com/abstract=987796>
- [10] M.Loretan, P.Phillips : *Testing the Covariance Stationarity of Heavy-tailed Time Series*, Journal of Empirical Finance 1, 211-248, (1994)
- [11] www.nbs.sk
- [12] www.nbs.sk/LEGS/2007/OPAT4-07.pdf
- [13] <http://cran.r-project.org>

Príloha

Výsledky modelovania v prípade blokových maxím resp. miním, na 9 ročnom intervale:

99-07	MAX	MIN
dĺžka blokov	$n = 30$	$n = 30$
počet dát	75	75
NLLH	-35.95195	-56.18375
MLE($\hat{\mu}$)	0.2075548	-0.21782395
MLE($\hat{\sigma}$)	0.1174083	0.08839088
MLE($\hat{\xi}$)	0.1480449	0.18012497
SE($\hat{\mu}$)	0.01529029	0.011534787
SE($\hat{\sigma}$)	0.01173974	0.009023192
SE($\hat{\xi}$)	0.08919004	0.088964744

Tabuľka 17:

Výsledky modelovania v prípade blokových maxím resp. miním, na 4 ročnom intervale:

04-07	MAX	MIN
dĺžka blokov	$n = 4$	$n = 4$
počet dát	250	250
NLLH	-232.3974	-224.6269
MLE($\hat{\mu}$)	0.05896385	-0.08325923
MLE($\hat{\sigma}$)	0.07680751	0.08224684
MLE($\hat{\xi}$)	0.10568387	0.05981058
SE($\hat{\mu}$)	0.005399915	0.005623204
SE($\hat{\sigma}$)	0.004051089	0.004015193
SE($\hat{\xi}$)	0.043066117	0.029992449

Tabuľka 18:

Výsledky modelovania POT pre maximá, na 9 ročnom intervale:

99-07	MAX			
prahová hodnota	$u = 0.1$	$u = 0.13$	$u = 0.16$	$u = 0.22$
počet prekročení	341	254	191	90
NLLH	-437.8397	-328.0621	-250.7835	-98.4393
MLE($\hat{\sigma}$)	0.09164872	0.08683131	0.07820646	0.1020499
MLE($\hat{\xi}$)	0.10592143	0.15239182	0.23549207	0.1884983
SE($\hat{\sigma}$)	0.007006436	0.008005732	0.008974736	0.01696027
SE($\hat{\xi}$)	0.054289726	0.068269852	0.090737605	0.13010676

Tabuľka 19:

Výsledky modelovania POT pre minimá, na 9 ročnom intervale:

99-07	MIN			
prahová hodnota	$u = 0.13$	$u = 0.14$	$u = 0.18$	$u = 0.22$
počet prekročení	281	247	153	96
NLLH	-405.481	-352.6121	-216.1071	-130.7053
MLE($\hat{\sigma}$)	0.07611605	0.07714592	0.07296883	0.07451031
MLE($\hat{\xi}$)	0.13246517	0.13441429	0.20544535	0.23514234
SE($\hat{\sigma}$)	0.006393555	0.00686726	0.008932807	0.01152168
SE($\hat{\xi}$)	0.059704408	0.06291413	0.093470260	0.11869107

Tabuľka 20:

Výsledky modelovania POT pre maximá, na 4 ročnom intervale:

04-07	MAX	
prahová hodnota	$u = 0.1$	$u = 0.16$
počet prekročení	153	86
NLLH	-207.4704	-125.5308
MLE($\hat{\sigma}$)	0.08885341	0.06912939
MLE($\hat{\xi}$)	0.06484432	0.21207075
SE($\hat{\sigma}$)	0.009455165	0.01131357
SE($\hat{\xi}$)	0.069463482	0.12530155

Tabuľka 21:

Výsledky modelovania POT pre minimá, na 4 ročnom intervale:

04-07	MIN	
prahová hodnota	$u = 0.12$	$u = 0.16$
počet prekročení	155	97
NLLH	-221.3439	-136.7862
MLE($\hat{\sigma}$)	0.07259293	0.06768552
MLE($\hat{\xi}$)	0.19474007	0.28293437
SE($\hat{\sigma}$)	0.00852536	0.01088347
SE($\hat{\xi}$)	0.087163099	0.12804257

Tabuľka 22: