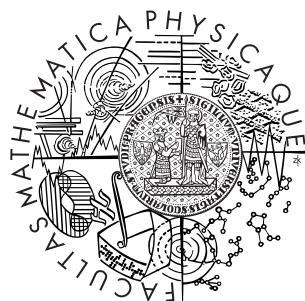


Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE



Daniel Veselý

Teorie extremálních rozdělení ve financích

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: Mgr. Karel Vaníček,
Studijní program: Matematika, Obecná matematika

2009

Děkuji Mgr. Karlu Vaníčkovi za vedení, konzultace a pomoc při vytváření bakalářské práce.

Prohlašuji, že jsem svou bakalářskou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 31.7.2009

Daniel Veselý

Obsah

1 Úvod	5
2 Rodiny extremálních rozdělení a jejich charakteristiky	7
2.1 Úvod a značení	7
2.2 Konvergence maxima	8
2.3 Maximum domain of attraction	11
2.4 Maximum domain of attraction Fréchetovy distribuční funkce	12
2.5 Maximum domain of attraction Weibullový distribuční funkce	15
2.6 Maximum domain of attraction Gumbelovy distribuční funkce	16
2.7 Obecná extremální distribuční funkce a obecná Paretova distribuční funkce	18
3 Parametrické a neparametrické odhadování charakteristik extremálních rozdělení	23
3.1 Úvod	23
3.2 Metody grafické analýzy dat	24
3.3 Parametrické odhadování charakteristik rozdělení	28
3.4 Odhadování pomocí maximum domain of attraction	30
3.5 Odhadování chvostů pomocí obecné Paretovy distribuční funkce	36
4 Míry rizik a jejich vlastnosti	39
4.1 Value-at-Risk - VaR	39
4.2 Expected shortfall - ES	40
5 Aplikace na reálná data	42
5.1 Úvod	42
5.2 Analýza historických cen akcií Bank of America Corporation	43
Literatura	57

Název práce: Teorie extremálních rozdělení ve financích

Autor: Daniel Veselý

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: Mgr. Karel Vaníček

e-mail vedoucího: karel.vanicek@ge.com

Abstrakt: V této práci studujeme použití teorie extremálních rozdělení (EVT) na měření extrémního rizika finančních aktiv. Nejdříve shrneme základní poznatky z teorie extremálních rozdělení. Popíšeme maximum domain of attraction Fréchetova, Weibullovu a Gumbelovu rozdělení, jejich sjednocení v obecnou extremální distribuční funkci (GEV) a odvodíme obecnou Paretovu distribuční funkci (GPD). Poté se zaměříme na několik metod odhadů charakteristik extremálních rozdělení. Mezi těmito odhady popíšeme některé parametrické a také neparametrické metody. Tyto metody později využijeme na odhad chvostu distribuční funkce ztrát a na stanovení dvou měr rizika. Konkrétně na Value-at-Risk a expected shortfall. Na závěr aplikujeme zmíněné odhady na reálná data z finančních trhů.

Klíčová slova: extrémní riziko, GEV, GPD, EVT

Title: Extreme Value Theory in Finance

Author: Daniel Veselý

Department: Department of Probability and Mathematical Statistics

Supervisor: Mgr. Karel Vaníček

Supervisor's e-mail address: karel.vanicek@ge.com

Abstract: In this work we study the application of extreme value theory (EVT) on measuring extreme risk of financial assets. First we describe basic theoretical knowledge of extreme value theory. We describe maximum domain of attraction of Fréchet, Weibull and Gumbel distribution, their unifying into generalised extreme value distribution function (GEV) and we derive generalised Pareto distribution function (GPD). Then we focus on several methods of estimating characteristics of extreme distributions. Among these estimations we describe some parametric and also non parametric methods. Then we use these methods on tail estimation of distribution function of losses and on two risk measures, Value-at-Risk and expected shortfall. In conclusion we apply described methods on real data from financial markets.

Keywords: extreme risk, GEV, GPD, EVT

Kapitola 1

Úvod

V této práci se budeme zabývat studiem extrémních jevů ve financích, tedy jevů, které se vyskytují pouze výjimečně, avšak jejichž vliv na nás je obrovský. Extrémní jevy se vyskytují ve všech oblastech. My se zaměříme na extrémní ztráty z finančních aktiv. Takovéto ztráty mohou mít združující důsledky a proto je důležité jim věnovat pozornost. V řeči matematiky budeme ztráty modelovat jako náhodné veličiny X_1, \dots, X_n , které mají distribuční funkci F . Extremní ztrátu potom chápeme jako náhodnou veličinu X , která nabude hodnoty z chvostu své distribuční funkce F . K modelování chvostu distribuční funkce ztrát F budeme používat teorii extremálních rozdělení. Teorii extremálních rozdělení popisuje druhá kapitola. V té shrneme základní poznatky, které budeme potřebovat dále k aplikaci teorie na reálná data. Rozdělíme zde distribuční funkce do tří rodin, které určují tři základní extremální distribuční funkce: Fréchetova, Weibullova a Gumbelova. Další důležitou distribuční funkcí, kterou zde odvodíme je obecná Paretova distribuční funkce. Ve třetí kapitole popíšeme metody pro parametrický a neparametrický odhad charakteristik extremálních rozdělení z naměřených dat. Parametrické odhady provedeme pomocí metody maximální věrohodnosti a pomocí metody pravděpodobnostně vážených momentů. Neparametrické metody rozdělíme na metodu pomocí maximum domain of attraction, kterou provedeme pomocí Pickandova, Hillova a Dekkers-Einmahl-de Haanova odhadu, a na odhad parametrů obecné Paretovy distribuční funkce. Uvedeme zde také některé grafické metody pro analýzu chvostu. Ve čtvrté kapitole popíšeme míry, pomocí kterých lze měřit riziko extrémních jevů. Tyto míry se nazývají Value-at-Risk a expected shortfall. Také zde uvedeme metody, jak tyto míry rizika

určit pomocí teorie extremálních rozdělení. V páté kapitole potom aplikujeme popsanou teorii na reálná data z finančních trhů a stanovíme jejich míry extrémního rizika. Konkrétně budeme studovat data akcií Bank of America Corporation. Při analýze budeme klást důraz na vliv globální finanční krize, která začala v roce 2007 ve Spojených státech amerických, na naměřená data.

Kapitola 2

Rodiny extremálních rozdělení a jejich charakteristiky

2.1 Úvod a značení

V této kapitole se budeme zabývat základními poznatkami z teorie extremálních rozdělení. Naším cílem bude zkoumat vlastnosti konvergence normovaných a centralizovaných maxim z náhodných výběrů z nějakého rozdělení a vlastnosti jejich limitních rozdělení. Zaměříme se na to, jaké vlastnosti distribuční funkce F zaručují existenci nedegenerovaného limitního rozdělení maxim z náhodného výběru s touto distribuční funkcí. Dále prozkoumáme jaká rozdělení mohou být limitní rozdělení maxim náhodného výběru s nějakou distribuční funkcí F . Nejdůležitějším výsledkem této části je Fisher-Tippettova věta, která určuje tři možná nedegenerovaná limitní rozdělení pro centralizovaná a normovaná maxima náhodného výběru. Tato tři rozdělení se nazývají extremální rozdělení. Poté se zaměříme na to, jaké vlastnosti distribuční funkce F implikují, že maxima z výběru s touto distribuční funkcí konvergují k dané extremální distribuční funkci. Současně nás bude samozřejmě také zajímat, jak můžeme volit posloupnosti centralizujících a normujících konstant. Nakonec se zaměříme na tzv. obecnou extremální distribuční funkci, která bude velmi důležitá při pozdější aplikaci teorie na reálná data.

V celé této kapitole bude X_1, X_2, \dots posloupnost iid náhodných veličin se známou df F . Budeme vyšetřovat vlastnosti výběrového maxima

$$M_1 = X_1, M_n = \max(X_1, \dots, X_n), n \geq 2.$$

Odpovídající vlastnosti výběrového minima můžeme zjistit pomocí transformace:

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

Dále si označíme pravý chrost distribuční funkce F jako $\bar{F} = 1 - F$. Číslo $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$ budeme nazývat hraniční hodnota distribuční funkce F .

2.2 Konvergence maxima

Nejdříve odvodíme distribuční funkci maxima M_n :

$$P(M_n \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = F^n(x),$$

Extrémní jevy se odehrávají blízko hraniční hodnoty x_F df F . Proto jsou intuitivně vlastnosti M_n závislé na vlastnostech pravého chrostu df F .

Pro všechna $x < x_F$ platí:

$$P(M_n \leq x) = F^n(x) \rightarrow 0, n \rightarrow \infty,$$

pokud je $x_F < \infty$ potom pro $x \geq x_F$ platí:

$$P(M_n \leq x) = F^n(x) = 1.$$

Tedy dostáváme, že $\lim_{n \rightarrow \infty} M_n = x_F$ v pravděpodobnosti a skoro jistě. Tento poznatek nám toho však o vlastnostech M_n mnoho neprozradí. Proto budeme dále studovat slabou konvergenci centralizovaných a normovaných maxim. Tedy se budeme zabývat vztahem

$$\frac{(M_n - d_n)}{c_n} \xrightarrow{d} H, n \rightarrow \infty$$

pro nějakou nedegenerovanou distribuční funkci H a konstanty $c_n > 0$, $d_n \in \mathbb{R}$. Tedy budeme se zajímat o pravděpodobnosti:

$$P\left(\frac{M_n - d_n}{c_n} \leq x\right).$$

Tento výraz můžeme dále upravit na

$$P(M_n \leq u_n(x)),$$

kde $u_n(x) = c_n * x + d_n$.

Nejdříve se budeme zabývat otázkou, jaké vlastnosti distribuční funkce F zajišťují existenci limity $P(M_n \leq u_n(x))$ pro $n \rightarrow \infty$. Jak jsme již intuitivně přepokládali, existence této limity bude záviset na vlastnostech pravého chvostu \bar{F} distribuční funkce F . Tyto vlastnosti popisují následující dvě tvrzení, která jsou základními stavebními kameny v teorii extremálních rozdělení.

Tvrzení 2.1. *Pro $\tau \in [0, \infty]$ a posloupnost (u_n) reálných čísel jsou následující výrazy ekvivalentní:*

$$\begin{aligned} P(M_n \leq u_n(x)) &\rightarrow e^{-\tau} \\ n\bar{F}(u_n) &\rightarrow \tau. \end{aligned}$$

Důkaz tohoto tvrzení je uveden v [2] na straně 116.

Tvrzení 2.2. *Nechť F je distribuční funkce s $x_F \leq \infty$ a nechť $\tau \in (0, \infty)$. Potom existuje posloupnost (u_n) splňující $n\bar{F}(u_n) \rightarrow \tau$ tehdy a jen tehdy, když*

$$\lim_{x \uparrow x_F} \frac{\bar{F}(x)}{\bar{F}(x-)} = 1$$

$$a \bar{F}(x_F-) = 1.$$

Toto tvrzení nalezneme v [2] na straně 117. Důsledkem těchto dvou tvrzení jsou omezení na distribuční funkce, ke kterým existuje nedegenerované limitní rozdělení pro M_n . Například pro distribuční funkce s $x_F < \infty$, které mají v bodě x_F skok, takovéto rozdělení neexistuje. Tento fakt vyřadí z našeho uvažování všechna diskrétní rozdělení s konečnou množinou hodnot. Pro diskrétní náhodnou veličinu s $x_F = \infty$ můžeme $\lim_{x \uparrow x_F} \frac{\bar{F}(x)}{\bar{F}(x-)} = 1$ přepsat na $\lim_{n \rightarrow \infty} \frac{\bar{F}(n)}{\bar{F}(n-1)} = 1$. Pomocí tohoto vztahu můžeme vyloučit nedegenerovaná limitní rozdělení pro M_n z náhodného výběru z Poissonova, geometrického a negativně binomického rozdělení. A dále všechna diskrétní rozdělení s $x_F = \infty$, pro která klesá $\bar{F}(n)$ příliš rychle. V následující definici popíšeme speciální třídu náhodných veličin. Tyto veličiny se později ukáží jako jediná možná limitní rozdělení pro konvergenci M_n z nějakého rozdělení.

Definice 2.1. Nedegenerovaná náhodná veličina X se nazývá max-stabilní, pokud pro všechna $n \geq 2$ a pro X_1, X_2, \dots, X_n iid se stejným rozdělením jako X platí:

$$\max(X_1, X_2, \dots, X_n) \stackrel{d}{=} c_n X + d_n.$$

Tedy platí, že každá distribuční funkce max-stabilní náhodné veličiny je limitní distribuční funkce pro výběrová maxima z nějakého náhodného výběru. Následující tvrzení navíc říká, že rodina max-stabilních distribučních funkcí jsou jediná možná limitní rozdělení pro výběrová maxima z nějakého náhodného výběru.

Tvrzení 2.3. Rodina max-stabilních distribučních funkcí se shoduje s rodinou všech možných (nedegenerovaných) limitních rozdělení pro (normovaná a centralizovaná) výběrová maxima.

Důkaz tohoto tvrzení je uveden v [2] na straně 121. Nyní uvedeme tzv. Fisher-Tippettovu větu, která je hlavním výsledkem této podkapitoly.

Tvrzení 2.4. Nechť (X_n) je posloupnost iid náhodných veličin. Pokud existují normující konstanty $c_n > 0$, $d_n \in \mathbb{R}$ a nedegenerovaná distribuční funkce H taková, že

$$\frac{(M_n - d_n)}{c_n} \xrightarrow{d} H, n \rightarrow \infty,$$

potom je H jedna ze tří následujících distribučních funkcí:

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

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

$$Gumbel : \Lambda(x) = \exp(-e^{-x}), x \in \mathbb{R}.$$

Tyto tři distribuční funkce se nazývají extremální distribuční funkce. Odpovídající náhodné veličiny se nazývají extremální náhodné veličiny.

Náznak důkazu tohoto tvrzení je uveden v [2] na straně 122. Extremální distribuční funkce jsou spolu spojeny podle následujícího vztahu pro $x > 0$:

$$X \text{ má df } \phi_\alpha(x) \iff \ln(X^\alpha) \text{ má df } \Lambda(x) \iff -X^{-1} \text{ má df } \psi_\alpha(x).$$

2.3 Maximum domain of attraction

V této kapitole budeme řešit otázku: Jaké vlastnosti df F implikují, že maxima z náhodného výběru s touto df (po znormování a centralizování) konvergují slabě k extremální distribuční funkci H ? Dále si řekneme, jak můžeme volit normalizační konstanty a zda jsou tyto konstanty jednoznačně určeny. První otázka, kterou v této kapitole zodpovíme je: Může se stát, že pomocí různých normalizačních konstant mohou maxima ze stejného náhodného výběru konvergovat slabě k různým extremálním distribučním funkcím? Odpověď na tuto otázku nám dá následující tvrzení, které je uvedeno v [2] na straně 554.

Tvrzení 2.5. *Nechť A, B, A_1, A_2, \dots jsou náhodné veličiny a $b_k > 0$, $\beta_k > 0$, $a_k \in \mathbb{R}$, $\alpha_k \in \mathbb{R}$ jsou konstanty. Pokud platí:*

$$\frac{A_n - a_n}{b_n} \xrightarrow{d} A.$$

Potom vztah:

$$\frac{A_n - \alpha_n}{\beta_n} \xrightarrow{d} B$$

platí, právě když:

$$\lim_{n \rightarrow \infty} \frac{b_n}{\beta_n} = b \in [0, \infty), \quad \lim_{n \rightarrow \infty} \frac{a_n - \alpha_n}{\beta_n} = a \in \mathbb{R}.$$

Potom $B \stackrel{d}{=} bA + a$ a a, b jsou jednoznačně určeny. A je nedegenerovaná, právě když $b > 0$ a potom jsou A a B stejného typu.

Tedy limitní rozdělení pro maxima z náhodného výběru je jednoznačně určeno až na afinní transformaci. Před zodpovězením dalších otázek definujeme maximum domain of attraction (MDA).

Definice 2.2. *Řekneme, že náhodná veličina X (odpovídající distribuční funkce F) patří do maximum domain of attraction extremální distribuční funkce H , pokud existují konstanty $c_n > 0$, $d_n \in \mathbb{R}$ takové, že platí*

$$\frac{(M_n - d_n)}{c_n} \xrightarrow{d} H, n \rightarrow \infty.$$

V tomto případě píšeme: $X \in MDA(H)$ ($F \in MDA(H)$)).

Definice 2.3. \bar{F} je stejně se měnící s indexem $-\alpha$ pro nějaké $\alpha \geq 0$, pokud

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(xt)}{\bar{F}(x)} = t^{-\alpha}, t > 0,$$

potom pišeme $\bar{F} \in R_{-\alpha}$.

Definice 2.4. Distribuční funkce F a H mají ekvivalentní chvosty, pokud $x_F = x_H$ a platí:

$$\lim_{x \uparrow x_F} \frac{\bar{F}(x)}{\bar{H}(x)} = c$$

pro nějakou konstantu $0 < c < \infty$.

Distribuční funkce, které mají ekvivalentní chvosty, patří do stejné $MDA(H)$ a dokonce platí, že pro ně můžeme volit stejné normovací konstanty. Výběrové maximum M_n je empirická verze $(1 - \frac{1}{n})$ -kvantilu distribuční funkce F . Proto budeme volit normující konstanty $c_n > 0$ jako $(1 - \frac{1}{n})$ -kvantil. Kvantilovou funkci budeme dále značit jako $F^-(t) = \{\inf x \in \mathbb{R} : F(x) \geq t\}$, $0 < t < 1$. Následující tvrzení je důsledkem Tvrzení 2.1 a poslouží nám k některým odhadům v dalších kapitolách.

Tvrzení 2.6 (Charakterizace $MDA(H)$).

Distribuční funkce F patří do $MDA(H)$ s normujícími konstantami $c_n > 0, d_n \in \mathbb{R}$, právě když

$$\lim_{n \rightarrow \infty} n\bar{F}(c_n x + d_n) = -\ln(H(x)), x \in \mathbb{R}.$$

Pokud $H(x) = 0$ potom je limita rovna ∞ .

2.4 Maximum domain of attraction Fréchetovy distribuční funkce

V této kapitole charakterizujeme distribuční funkce, které patří do $MDA(\Phi_\alpha)$ pro $\alpha > 0$. Protože Fréchetova distribuční funkce má tvar:

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

Platí, že $\bar{F}(x) = 1 - \exp(-x^{-\alpha})$ a podle Taylorova rozvoje platí, že $\bar{F}(x) \sim x^{-\alpha}$. Tedy chvost $\Phi_\alpha(x)$ klesá jako mocnina a patří mezi rozdělení

s těžkými chvosty. Následující tvrzení nám prozradí, že všechny distribuční funkce, které patří do $MDA(\Phi_\alpha)$, mají těžký chvost, který klesá jako mocnina. Důkaz tohoto tvrzení je uveden v [2] na straně 131.

Tvrzení 2.7. *Distribuční funkce F patří do $MDA(\Phi_\alpha)$ pro $\alpha > 0$ právě když $\bar{F}(x) = x^{-\alpha}L(x)$. Kde L je kladná Lebesgueovsky měřitelná funkce na $(0, \infty)$ a $L \in R_0$. Pokud $F \in MDA(\Phi_\alpha)$, potom*

$$c_n^{-1}M_n \xrightarrow{d} \Phi_\alpha,$$

kde normující konstanty c_n můžeme volit jako $F^\leftarrow(1 - \frac{1}{n})$. Pozn. funkci L s těmito vlastnostmi budeme nazývat "pomalu se měnící funkce L ".

Důsledkem tohoto tvrzení je, že všechny distribuční funkce $F \in MDA(\Phi_\alpha)$ mají $x_F = \infty$. Dalším důsledkem tvrzení je, že pro $X \in MDA(\Phi_\alpha)$ je $E(X^+)^{\delta} = \infty$ pro $\delta > \alpha$. Tedy jsou to rozdělení s velmi těžkými chvosty a $x_F = \infty$, která se používají k modelování velmi volatilních trhů, akcií apod. Dále platí

$$F \in MDA(\Phi_\alpha) \Leftrightarrow \bar{F} \in R_{-\alpha},$$

což je jednoduchá charakterizace $MDA(\Phi_\alpha)$ podle tvaru chvostu. Další možná charakterizace $MDA(\Phi_\alpha)$ je pomocí tzv. von Miseho podmínky.

Tvrzení 2.8. *Nechť je F absolutně spojitá distribuční funkce s hustotou f . Pokud platí:*

$$\lim_{x \rightarrow \infty} \frac{xf(x)}{\bar{F}(x)} = \alpha > 0,$$

potom $F \in MDA(\Phi_\alpha)$.

Toto tvrzení je odvozeno v [2] na straně 132 a lze podle něj charakterizovat $MDA(\Phi_\alpha)$ pomocí von Miseho podmínky:

$MDA(\Phi_\alpha)$ se skládá z distribučních funkcí splňujících von Miseho podmínu a z distribučních funkcí s nimi chvostově ekvivalentních.

Příklad 2.1 (distribuční funkce z $MDA(\Phi_\alpha)$).

Do $MDA(\Phi_\alpha)$ patří například Paretovo a Cauchyho rozdělení.

a) Paretovo rozdělení

s hustotou

$$f(x) = \frac{a}{b} \left(\frac{b}{x}\right)^{a+1}, a > 0, b > 0, x \geq b$$

distribuční funkce má tvar

$$F(x) = 1 - \left(\frac{b}{x}\right)^a, a > 0, b > 0, x \geq b$$

tedy chvost je

$$\bar{F}(x) = \left(\frac{b}{x}\right)^a$$

podle první charakterizace $MDA(\Phi_\alpha)$ dostáváme:

$$\bar{F}(x) = \left(\frac{b}{x}\right)^a \sim \frac{1}{x^a} \in R_{-a} \Rightarrow F \in MDA(\Phi_a)$$

podle druhé charakterizace $MDA(\Phi_\alpha)$ dostáváme:

von Miseho podmínka:

$$\lim_{x \rightarrow \infty} \frac{xf(x)}{\bar{F}(x)} = \lim_{x \rightarrow \infty} \frac{\frac{ax}{b} \left(\frac{b}{x}\right)^{a+1}}{\left(\frac{b}{x}\right)^a} = \lim_{x \rightarrow \infty} \frac{a \left(\frac{b}{a}\right)^a}{\left(\frac{b}{a}\right)^a} = a \Rightarrow F \in MDA(\Phi_a).$$

b) Standardní Cauchyho rozdělení

s hustotou

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

distribuční funkce má tvar

$$F(x) = \frac{\arctan(x) + \frac{\pi}{2}}{\pi}, x \in \mathbb{R}$$

tedy chvost je

$$\bar{F}(x) = 1 - \frac{\arctan(x) + \frac{\pi}{2}}{\pi}, x \in \mathbb{R}$$

podle první charakterizace $MDA(\Phi_\alpha)$ dostáváme:

$$\bar{F}(x) \sim \frac{1}{\pi x} \in R_{-1} \Rightarrow F \in MDA(\Phi_1)$$

podle druhé charakterizace $MDA(\Phi_\alpha)$ dostáváme:
von Miseho podmínka:

$$\lim_{x \rightarrow \infty} \frac{xf(x)}{\bar{F}(x)} = \lim_{x \rightarrow \infty} \frac{\frac{x}{\pi(1+x^2)}}{\frac{1}{\pi x}} = \lim_{x \rightarrow \infty} \frac{x^2}{1+x^2} = 1 \Rightarrow F \in MDA(\Phi_1).$$

Další vlastnosti Cauchyho a Paretova rozdělení jsou uvedeny v [1] na straně 25 a 27.

2.5 Maximum domain of attraction Weibullový distribuční funkce

V této kapitole charakterizujeme distribuční funkce, které patří do $MDA(\Psi_\alpha)$ pro $\alpha > 0$. Důležitým poznatkem je, že všechny distribuční funkce z $MDA(\Psi_\alpha)$ mají konečný hraniční bod x_F . MDA Weibullový a Frechetovy distribuční funkce jsou si velice podobné. To nám napovídá i vztah $\Psi_\alpha(-\frac{1}{x}) = \Phi_\alpha(x), x > 0$. První charakterizace $MDA(\Psi_\alpha)$ podle tvaru chvostu F zní:

$$F \in MDA(\Psi_\alpha) \Leftrightarrow x_F < \infty, \bar{F}(x_F - \frac{1}{x}) \in R_{-\alpha}.$$

Normující konstanty zde můžeme volit jako: $c_n = x_F - F^\leftarrow(1 - \frac{1}{n})$ a $d_n = x_F$. Modifikací Tvrzení 2.8 můžeme také pro $MDA(\Psi_\alpha)$ stanovit von Miseho podmínu:

Tvrzení 2.9. *Nechť je F absolutně spojitá distribuční funkce s hustotou f , která je kladná na nějakém konečném intervalu (z, x_F) . Pokud platí:*

$$\lim_{x \uparrow x_F} \frac{(x_F - x)f(x)}{\bar{F}(x)} = \alpha > 0,$$

potom $F \in MDA(\Psi_\alpha)$.

Opět můžeme $MDA(\Psi_\alpha)$ charakterizovat pomocí von Miseho podmínky:

$MDA(\Psi_\alpha)$ se skládá z distribučních funkcí splňujících von Miseho podmínu a z distribučních funkcí s nimi chvostově ekvivalentních.

Do $MDA(\Psi_\alpha)$ patří například rovnoměrné a beta rozdělení.

Příklad 2.2 (rovnoměrné rozdělení na $(0,1)$).

hustota $f(x) = 1, x \in (0, 1)$

distribuční funkce $F(x) = x, x \in (0, 1)$

chvost $\bar{F}(x) = 1 - x, x \in (0, 1)$

$x_F = 1$

Pomocí charakterizace $MDA(\Psi_\alpha)$ podle tvaru chvostu dostaneme:

$$\bar{F}(x_F - \frac{1}{x}) = \frac{1}{x} \in R_{-1} \Rightarrow F \in MDA(\Psi_1).$$

Pomocí von Miseho podmínky dostaneme:

$$\lim_{x \uparrow 1} \frac{(1-x)1}{1-x} = 1 \Rightarrow F \in MDA(\Psi_1).$$

Normující konstanty můžeme zvolit jako $c_n = x_F - F^{-1}(1 - \frac{1}{n}) = 1 - 1 + \frac{1}{n} = \frac{1}{n}$ a $d_n = x_F = 1$. Pak platí:

$$(M_n - 1)n \rightarrow \Psi_1.$$

2.6 Maximum domain of attraction Gumbelovy distribuční funkce

$MDA(\Lambda)$ obsahuje širokou škálu distribučních funkcí. Bohužel, zde neexistuje jednoduchá charakterizace $MDA(\Lambda)$ pomocí $R_{-\alpha}$ jako u Fréchetovy a Weibullové distribuční funkce. Podle Taylorova rozvoje platí $1 - \Lambda(x) = \bar{F}(x) \sim e^{-x}, x \rightarrow \infty$. Otázkou je, jak moc se může chvost distribuční funkce F asymptoticky lišit od e^{-x} , aby F patřila do $MDA(\Lambda)$. Platí, že $MDA(\Lambda)$ obsahuje distribuční funkce s různě těžkými chvosty. Obsahuje rozdělení se středně těžkými chvosty jako je logaritmicko-normální rozdělení až po rozdělení s lehkými chvosty jako je normální rozdělení. Navíc v $MDA(\Lambda)$ mohou být rozdělení s $x_F = \infty$ tak také s $x_F < \infty$. Následující popis distribučních funkcí nám pomůže charakterizovat $MDA(\Lambda)$.

Definice 2.5 (Von Miseho funkce). *Nechť F je distribuční funkce, která je spojitá v $x_F \leq \infty$. Pokud existuje konstanta $z < x_F$ taková, že F lze*

reprezentovat jako:

$$\bar{F}(x) = c \exp\left\{-\int_z^x \frac{1}{a(t)} dt\right\}, z < x < x_F,$$

kde c je nějaká kladná konstanta, $a(\cdot)$ je kladná a absolutně spojitá funkce (vzhledem k Lebesgueově míře) s hustotou a' a $\lim_{x \uparrow x_F} a'(x) = 0$.

Potom se F nazývá von Miseho funkce, funkce $a(\cdot)$ se nazývá pomocná funkce F .

Dále pro nás bude $\bar{F}(x) \in R_{-\infty}$ znamenat

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(xt)}{\bar{F}(x)} = \begin{cases} 0, & t > 1 \\ \infty, & 0 < t < 1. \end{cases}$$

Tvrzení 2.10 (Vlastnosti von Miseho funkce). *Každá von Miseho funkce F je absolutně spojitá na (z, x_F) s kladnou hustotou f . Pomocná funkce může být volena jako $a(x) = \frac{\bar{F}(x)}{f(x)}$. Navíc platí:*

a) pokud $x_F = \infty$, potom $\bar{F} \in R_{-\infty}$ a

$$\lim_{x \rightarrow \infty} \frac{xf(x)}{\bar{F}(x)} = \infty.$$

b) pokud $x_F < \infty$, potom $\bar{F}(x_F - \frac{1}{x}) \in R_{-\infty}$ a

$$\lim_{x \uparrow x_F} \frac{(x_F - x)f(x)}{\bar{F}(x)} = \infty.$$

Důkaz tohoto tvrzení je uveden v [2] na straně 141. Platí, že každá von Miseho funkce je prvkem $MDA(\Lambda)$. Normující konstanty můžeme volit jako $d_n = F^\leftarrow(1 - \frac{1}{n})$ a $c_n = a(d_n)$, kde $a(\cdot)$ je pomocná funkce F . Von Miseho funkce však nejsou veškeré distribuční funkce, které patří do $MDA(\Lambda)$. Pro kompletní charakterizaci $MDA(\Lambda)$ musíme k von Miseho funkci přidat i všechny distribuční funkce s nimi chvostově ekvivalentní. Tedy máme kompletní charakterizaci $MDA(\Lambda)$:

$MDA(\Lambda)$ se skládá z von Miseho funkcí a z distribučních funkcí s nimi chvostově ekvivalentních.

Narozdíl od $MDA(\Phi_\alpha)$ platí pro $F \in MDA(\Lambda)$ s $x_F = \infty$, že $F \in R_{-\infty}$, a tedy $E(X^+)^{\alpha} < \infty$ pro všechna $\alpha > 0$, kde $X^+ = \max(X, 0)$. Příklady rozdělení, které patří do $MDA(\Lambda)$, jsou například normální rozdělení, logaritmicko-normální rozdělení a exponenciální rozdělení. Definice a vlastnosti těchto rozdělení lze najít v [1].

Příklad 2.3 (Exponenciální rozdělení).

exponenciální rozdělení má hustotu $f(x) = \lambda \exp(-\lambda x), x > 0, \lambda > 0$

distribuční funkce je $F(x) = 1 - \exp(-\lambda x), x > 0, \lambda > 0$

tedy platí, že chvost má tvar $\bar{F}(x) = \exp(-\lambda x), x > 0, \lambda > 0$ a F je von Miseho funkce s $a(\cdot) = \frac{1}{\lambda}, \lambda > 0$.

2.7 Obecná extremální distribuční funkce a obecná Paretova distribuční funkce

Pro další potřeby zde uvedeme obecnou extremální distribuční funkci, která pomocí jednoho parametru ξ sjednotí předchozí extremální distribuční funkce do jedné. Pro parametr ξ používáme následující interpretaci:

$\xi = \frac{1}{\alpha} > 0$ odpovídá Fréchetově distribuční funkci Φ_α

$\xi = 0$ odpovídá Gumbelově distribuční funkci Λ

$\xi = -\frac{1}{\alpha} < 0$ odpovídá Weibullově distribuční funkci Ψ_α .

Potom obecná extremální distribuční funkce má tvar:

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

kde $1 + \xi x > 0$. Později budeme používat obecnou extremální distribuční funkci $H_{\xi, \mu, \psi}$, kde nahradíme proměnnou x v H_ξ pomocí $\frac{x-\mu}{\psi}$. Následující tvrzení charakterizuje $MDA(H_\xi)$. Navíc je základem pro některé statistické aplikace teorie extremálních rozdělení. Náznak jeho důkazu je uveden v [2] na straně 159.

Tvrzení 2.11 (Charakterizace $MDA(H_\xi)$).

Pro $\xi \in \mathbb{R}$ jsou následující tvrzení ekvivalentní:

a) $F \in MDA(H_\xi)$

b) Existuje kladná měřitelná funkce $a(\cdot)$ taková, že pro $1 + \xi x > 0$ platí:

$$\lim_{u \uparrow x_F} \frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \begin{cases} (1 + \xi x)^{-\xi^{-1}}, & \xi \neq 0 \\ e^{-x}, & \xi = 0 \end{cases}$$

c) Pro $x, y > 0, y \neq 1$ platí:

$$\lim_{s \rightarrow \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} = \begin{cases} \frac{x^\xi - 1}{y^\xi - 1}, & \xi \neq 0 \\ \frac{\ln(x)}{\ln(y)}, & \xi = 0 \end{cases}$$

kde $U(t) = \bar{F}(1 - \frac{1}{t}), t > 0$.

Pokud je X náhodná veličina s distribuční funkcí $F \in MDA(H_\xi)$, potom můžeme bod b) z předcházejícího tvrzení interpretovat jako:

$$\lim_{u \uparrow x_F} P\left(\frac{X - u}{a(u)} > x \mid X > u\right) = \begin{cases} (1 + \xi x)^{\frac{-1}{\xi}}, & \xi \neq 0 \\ e^{-x}, & \xi = 0. \end{cases}$$

Tento vztah nám dává distribuční approximaci pro normovaný přesah přes vysoký práh u s normativní funkcí $a(u)$. Tato approximace je důležitá pro měření rizika pro finanční instrumenty a motivuje nás k následujícím úvahám. Nejdříve si definujeme distribuční funkci přesahu a funkci středního přesahu.

Definice 2.6 (Distribuční funkce přesahu, funkce středního přesahu). *Nechť X je náhodná veličina s distribuční funkcí F a hraničním bodem x_F . Pro pevné $u < x_F$ se*

$$F_u(x) = P(X - u \leq x \mid X > u), x \geq 0$$

nazývá distribuční funkce přesahu náhodné veličiny X přes práh u . Funkce

$$e(u) = E(X - u \mid X > u)$$

se nazývá funkce středního přesahu náhodné veličiny X .

Funkce $e(u)$ se také někdy nazývá střední zbytková doba života (mean residual lifetime) a hraje důležitou roli ve statistických aplikacích teorie extremálních hodnot. Pomocí grafu empirické $e(u)$ lze posoudit, zda naměřená data pocházejí z rozdělení s těžkým chvostem. Tento graf se také používá při odhadu parametrů obecné extremální distribuční funkce $H_{\xi,\mu,\psi}$. V neposlední řadě slouží tato funkce k výpočtu tzv. očekávané extrémní ztráty (expected shortfall), která se používá k měření rizika finančních instrumentů. Další vlastnosti této funkce představíme v kapitole Míry rizika a podrobněji je popsána v [2].

Dále si označíme limitní výsledek z Tvrzení 2.11 b) pro přípustná x jako obecnou Paretovu distribuční funkci.

Definice 2.7 (Obecná Paretova distribuční funkce(GPD)).

Distribuční funkci:

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

kde $x > 0$ pro $\xi \geq 0$ a $0 \leq x \leq \frac{-1}{\xi}$ pro $\xi < 0$. Budeme nazývat obecná Paretova distribuční funkce.

Někdy budeme místo $G_\xi(x)$ používat značení $G_{\xi,\nu,\beta}$, kde nahradíme argument x pomocí $\frac{x-\nu}{\beta}$, $\nu \in \mathbb{R}$, $\beta > 0$. Obecná Paretova distribuční funkce je limitní distribuční funkce pro normovaný přesah přes vysoký práh, odhadem jejích parametrů se budeme zabývat v kapitole Parametrické a neparametrické odhady charakteristik rozdělení. Důležitou roli ve statistických aplikacích hraje distribuční funkce $G_{\xi,0,\beta}$. Následující tvrzení shrnuje základní pravděpodobnostní vlastnosti $G_{\xi,0,\beta}$ a jeho důkaz lze najít v [2] na straně 166.

Tvrzení 2.12 (Vlastnosti $G_{\xi,0,\beta}$).

a) Nechť X má $G_{\xi,0,\beta}$. Potom $EX < \infty$ právě tehdy, když je $\xi < 1$. V tomto případě platí:

$$E(1 + \frac{\xi}{\beta}X)^{-r} = \frac{1}{1 + \xi r}, r > \frac{-1}{\xi}$$

$$E(\ln(1 + \frac{\xi}{\beta}X))^k = \xi^k k!, k \in \mathbb{N}$$

$$EX(\overline{G_{\xi,0,\beta}}(X))^r = \frac{\beta}{(r+1-\xi)(r+1)}, \frac{r+1}{|\xi|} > 0.$$

Pokud $\xi < \frac{1}{r}$ pro $r \in \mathbb{N}$, potom:

$$EX^r = \frac{\beta^r}{\xi^{r+1}} \frac{\Gamma(\xi^{-1} - r)}{\Gamma(1 + \xi^{-1})} r!.$$

b) Pro každé $\xi \in \mathbb{R}$, $F \in MDA(H_\xi)$ právě když

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

pro nějakou kladnou funkci β .

c) Nechť $x_i \in D(\xi, \beta)$, $i = 1, 2$, kde

$D(\xi, \beta) = [0, \infty)$, $\xi \geq 0$ a $D(\xi, \beta) = [0, -\frac{-\beta}{\xi})$, $\xi < 0$ potom platí:

$$\frac{\overline{G}_{\xi,0,\beta}(x_1 + x_2)}{\overline{G}_{\xi,0,\beta}(x_1)} = \overline{G}_{\xi,0,\beta+\xi x_1}(x_2).$$

d) Nechť N je náhodná veličina s rozdělením $Po(\lambda)$ a nezávislá s iid posloupností (X_n) , kde X_n má distribuční funkci $G_{\xi,0,\beta}$ a nechť $M_N = \max(X_1, \dots, X_N)$. Potom platí:

$$P(M_N \leq x) = \exp\left\{-\lambda\left(1 + \xi \frac{x}{\beta}\right)^{\frac{-1}{\xi}}\right\} = H_{\xi, \mu, \psi(x)},$$

kde $\mu = \frac{\beta(\lambda^\xi - 1)}{\xi}$ a $\psi = \beta \lambda^\xi$.

e) Nechť X má distribuční funkci $G_{\xi,0,\beta}$, kde $\xi < 1$. Potom pro $u < x_F$ platí:

$$e(u) = E(X - u \mid X > u) = \frac{\beta + \xi u}{1 - \xi}, \beta + \xi u > 0.$$

Vlastnosti $G_{\xi,0,\beta}$ z tohoto tvrzení jsou základem statistických aplikací extremálních hodnot.

Vlastnost b) nám říká, že $G_{\xi,0,\beta}$ approximuje distribuční funkci přesahu $F_u(x)$, pro dostatečně vysoký prah u . Funkce β je odhadnuta pomocí naměřených dat, tedy platí:

$$\bar{F}_u(x) = P(X - u > x | X > u) \approx \bar{G}_{\xi,0,\beta(u)}(x), x > 0$$

nebo ekvivaletně pro $x > u$,

$$P(X > x | X > u) \approx \bar{G}_{\xi,u,\beta(u)}(x).$$

Návod na určení dostatečně vysokého prahu u nám dává bod e). Podle tohoto tvrzení je $e(u)$ z $G_{\xi,0,\beta}$ lineární. Pomocí konstrukce empirické $e_n(u)$, kterou popíšeme podrobně v příští kapitole získáme grafickou metodu na odhad dostatečně vysokého prahu u . Budeme vybírat taková u , pro která bude $e_n(u)$ zhruba lineární.

Kapitola 3

Parametrické a neparametrické odhady charakteristik extremálních rozdělení

3.1 Úvod

V této kapitole se budeme zabývat aplikací teorie extremálních hodnot na reálná data. Nejdříve popíšeme některé metody pro grafickou analýzu naměřených dat. Pomocí nich budeme testovat, zda daná data pocházejí z určitého rozdělení, a jestli se jedná o data z rozdělení s těžkým chvostem. Dále bude naším cílem stanovit odhady vysokých kvantilů \hat{x}_p a odhad chvostu $\hat{F}(x)$ pro náhodný výběr X_1, \dots, X_n z nějakého rozdělení. Metody pro tyto odhady rozdělíme do dvou skupin. Parametrické odhady, které budeme provádět pomocí metody maximální věrohodnosti a metody pravděpodobnostně vážených momentů a neparametrické odhady, které budeme provádět pomocí metody maximum domain of attraction a pomocí obecné Paretovy distribuční funkce. V celé této kapitole budeme používat značení

$$H_{\xi,\mu,\psi}(x) = \exp\left\{-\left(1 + \xi \frac{x - \mu}{\psi}\right)^{\left(\frac{-1}{\xi}\right)}\right\}, 1 + \xi \frac{x - \mu}{\psi} > 0$$

pro obecnou extremální distribuční funkci. Pro případ $\xi = 0$ odpovídá $H_{0,\mu,\psi}(x) = \exp\left\{-e^{\frac{-(x-\mu)}{\psi}}\right\}$, $x \in \mathbb{R}$. Gumbelově distribuční funkci. Pro zjednodušení zápisu zavedeme parametr $\theta = (\xi, \mu, \psi) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+$.

3.2 Metody grafické analýzy dat

Kvantilové grafy

Tato metoda se zabývá následujícím problémem: Najít distribuční funkci F , která dobře odpovídá rozdělení, z kterého pocházejí naměřená iid data X_1, \dots, X_n . Základem této metody je následující úvaha. Pro náhodný výběr X_1, \dots, X_n z rozdělení se spojitou distribuční funkcí F platí, že náhodné veličiny $U_i = F(X_i)$, $i = 1 \dots n$ jsou iid s rovnoramenným rozdělením na $(0,1)$. Navíc pro uspořádaný náhodný výběr $X_{n,n} \leq \dots \leq X_{1,n}$ platí

$F(X_{k,n}) \stackrel{d}{=} U_{k,n}$, $k = 1 \dots n$. Z toho vyplývá, že $EF(X_{k,n}) = \frac{n-k+1}{n+1}$, $k = 1 \dots n$. Tato úvaha nás přivádí na myšlenku vykreslit graf:

$$\{(X_{n,k}, F^{-1}\left(\frac{n-k+1}{n+1}\right)) : k = 1 \dots n\}.$$

Tento graf se nazývá kvantilový graf. Existují i jiné varianty tohoto grafu, které jsou typu:

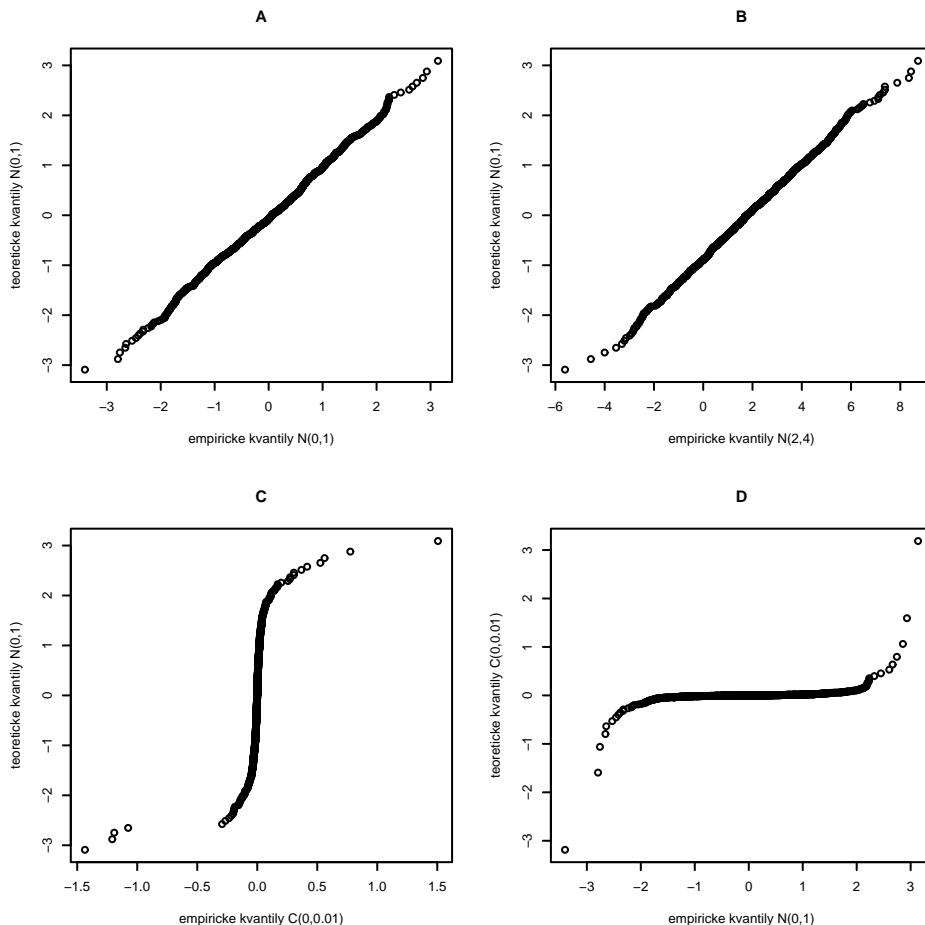
$$\{(X_{n,k}, F^{-1}(p_{k,n})) : k = 1 \dots n\},$$

kde $p_{k,n} = \frac{n-k+\delta_k}{n+\gamma_k}$ pro nějaké konstanty (δ_k, γ_k) . Často se volí například $(\delta_k = 0.5, \gamma_k = 0)$. Pokud naše data pocházejí z rozdělení s distribuční funkcí F potom by měl být kvantilový graf zhruba lineární. Pokud data pocházejí ze třídy distribučních funkcí $F\left(\frac{x-\mu}{\psi}\right)$, kde μ je parametr polohy a ψ je parametr měřítka, potom by měl být kvantilový graf také zhruba lineární s absolutním členem μ a směrnicí ψ . Tyto parametry lze odvodit pomocí lineární regrese. Pokud data pocházejí z rozdělení s těžším chvostem než má distribuční funkce F , potom se bude kvantilový graf stáčet nahoru vlevo a dolů vpravo. Naopak pokud má F těžší chrost, bude se kvantilový graf stáčet dolů vlevo a nahoru vpravo. Pokud budeme chtít testovat, zda naše data pocházejí z $H_{\xi, \mu, \psi}$, budeme nejdříve muset odhadnout parametr ξ pomocí metod popsaných dále.

Funkce středního přesahu

V minulé kapitole jsme definovali funkci středního přesahu jako

$$e(u) = E(X - u \mid X > u), 0 \leq u \leq x_F.$$



Obrázek 3.1: A-kvantilový graf simulovaných dat z normálního rozdělení se střední hodnotou 0 a rozptylem 1. B-kvantilový graf simulovaných dat z normálního rozdělení se střední hodnotou 2 a rozptylem 4. C-kvantilový graf simulovaných dat z Cauchyho rozdělení s parametry 0 a 0.01. Všechny proti kvantilům normálního rozdělení se střední hodnotou 0 a rozptylem 1. D-kvantilový graf simulovaných dat z normálního rozdělení se střední hodnotou 0 a rozptylem 1, proti kvantilům Cauchyho rozdělení s parametry 0 a 0.01.

Pokud je X kladná náhodná veličina s distribuční funkcí F a konečnou střední hodnotou potom:

$$e(u) = \frac{1}{\bar{F}(u)} \int_u^{x_F} \bar{F}(x) dx, 0 \leq u \leq x_F.$$

V minulé kapitole jsme ukázali, že pro obecnou Paretovu distribuční funkci je $e(u)$ lineární s kladnou směrnicí. Snadno se ukáže, že pro náhodnou veličinu X s exponenciálním rozdělením s parametrem λ je $e(u) = \frac{1}{\lambda}$ tedy konstantní. Pro distribuční funkce s lehkým chvostem typu

$\bar{F}(x) \sim \exp\{-x^a\}, a > 1$ klesá $e(u)$ k nule. Pro náhodnou veličinu X s těžkým chvostem $e(u)$ typicky roste k nekonečnu. Její tvar je většinou mezi konstantou (pro exponenciální rozdělení) a lineární funkci s kladnou směrnicí (pro Paretovo rozdělení).

Náš grafický test bude založen na empirické funkci středního přesahu $e_n(u)$. Tento test nám pomůže odhalit, zda naměřená data pocházejí z rozdělení s těžkým chvostem. Nechť X_1, \dots, X_n jsou iid náhodné veličiny s distribuční funkcí F . Empirickou distribuční funkci budeme značit F_n a

$\Delta_n(u) = \{i : i = 1, \dots, n, X_i > u\}$, potom

$$e_n(u) = \frac{1}{\bar{F}(u)} \int_u^{\infty} \bar{F}_n(y) dy = \frac{1}{\text{card } \Delta_n(u)} \sum_{i \in \Delta_n(u)} (X_i - u), u \geq 0.$$

Graf střední hodnoty přesahu má tvar:

$$\{(X_{k,n}; e_n(X_{k,n})) : k = 1, \dots, n\}.$$

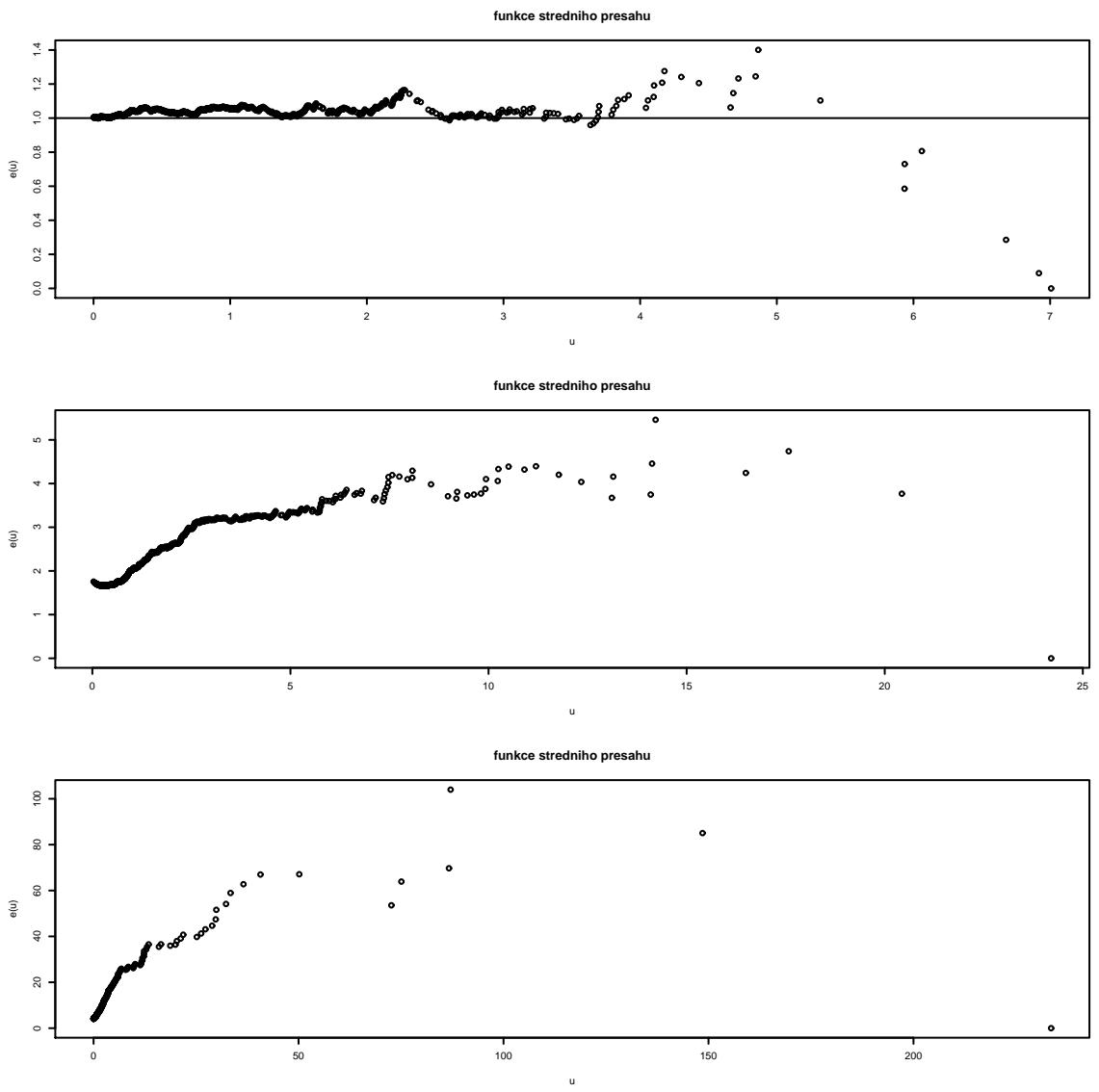
Poměr maxima a součtu

Tato grafická metoda nám dává jednoduchý nástroj na odhad těžkosti chvostu a odhad řádu konečných momentů. Opět předpokládáme, že X_1, X_2, \dots jsou iid náhodné veličiny a definujeme hodnoty:

$$S_n(p) = |X_1|^p + \dots + |X_n|^p, M_n(p) = \max(|X_1|^p, \dots, |X_n|^p), n \geq 1.$$

Dále definujeme $R_n(p) = \frac{M_n(p)}{S_n(p)}$, $n \geq 1, p \geq 0$. Pro takto definované $R_n(p)$ platí:

$$R_n(p) \xrightarrow{s.j.} 0 \Leftrightarrow E |X|^p < \infty.$$



Obrázek 3.2: Graf funkce středního přesahu simulovaných dat. Nahoře exponenciální rozdělení s parametrem 1, uprostřed logaritmicko-normální rozdělení s parametry 0 a 1, dole pravý chvost Cauchyho rozdělení s parametry 1 a 0.

Pro aplikaci na naměřená data vykreslíme $R_n(p)$ pro různé hodnoty p . Pokud $E |X|^p < \infty$, potom by mělo být $R_n(p)$ malé pro vysoké hodnoty n . Naopak pokud vykazuje graf $R_n(p)$ na pravé straně odchylku od nuly, značí to, že $E |X|^p = \infty$.

3.3 Parametrické odhady charakteristik rozdělení

V této kapitole budeme provádět parametrické odhady pro náhodný výběr X_1, \dots, X_n z H_θ . Protože nelze předpokládat, že všechna naměřená data budou mít přesně distribuční funkci H_θ , provedeme nejdříve transformaci náhodného výběru. Rozdělíme si naměřená data do s rozměrných vektorů. Tyto vektory symbolizují data naměřená v časových intervalech délky s . Potom mají data tvar :

$$\begin{aligned}\mathbf{X}^{(1)} &= (X_1^{(1)}, \dots, X_s^{(1)}) \\ \mathbf{X}^{(2)} &= (X_1^{(2)}, \dots, X_s^{(2)}) \\ &\vdots \\ \mathbf{X}^{(n)} &= (X_1^{(n)}, \dots, X_s^{(n)}).\end{aligned}$$

Předpokládáme, že vektory $\mathbf{X}^{(i)}$ jsou navzájem nezávislé, avšak uvnitř každého vektoru $\mathbf{X}^{(i)}$ mohou být data závislá. Náš nový výběr z H_θ se poté bude skládat z $X_i = \max(X_1^{(i)}, \dots, X_s^{(i)})$.

Metoda maximální věrohodnosti

Předpokládáme, že X_1, \dots, X_n je náhodný výběr z rozdělení s distribuční funkcí H_θ . Tedy toto rozdělení je absolutně spojité s hustotou h_θ . V tomto případě je možné odhadnout parametr θ pomocí metody maximální věrohodnosti. Definujeme funkci:

$$L(\theta, X) = \prod_{i=1}^n h_\theta(X_i) I_{\{1 + \xi \frac{X_i - \mu}{\psi} > 0\}}.$$

Dále označíme $l(\theta, X) = \ln(L(\theta, X))$, potom odhad pomocí metody maximální věrohodnosti pro parametr θ se rovná

$$\widehat{\theta}_n = \arg \max_{\theta \in \Theta} l(\theta, X).$$

Maximum určíme zderivováním $l(\theta, X)$ podle parametrů ξ, μ, ψ a položíme vzniklé výrazy rovny nule. Bohužel neexistuje explicitní vyjádření parametrů z těchto rovnic, proto musíme k jejich vyřešení použít numerické metody.

Metoda pravděpodobnostně vážených momentů

Tato metoda je založená na porovnávání pravděpodobnostně vážených momentů náhodné veličiny X , která má distribuční funkci H_θ a pravděpodobnostně vážených empirických momentů. Definujeme

$$w_r(\theta) = E(XH_\theta^r(X)), r \in N_0,$$

kde X má distribuční funkci H_θ s parametrem $\theta = (\xi, \mu, \psi)$. Pokud je $\xi \geq 1$, potom je $w_0 = EX = \infty$. Proto se omezíme pouze na případy, kdy je $\xi < 1$. Dále definujeme empirický $\widehat{w}_r(\theta)$.

$$\widehat{w}_r(\theta) = \int_{-\infty}^{+\infty} xH_\theta^r(x) dF_n(x), r \in N_0,$$

kde F_n je empirická distribuční funkce odpovídající datům X_1, \dots, X_n . Abychom určili θ vyřešíme rovnice

$$w_r(\theta) = \widehat{w}_r(\theta), r = 0, 1, 2.$$

Po dosazení dostáváme

$$\widehat{w}_r(\theta) = \frac{1}{n} \sum_{j=1}^n X_{j,n} H_\theta^r(X_{j,n}) = \frac{1}{n} \sum_{j=1}^n X_{j,n} U_{j,n}^r, r = 0, 1, 2,$$

kde $U_{j,n}$ je uspořádaný náhodný výběr z iid posloupnosti X_1, \dots, X_n s rovnoměrným rozdělením na $(0,1)$. $U_{j,n}^r$ může být nahrazeno $EU_{j,n}^r$. Pro výpočet $w_r(\theta)$ použijeme vzorec:

$$w_r(\theta) = \int_{-\infty}^{+\infty} xH_\theta^r(x) dH_\theta(x) = \int_0^1 H_\theta^\leftarrow(y) y^r dy$$

kde pro $0 < y < 1$

$$H_\theta^\leftarrow(y) = \begin{cases} \mu - \frac{\psi}{\xi}(1 - (-\ln y)^{-\xi}), & \xi \neq 0 \\ \mu - \psi \ln(-\ln y), & \xi = 0. \end{cases}$$

Pro $\xi < 1$ a $\xi \neq 0$ vyjde

$$w_r(\theta) = \frac{1}{1+r} \left\{ \mu - \frac{\psi}{\xi} (1 - \Gamma(1-\xi)(1+r)^\xi) \right\},$$

kde Γ značí gamma funkci $\Gamma(t) = \int_0^\infty e^{-u} u^{t-1} du, t > 0$. Dále dostáváme:

$$w_0(\theta) = \mu - \frac{\psi}{\xi} (1 - \Gamma(1-\xi)),$$

$$2w_1(\theta) - w_0(\theta) = \frac{\psi}{\xi} \Gamma(1-\xi)(2^\xi - 1),$$

$$3w_2(\theta) - w_0(\theta) = \frac{\psi}{\xi} \Gamma(1-\xi)(3^\xi - 1),$$

a tedy:

$$\frac{3w_2(\theta) - w_0(\theta)}{2w_1(\theta) - w_0(\theta)} = \frac{3^\xi - 1}{2^\xi - 1}.$$

Dosazením empirických $\hat{w}_r(\theta)$ získáme z poslední rovnosti odhad $\hat{\xi}$. Poté dopočteme zbylé parametry pomocí ztahů:

$$\hat{\psi} = \frac{(2\hat{w}_1 - \hat{w}_0)\hat{\xi}}{\Gamma(1-\hat{\xi})(2^{\hat{\xi}} - 1)},$$

$$\hat{\mu} = \hat{w}_0 + \frac{\hat{\psi}}{\hat{\xi}} (1 - \Gamma(1-\hat{\xi})).$$

3.4 Odhadý pomocí maximum domain of attraction

V této kapitole budeme předpokládat, že máme náhodný výběr X_1, \dots, X_n s distribuční funkcí $F \in MDA(H_\xi), \xi \in \mathbb{R}$. Oproti minulé kapitole tedy nepředpokládáme, že F je přesně H_ξ . Předpoklad $F \in MDA(H_\xi)$ je obecnější, a tudíž lze použít na širší škálu dat. Budeme se zabývat dvěma hlavními body. Nejdříve si představíme tři metody pro odhad parametru $\xi \in \mathbb{R}$, popíšeme jejich statistické vlastnosti a porovnáme je. Poté odvodíme odhad pro normující konstanty c_n, d_n . Výsledky použijeme k

odhadu vysokých kvantilů x_p , často mimo dosah naměřených dat. Tedy pro $p > 1 - \frac{1}{n}$. A také pro odhad chvostu $\bar{F}(x)$ pro vysoké hodnoty x .

Podle Tvrzení 2.6 $F \in MDA(H_\xi)$, právě když

$$\lim_{n \rightarrow \infty} n\bar{F}(c_n x + d_n) = -\ln(H_\xi(x)),$$

tedy pro velká $u = c_n x + d_n$ platí:

$$n\bar{F}(u) \approx (1 + \xi \frac{u - d_n}{c_n})^{\frac{-1}{\xi}},$$

tedy odhad chvostu má tvar:

$$\widehat{\bar{F}(u)} = \frac{1}{n} \left(1 + \widehat{\xi} \frac{u - \widehat{d}_n}{\widehat{c}_n}\right)^{\frac{-1}{\widehat{\xi}}},$$

pro odpovídající odhadu $\widehat{\xi}, \widehat{c}_n, \widehat{d}_n$. Protože vlastnost $F \in MDA(H_\xi)$ je vlastností chvostu $\bar{F}(x)$ bude odhad parametru ξ založen na k největších pozorování uspořádáného náhodného výběru $X_{k,n} \leq \dots \leq X_{1,n}$. Určení vhodného k závisí na rozsahu náhodného výběru n a mělo by splňovat následující vlastnosti:

$$a) k(n) \rightarrow \infty$$

$$b) \frac{n}{k(n)} \rightarrow \infty.$$

Na volbě $k(n)$ také závisí statistické vlastnosti odhadu pro parametr ξ , které rozebereme později. Ze vzorce pro odhad chvostu můžeme také odvodit odhad kvantilu $x_p = F^\leftarrow(p)$:

$$\widehat{x}_p = \widehat{d}_n + \frac{\widehat{c}_n}{\widehat{\xi}} ((n(1-p))^{-\widehat{\xi}} - 1).$$

Protože hodnoty c_n, d_n jsou definovány pomocí kvantilů $F^\leftarrow(1 - \frac{1}{n})$, a tedy jejich určení je ekvivalentní s odhadem kvantilů x_p na hranici našich dat, použijeme při odhadu kvantilů následující trik. Budeme předpokládat, že $\frac{n}{k} \in \mathbb{N}$. Při odhadu x_p se posuneme k podposloupnosti $\frac{n}{k}$, kde $k = k(n)$ splňuje podmínky a) a b). Odhad kvantilu poté bude mít tvar:

$$\widehat{x}_p = \widehat{d}_{\frac{n}{k}} + \frac{\widehat{c}_{\frac{n}{k}}}{\widehat{\xi}} \left(\left(\frac{n}{k}(1-p)\right)^{-\widehat{\xi}} - 1\right).$$

Pro statistické účely pro nás může být zajímavé určit následující hodnoty:

$$x_{p,r} = F^{-1}(p^{\frac{1}{r}}), r \in \mathbb{N}.$$

Interpretace $x_{p,r}$ je následující:

$$p = F^r(x_{p,r}) = P(\max(X_{n+1}, \dots, X_{n+r}) \leq x_{p,r}),$$

tedy $x_{p,r}$ je hodnota, která s danou pravděpodobností p , nebude v následujících r pozorováních překonána. Odhad $x_{p,r}$ má tvar:

$$\widehat{x}_{p,r} = \widehat{d}_{\frac{n}{k}} + \frac{\widehat{c}_{\frac{n}{k}}}{\widehat{\xi}} \left(\left(\frac{n}{k} (1 - p^{\frac{1}{r}}) \right)^{-\widehat{\xi}} - 1 \right).$$

Tedy pokud máme odvozen odhad pro \widehat{x}_p můžeme jednoduše odhadnout i $\widehat{x}_{p,r}$.

Odhady parametru ξ

1. Pickandův odhad pro $\xi \in \mathbb{R}$

Pickandův odhad je odvozen od Tvrzení 2.11 c) a má tvar:

$$\widehat{\xi}_{k,n}^{(P)} = \frac{1}{\ln(2)} \ln \left(\frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}} \right).$$

Tvrzení 3.1 (Vlastnosti Pickanova odhadu).

Nechť (X_n) je iid posloupnost náhodných veličin s distribuční funkcí $F \in MDA(H_\xi)$, $\xi \in \mathbb{R}$. Nechť $\widehat{\xi}^{(P)} = \widehat{\xi}_{k,n}^{(P)}$ je Pickandův odhad, potom platí:
a) (slabá konzistence) Pokud $k \rightarrow \infty, \frac{k}{n} \rightarrow 0$, pro $n \rightarrow \infty$, potom:

$$\widehat{\xi}^{(P)} \xrightarrow{P} \xi, n \rightarrow \infty.$$

b) (silná konzistence) Pokud $\frac{k}{n} \rightarrow 0, \frac{k}{\ln(\ln(n))} \rightarrow \infty$, pro $n \rightarrow \infty$, potom:

$$\widehat{\xi}^{(P)} \xrightarrow{s.j.} \xi, n \rightarrow \infty.$$

Toto tvrzení je uvedeno v [2] na straně 328. Pro výběr optimálního $k(n)$ neexistuje obecně nejlepší řešení. Tento výběr budeme provádět pomocí grafu $\{(k, \widehat{\xi}_{k,n}^{(P)}) : k = 1, \dots, n\}$. Intuitivně bychom měli vybrat $\widehat{\xi}_{k,n}^{(P)}$ z takových hodnot k , kde je graf zhruba horizontální a k je vysoké, odpovídající podmínkám a), b).

2. Hillův odhad pro $\xi = \frac{1}{\alpha} > 0$

Nechť X_1, \dots, X_n je náhodný výběr s distribuční funkcí $F \in MDA(\Phi_\alpha)$, $\alpha > 0$ tedy $\bar{F}(x) = x^{-\alpha}L(x)$, pro nějakou pomalu se měnící funkci L, potom Hillův odhad parametru α má tvar:

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

kde $k = k(n)$ je zvoleno v závislosti na n , podobně jako u Pickandova odhadu. Distribuční funkce z $MDA(\Phi_\alpha)$, $\alpha > 0$ jsou hlavními příklady distribučních funkcí s těžkými chvosty. Empirické studie chvostů denních logaritmických návratů ve financích odhalují nejčastěji hodnoty α mezi 3 a 4. Tedy pro modelování ve financích lze právě takovéto chvosti očekávat. Hillův odhad lze odvodit několika různými způsoby, což dokazuje, že je to odhad velmi přirozený. Jedna z metod, pomocí kterých lze Hillův odhad odvodit se zakládá na metodě maximální věrohodnosti. Toto odvození je založeno na k nejvyšších náhodných veličinách uspořádného náhodného výběru. Pomocí tohoto postupu lze odvodit i odhad pro chvost $\bar{F}(x)$ pro vysoká x a odhad pro vysoké kvantily x_p . Tyto odhady mají následující tvar:

$$\begin{aligned}\widehat{\bar{F}}(x) &= \frac{k}{n} \left(\frac{x}{X_{k,n}} \right)^{-\widehat{\alpha}_{k,n}^{(H)}} \\ \widehat{x}_p &= \left(\frac{n}{k} (1-p) \right)^{-1/\widehat{\alpha}_{k,n}^{(H)}} X_{k,n}.\end{aligned}$$

Pomocí těchto odhadů můžeme odvodit odhad pro distribuční funkci přesahu $F_u(x-u)$, $x \geq u$ pomocí $F_u(x-u) = 1 - \frac{\bar{F}(x)}{\bar{F}(u)}$.

Vlastnosti Hillova odhadu shrnuje následující tvrzení:

Tvrzení 3.2 (Vlastnosti Hillova odhadu).

Nechť (X_n) je iid posloupnost náhodných veličin s distribuční funkcí F splňující pro nějaké $\alpha > 0$ a $L \in R_0$,

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

Nechť $\widehat{\alpha}^{(H)} = \widehat{\alpha}_{k,n}^{(H)}$ je Hillův odhad parametru α , potom platí:
 a) (slabá konzistence) Pokud $k \rightarrow \infty, \frac{k}{n} \rightarrow 0$, pro $n \rightarrow \infty$, potom:

$$\widehat{\alpha}^{(H)} \xrightarrow{P} \alpha.$$

b) (silná konzistence) Pokud $\frac{k}{n} \rightarrow 0, \frac{k}{\ln(\ln(n))} \rightarrow \infty$, pro $n \rightarrow \infty$, potom:

$$\widehat{\alpha}^{(H)} \xrightarrow{s.j.} \alpha.$$

V předchozím tvrzení lze formulovat slabou konzistenci obecněji pro striktně stacionární procesy s dalšími podmínkami na slabou závislost namísto předpokladu iid. V obecnější formě lze toto tvrzení nalézt v [2] na straně 337.

3. Dekkers-Einmahl-de Haanův odhad pro $\xi \in \mathbb{R}$

Dekkers-Einmahl-de Haanův odhad je rozšíření Hillova odhadu na celou třídu $H_\xi, \xi \in \mathbb{R}$. Pro zjednodušení budeme předpokládat, že $x_F > 0$. Potom Dekkers-Einmahl-de Haanův odhad má tvar:

$$\widehat{\xi} = 1 + H_n^{(1)} + \frac{1}{2} \left(\frac{(H_n^{(1)})^2}{H_n^{(2)}} - 1 \right)^{-1},$$

kde

$$H_n^{(1)} = \frac{1}{k} \sum_{j=1}^k (\ln(X_{j,n}) - \ln(X_{k+1,n}))$$

$$H_n^{(2)} = \frac{1}{k} \sum_{j=1}^k (\ln(X_{j,n}) - \ln(X_{k+1,n}))^2.$$

Tyto výrazy mají pro $\frac{k}{n} \rightarrow \infty$ smysl protože $X_{k,n} \xrightarrow{s.j.} x_F > 0$.

Porovnání odhadů pro parametr ξ

Otázka, jaký odhad pro parametr ξ bychom měli použít, nemá jednoznačnou odpověď. Nejlepší volba odhadu závisí na vlastnostech distribuční funkce F . Pro $\xi = \frac{1}{\alpha} > 0$ a pro F splňující další vlastnosti, které jsou však těžko ověřitelné v praxi na reálných datech, má Hillův odhad nejmenší čtvercovou

odchylku. Pro obecné $\xi \in \mathbb{R}$ lze použít pouze Pickandův a Dekkers-Einmahl-de Haanův odhad. Přičemž pro $\xi > -2$ má Dekkers-Einmahl-de Haanův odhad menší rozptyl, než Pickandův odhad.

Odhad normujících konstant

Zaměříme se pouze na odhad normujících konstant pro náhodný výběr s distribuční funkcí $F \in MDA(H_\xi, \xi \geq 0)$. Tato třída distribučních funkcí pokrývá většinu případů ve financích a je spojením Fréchetovy a Gumbelovy MDA. Následující tvrzení je odvozeno v [2] na straně 346 a ukazuje, jak můžeme tyto dvě MDA sjednotit.

Tvrzení 3.3 (Sjednocení $MDA(\Phi_\alpha)$ a $MDA(\Lambda)$).

Nechť X_1, \dots, X_n je náhodný výběr s distribuční funkcí $F \in MDA(H_\xi)$, $\xi \geq 0$ a $x_F = \infty$ a normujícími konstantami $c_n > 0, d_n \in \mathbb{R}$. Potom X_1^*, \dots, X_n^* jsou iid s distribuční funkcí $F^* \in MDA(\Lambda)$ a pomocnou funkcí

$$a^* = \int_t^\infty \frac{\bar{F}^*(y)}{\bar{F}^*(t)} dy.$$

Normující konstanty můžeme zvolit jako:

$$d_n^* = (F^*)^{-1}(1 - \frac{1}{n}),$$

$$c_n^* = a^*(d_n^*) = \int_{d_n^*}^\infty \frac{\bar{F}^*(y)}{\bar{F}^*(d_n^*)} dy \sim n \int_{d_n^*}^\infty \bar{F}^*(y) dy.$$

Kde $x^* = \ln(1 \vee x), x \in \mathbb{R}$.

Nahrazením F^* pomocí empirické distribuční funkce F_n^* získáme odhady:

$$\hat{d}_{\frac{n}{k}}^* = X_{k+1,n}^*$$

$$\hat{c}_{\frac{n}{k}}^* = \frac{1}{k} \sum_{j=1}^k \ln(X_{j,n}) - \ln(X_{k+1,n}).$$

Přechodem zpět k distribuční funkci F lze odvodit odhady pro chvost a kvantily:

$$\widehat{\bar{F}(x)} = \frac{k}{n} \left(\frac{x}{X_{k+1,n}} \right)^{1/\hat{c}_{\frac{n}{k}}^*}$$

$$\widehat{x}_p = \left(\frac{n}{k}(1-p)\right)^{-\widehat{c}_{\frac{n}{k}}^*} X_{k+1,n}.$$

Tyto odhady jsme již dříve odvodili pomocí Hillova odhadu pro $\xi > 0$. Přechod od k k $k+1$ je asymptoticky nevýznamný.

3.5 Odhad chvostů pomocí obecné Paretovy distribuční funkce

V této kapitole budeme opět předpokládat, že X_1, \dots, X_n je náhodný výběr s distribuční funkcí $F \in MDA(H_\xi)$, $\xi \in \mathbb{R}$. Budeme se snažit najít odhad chvostu distribuční funkce F a odhady vysokých kvantilů pomocí obecné Paretovy distribuční funkce. K tomu použijeme následující postup. Nejdříve si zvolíme vysoký práh u a označíme

$$N_u = \text{card}\{i : i = 1, \dots, n, X_i > u\}$$

jako počet překonání u v X_1, \dots, X_n . Odpovídající přesahy označíme Y_1, \dots, Y_{N_u} . Připomeňme, že distribuční funkce přesahu X přes práh u je tvaru:

$$F_u(y) = P(X - u \leq y \mid X > u) = P(Y \leq y \mid X > u), y \geq 0.$$

Chvost distribuční funkce F můžeme popsat pomocí $\bar{F}_u(y)$ následujícím způsobem:

$$\bar{F}(u+y) = \bar{F}(u)\bar{F}_u(y).$$

Tedy odhad chvostu distribuční funkce $\bar{F}(x)$ pro vysoké hodnoty $x > u$ můžeme popsat pomocí odhadu nepodmíněného chvostu $\bar{F}(u)$ a podmíněného chvostu $\bar{F}_u(x-u)$. Přirozený odhad nepodmíněného chvostu $\bar{F}(u)$ je dán pomocí empirické distribuční funkce:

$$\widehat{\bar{F}}(u) = \bar{F}_n(u) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i > u\}} = \frac{N_u}{n}.$$

Odhad podmíněného chvostu $\bar{F}_u(x-u)$ odvodíme pomocí Tvrzení 2.12 b). Tedy pro u vysoké platí

$$\bar{F}_u(y) \approx \bar{G}_{\xi, 0, \beta(u)}(y).$$

Kde $\beta(u)$ je funkce práhu u . Máme-li stanoven vysoký práh u parametry $\xi, \beta(u)$ odhadneme z naměřených dat, která překonají práh u . Tedy tyto odhady závisí na daném u . Odhad $\widehat{F}_u(y)$ bude mít tvar:

$$\widehat{F}_u(y) = \widehat{G}_{\widehat{\xi}, 0, \widehat{\beta}}(y)$$

pro odpovídající odhady $\widehat{\xi} = \widehat{\xi}_{N_u} \widehat{\beta} = \widehat{\beta}_{N_u}$. Výsledný odhad pro chvost $\widehat{F}(u + y), y > 0$ má tvar:

$$\widehat{F}(u + y) = \frac{N_u}{n} \left(1 + \widehat{\xi} \frac{y}{\widehat{\beta}}\right)^{1/\widehat{\xi}}.$$

Odhad kvantilu má tvar:

$$\widehat{x}_p = u + \frac{\widehat{\beta}}{\widehat{\xi}} \left(\left(\frac{n}{N_u}(1-p)\right)^{-\widehat{\xi}} - 1\right).$$

Pro $\widehat{\xi} < 0$ odvodíme odhad pro hraniční bod x_F dosazením do odhadu pro $x_p, p = 1$:

$$\widehat{x}_F = u - \frac{\widehat{\beta}}{\widehat{\xi}}.$$

Pro zvolení optimálního práhu u použijeme graf empirické funkce středního přesahu

$$e_n(u) = \frac{1}{N_u} \sum_{i \in \Delta_n(u)} (X_i - u), u > 0,$$

kde $\Delta_n(u) = \{i : i = 1, \dots, n, X_i > u\}$. Podle Tvrzení 2.12 e) platí pro náhodnou veličinu X s distribuční funkcí $G_{\xi, 0, \beta}$:

$$e(u) = E(X - u \mid X > u) = \frac{\beta + \xi u}{1 - \xi}, \beta + \xi u > 0, \xi < 1.$$

A tedy $e(u)$ je lineární. Návod na zvolení optimálního prahu u podle předchozího zní:

Zvolme $u > 0$ tak, že $e_n(x)$ je zhruba lineární pro $x \geq u$.

Pro odhad parametrů ξ, β použijeme metodu maximální věrohodnosti: Předpokládejme, že X_1, \dots, X_n jsou iid náhodné veličiny s distribuční funkcí $F = G_{\xi, 0, \beta}$, potom hustota f má tvar

$$f(x) = \frac{1}{\beta} \left(1 + \xi \frac{x}{\beta}\right)^{-\frac{1}{\xi}-1}, x \in D(\xi, \beta).$$

Logaritmická rovnice maximální věrohodnosti má tvar

$$l((\xi, \beta); \mathbf{X}) = -n \ln \beta - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^n \ln \left(1 + \frac{\xi}{\beta} X_i\right).$$

V našem případě budeme předpokládat, že distribuční funkci $G_{\xi,0,\beta}$ mají přesahy Y_1, \dots, Y_{N_u} . Výsledné věrohodnostní rovnice vyřešíme pomocí reparametrisace $(\xi, \beta) \rightarrow (\xi, \tau)$, kde $\tau = -\frac{\xi}{\beta}$. Vyjde nám řešení:

$$\widehat{\xi} = \frac{1}{N_u} \sum_{i=1}^{N_u} \ln(1 - \tau Y_i),$$

$$\frac{1}{\widehat{\tau}} + \frac{1}{N_u} \left(\frac{1}{\widehat{\xi}} + 1\right) \sum_{i=1}^{N_u} \frac{Y_i}{1 - \widehat{\tau} Y_i} = 0.$$

Další informace o aplikaci obecné Paretovy distribuční funkce na odhad chvostu a míry rizika lze nalézt v [3].

Kapitola 4

Míry rizik a jejich vlastnosti

V této kapitole budeme opět předpokládat, že X_1, X_2, \dots jsou iid náhodné veličiny s neznámou distribuční funkcí F . Tato data budou reprezentovat určitá rizika, která se dají interpretovat mnoha různými způsoby podle potřeby. V tomto případě budeme předpokládat, že jsou to denní ztráty z nějakého finančního aktiva nebo portfolia a tedy F je distribuční funkce denní ztráty. Měřením rizika budeme rozumět ohodnocení distribuční funkce F číslem, které představuje míru rizika. Nejjednoduššími charakteristikami rizika pro naměřená data X_1, \dots, X_n mohou být například výběrový průměr nebo výběrový rozptyl. Tato čísla určitým způsobem charakterizují riziko, avšak neposkytují nám mnoho informací o riziku extrémních jevů. Extrémní jevy rozumíme vzácné avšak devastující ztráty. V našem modelu se extrémní jevy objeví, pokud náhodná veličina X_i nabude hodnoty z chvostu distribuční funkce F . Protože extrémní jevy jsou charakterizovány pomocí chvostu distribuční funkce F , je teorie extremálních rozdělení nejlepší způsob pro jejich modelování. Pro měření rizika extrémních jevů budeme používat dvě míry rizika:

- 1) Value-at-Risk, kterou budeme značit VaR
- 2) Expected shortfall, kterou budeme značit ES.

4.1 Value-at-Risk - VaR

VaR na hladině $\alpha \in (0, 1)$ se definuje jako:

$$VaR_\alpha = F^\leftarrow(\alpha),$$

kde F je distribuční funkce ztráty. Většinou se hodnota α volí $\alpha = 0,95$ nebo $\alpha = 0,99$. Potom mluvíme o 95%-ním(99%-ním) Value-at-Risk. VaR_α je tedy horní hranice ztráty, která bude překonána pouze s malou pravděpodobností $1 - \alpha$. V předchozí kapitole jsme si představili několik způsobů jak pomocí naměřených dat, která splňují určité předpoklady, vytvořit odhad pro vysoké kvantily distribuční funkce F . Zde napíšeme stručné shrnutí:

- 1) Pomocí parametrických odhadů, kde předpokládáme, že X_1, \dots, X_n je náhodný výběr s distribuční funkcí H_θ , odhadneme parametry $\hat{\xi}, \hat{\mu}, \hat{\psi}$. Potom odhad $VaR_\alpha, \alpha \in (0, 1)$ získáme invertováním distribuční funkce H_θ .

$$\widehat{VaR}_\alpha = \widehat{x}_\alpha = \hat{\mu} - \frac{\hat{\psi}}{\hat{\xi}}(1 - (-\ln \alpha)^{-\hat{\xi}}), \alpha \in (0, 1).$$

- 2) Pomocí metody Maximum domain of attraction, kde předpokládáme, že X_1, \dots, X_n je náhodný výběr s distribuční funkcí $F \in MDA(H_\xi), \xi \geq 0$. Potom pro dobře zvolené $k = k(n)$ získáme odhad $VaR_\alpha, \alpha \in (0, 1)$ jako:

$$\widehat{VaR}_\alpha = \widehat{x}_\alpha = \left(\frac{n}{k}(1 - \alpha)\right)^{-\hat{c}_n^*} X_{k+1, n}, \alpha \in (0, 1).$$

- 3) Pomocí obecné Paretovy distribuční funkce, kde předpokládáme, že X_1, \dots, X_n je náhodný výběr s distribuční funkcí $F \in MDA(H_\xi), \xi \in \mathbb{R}$. Potom získáme odhad $VaR_\alpha, \alpha \in (0, 1)$ jako:

$$\widehat{VaR}_\alpha = \widehat{x}_\alpha = u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{n}{N_u}(1 - \alpha)\right)^{-\hat{\xi}} - 1\right), \alpha \in (0, 1).$$

Typický závěr při určení $VaR_\alpha, \alpha \in (0, 1)$ zní: S pravděpodobností $1 - \alpha$ bude ztráta větší než VaR_α . Nevýhoda hodnoty v riziku je, že nám neříká nic o nominální velikosti ztráty, která VaR_α překoná.

4.2 Expected shortfall - ES

Expected shortfall nám dává odpověď na otázku: Pokud ztráta překročí VaR, jak velká by tato ztráta měla být? Tedy ES udává očekávanou velikost

ztráty, která překoná VaR. Definujeme jí jako:

$$ES_\alpha = E[X \mid X > VaR_\alpha], \alpha \in (0, 1).$$

Z tohoto vzorce je vidět úzký vztah mezi ES a funkcí středního přesahu $e(u) = E(X - u \mid X > u), 0 \leq u \leq x_F$. Tento vztah můžeme popsat jako:

$$ES_\alpha = e(VaR_\alpha) + VaR_\alpha.$$

Pomocí odhadu VaR_α a odhadu chvostu distribuční funkce $F \in MDA(H_\xi), \xi < 1$ můžeme odhadnout ES_α pro nezápornou náhodnou veličinu X pomocí vzorce:

$$\widehat{ES}_\alpha = \frac{1}{\overline{F}(VaR_\alpha)} \int_{VaR_\alpha}^{\infty} \overline{F}(x) dx + VaR_\alpha = \frac{1}{1 - \alpha} \int_{VaR_\alpha}^{\infty} \overline{F}(x) dx + VaR_\alpha.$$

Navíc pomocí obecné Paretovy distribuční funkce pro $F \in MDA(H_\xi), \xi < 1$ získáme odhad pro ES jako:

$$\widehat{ES}_\alpha = \frac{\widehat{VaR}_\alpha}{1 - \widehat{\xi}} + \frac{\widehat{\beta} - \widehat{\xi}u}{1 - \widehat{\xi}},$$

kde u je vysoký práh z Kapitoly 3.5. Více informací o aplikaci obecné Paretovy distribuční funkce na odhad míry rizika lze nalézt v [3].

Kapitola 5

Aplikace na reálná data

5.1 Úvod

V této kapitole aplikujeme teorii z předcházejících kapitol na data z finančních trhů. Konkrétně provedeme analýzu ceny akcií Bank of America Corporation (BAC), které jsou součástí indexu Dow Jones Industrial Average (DJIA). V době, kdy tato práce vznikla, tedy v létě roku 2009 celým světem otřásla finanční krize, která způsobila turbulence nejen na akciových trzích po celém světě. Proto provedeme analýzu dvou souborů dat. Nejdříve analyzujeme historická data akcií BAC do začátku krize. Poté provedeme analýzu historických cen akcií BAC až do současnosti, tedy včetně dat z období krize. Konkrétně budeme analyzovat denní logaritmické ztráty, které se spočítají následujícím vzorcem:

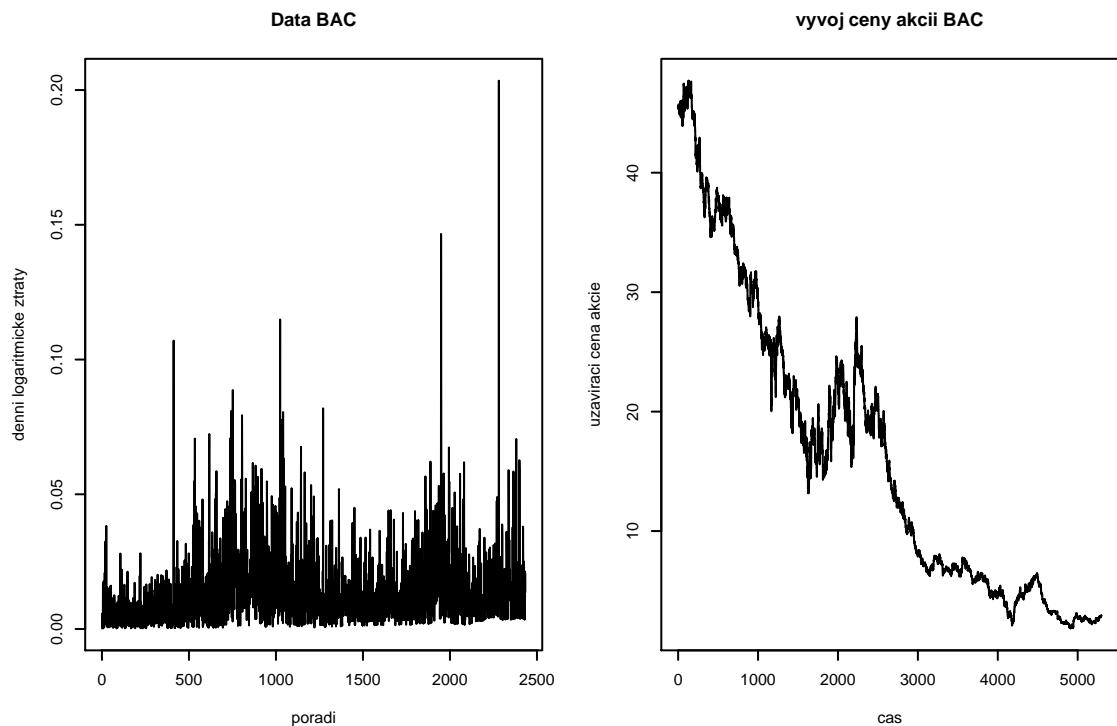
$$\ln(\text{close}(i)) - \ln(\text{close}(i-1)) = \ln\left(\frac{\text{close}(i)}{\text{close}(i-1)}\right),$$

kde $\text{close}(i)$ značí uzavírací cenu v den i . Protože nás budou zajímat ztráty budeme brát pouze $\ln(\text{close}(i)) - \ln(\text{close}(i-1)) < 0$. Bude nás zajímat, jak se změní chvost rozdělení, pokud přidáme data z období krize a také, jak velké ztráty v období krize nastaly vzhledem k předchozímu období.

5.2 Analýza historických cen akcií Bank of America Corporation

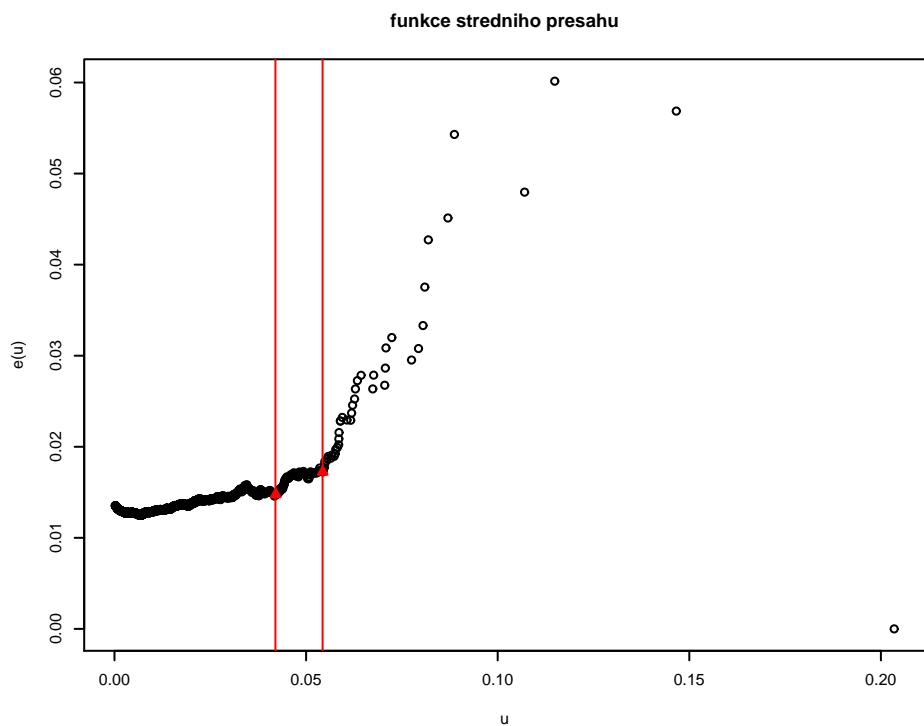
Analýza dat před začátkem finanční krize

Analýza denních očištěných (o dividendy a splity) logaritmických ztrát z akcií BAC z období 29.5.1986-29.5.2007. Výběr obsahuje 5296 denních návratů z toho 2433 ztrát. Zdroj dat:finance.yahoo.com. Na Obrázku 5.1 jsou zobrazeny denní logaritmické ztráty a vývoj ceny akcií BAC z tohoto období.



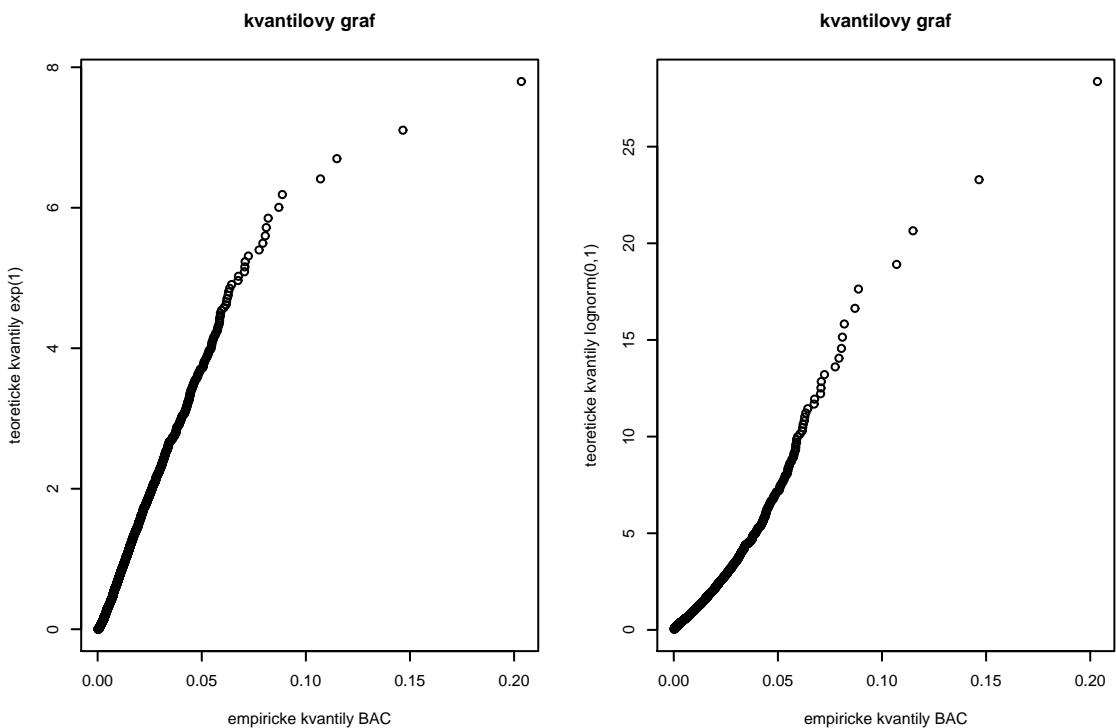
Obrázek 5.1: Vlevo jsou zobrazeny denní logaritmické ztráty z akcií BAC. Vpravo je graf vývoje ceny akcií BAC.

Pro odhad tvaru chvostu použijeme některé metody z kapitoly 3.2. Nejdříve si nakreslíme graf funkce středního přesahu.



Obrázek 5.2: Funkce středního přesahu- $e(u)$ akcií BAC. Zvýrazněny jsou hodnoty $u=0.042$ a $u=0.054$, v nichž směrnice grafu vzroste

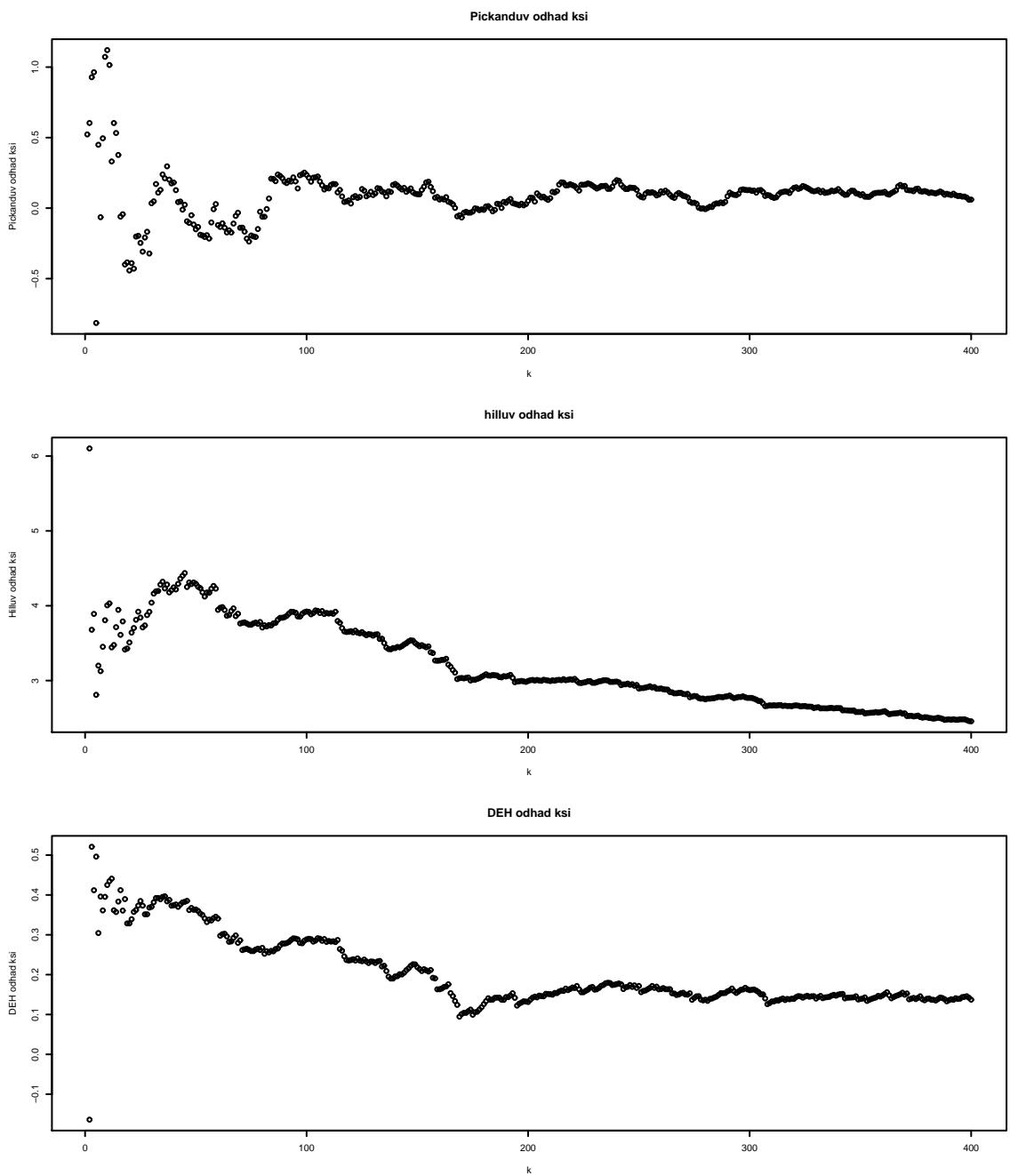
Graf funkce středního přesahu ukazuje změnu směrnice v oblasti $u = 0.042$ a $u = 0.054$. V těchto bodech výrazně vzroste směrnice a naznačuje těžký chvost rozdělení denních očištěných logaritmických ztrát. Podle tohoto grafu a podle faktu, že denní logaritmické ztráty jsou neomezené zprava můžeme usoudit, že naše rozdělení by mohlo patřit do $MDA(\Phi_\alpha)$. Další odhad tvaru chvostu dostaneme pomocí kvantilového grafu. Vykreslíme kvantily logaritmických ztrát proti kvantilům exponenciálního rozdělení s parametrem 1 a proti kvantilům logaritmicko-normálního rozdělení s parametry 0, 1.



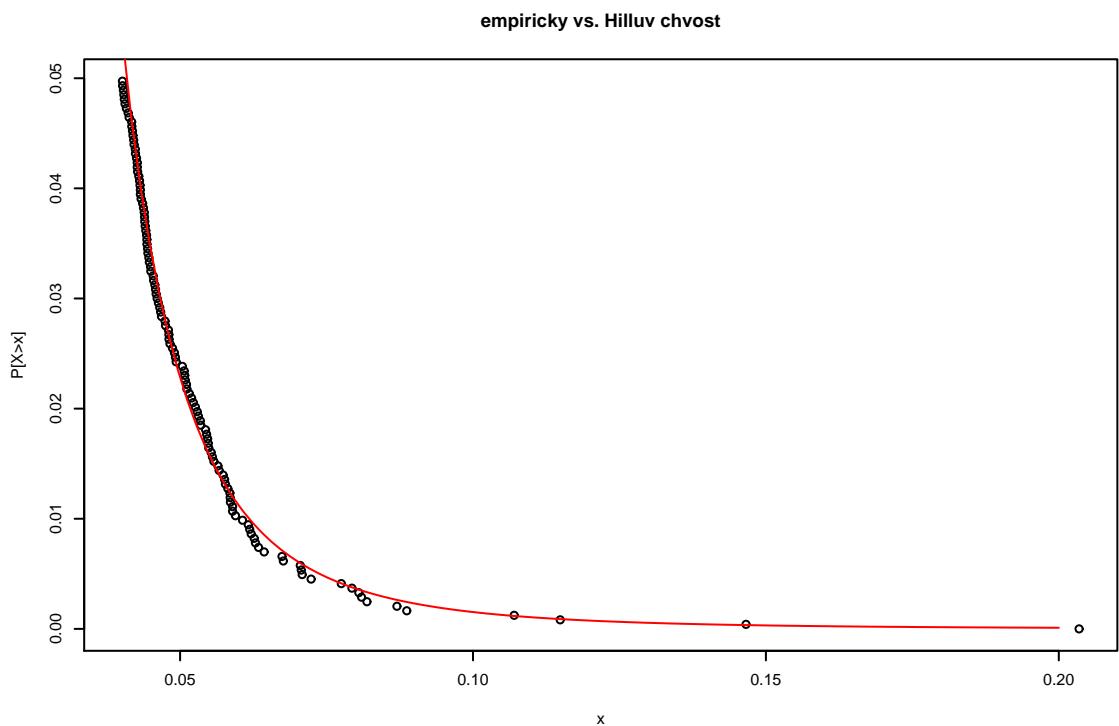
Obrázek 5.3: kvantilový graf akcií BAC. Vlevo jsou empirické kvantily vykresleny proti kvantilům exponenciálního rozdělení. Vpravo proti kvantilům logaritmicko-normálního rozdělení.

Graf vlevo ukazuje, že exponenciální rozdělení se velmi dobře shoduje s rozdělením denních logaritmických ztrát akcií BAC až na tvar pravého chvostu. Ten mají námi naměřená data výrazně těžší než má exponenciální rozdělení, což ukazuje změna tvaru grafu. Pravý horní roh pravého grafu dokonce ukazuje, že naše rozdělení má těžší pravý chrost než logaritmicko-normální rozdělení. Tedy celkově kvantilové grafy potvrzují naši hypotézu, že rozdělení denních logaritmických ztrát akcií BAC patří do $MDA(\Phi_\alpha)$. Dále budeme pokračovat odhadem parametru ξ v $MDA(H_\xi)$. Tento odhad provedeme pomocí vykreslení Pickandova, Hillova a DEH grafu - Obrázek 5.4. Protože naše dosavadní analýza naznačuje, že naše data pocházejí z rozdělení, které patří do $MDA(\Phi_\alpha)$, můžeme k odhadu parametru ξ použít i Hillův odhad. Navíc protože denní logaritmické ztráty jsou obecně neomezené zprava, můžeme z našeho uvažování vyloučit možnost $\xi < 0$.

Pickandův odhad ukazuje hodnotu ξ lehce nad nulou. V oblasti $k = 80 - 95$ odhaduje $\xi = 0.22$, druhá možnost se nabízí v oblasti $k = 210 - 250$, kde vychází $\xi = 0.16$. Protože jsou v případě, že je $\xi > 0$ přesnější Hillův a DEH odhad než Pickandův odhad, použijeme tato čísla hlavně jako utvrzení, že naše rozdělení patří do $MDA(\Phi_\alpha)$. DEH odhad nabízí opět několik možností pro hodnotu ξ . V oblasti $k = 95 - 115$ vychází hodnota $\xi = 0.29$. V oblasti $k = 115 - 135$ odhaduje hodnotu $\xi = 0.24$. A pokud bychom brali hodnoty $k > 200$ vyšlo by $\xi = 0.16$. Pokud se podíváme na Hillův odhad, nabízí se nám opět několik možností hodnoty α . Pro oblast $k = 90 - 115$ odpovídá hodnota $\alpha = 3.9 (\xi = 0.26)$. Pro oblast $k = 115 - 130$ odpovídá hodnota $\alpha = 3.6 (\xi = 0.28)$ a pro $k = 170 - 230$ odpovídá hodnota $\alpha = 3 (\xi = 0.33)$. Odhadu nám nabízí několik možností volby ξ . Protože v případě $\xi > 0$ je nejlepší použít Hillův odhad, zvolíme podle něj $k = 100$ a $\alpha = 3.9 (\xi = 0.26)$. Toto číslo je poměrně konzistentní i s DEH odhadem v podobné oblasti. Pomocí Hillova odhadu můžeme stanovit kvantily denních logaritmických ztrát pro vysoké hodnoty. Tímto způsobem určíme 95%, 99%, 99.9% a 99.99% kvantily tedy i VaR. Vyjde nám $\widehat{VaR}_{0.95} = 0.041$, $\widehat{VaR}_{0.99} = 0.062$, $\widehat{VaR}_{0.999} = 0.112$ a $\widehat{VaR}_{0.9999} = 0.201$. Pomocí Hillova odhadu můžeme také odhadnout chrost rozdělení. Na Obrázku 5.5 uvidíme, že chrost odvozený podle Hillova odhadu velmi dobře kopíruje empirický chrost našich dat.



Obrázek 5.4: Odhady parametru ξ pro denní logaritmické ztráty z akcií BAC.
Nahoře-Pickandův odhad parametru ξ . Uprostřed-Hillův odhad parametru α . Dole-DEH odhad parametru ξ .



Obrázek 5.5: Tečkovaně je vyznačen empirický chvost, červená čára značí odhad chvostu podle Hillova odhadu.

Pomocí toho odhadu můžeme vypočítat expected shortfall pro naše data. Výsledky jsou následující: $\widehat{ES}_{0.95} = 0.055$, $\widehat{ES}_{0.99} = 0.083$, $\widehat{ES}_{0.999} = 0.15$, $\widehat{ES}_{0.9999} = 0.271$. Druhou metodou, kterou můžeme použít k odhadu VaR, ES a chvostu rozdělení, je pomocí GPD. V této metodě se nejdříve vrátíme k Obrázku 5.2. Podle předchozí teorie si nejdříve musíme zvolit dostatečně velký práh, po kterém je graf $e(u)$ zhruba lineární. Pro tento účel se nám nabízejí dvě možnosti. Můžeme zvolit buďto $u = 0.042$, což vyústí v 109 přesahů, nebo zvolit hladinu $u = 0.053$, což vyústí v 48 přesahů. Protože při našem rozsahu dat 2433 tvoří 48 přesahů pouze necelé dvě procenta dat, zvolíme raději hladinu $u = 0.042$. Potom 109 přesahů tvoří téměř pět procent našich naměřených dat. Pomocí metody maximální věrohodnosti odhadneme parametry GPD $\xi = 0.27$, $\beta = 0.01$, což je konzistentní s naším odhadem parametru $\xi = 0.26$. Také odhady VaR a ES vycházejí velmi podobně. Srovnání odhadů pomocí Hillova odhadu a GPD nabízí následující tabulky.

druh	$\widehat{VaR}_{0.95}$	$\widehat{VaR}_{0.99}$	$\widehat{VaR}_{0.999}$	$\widehat{VaR}_{0.9999}$
Hill	0.041	0.062	0.112	0.201
GDP	NA	0.06	0.108	0.197

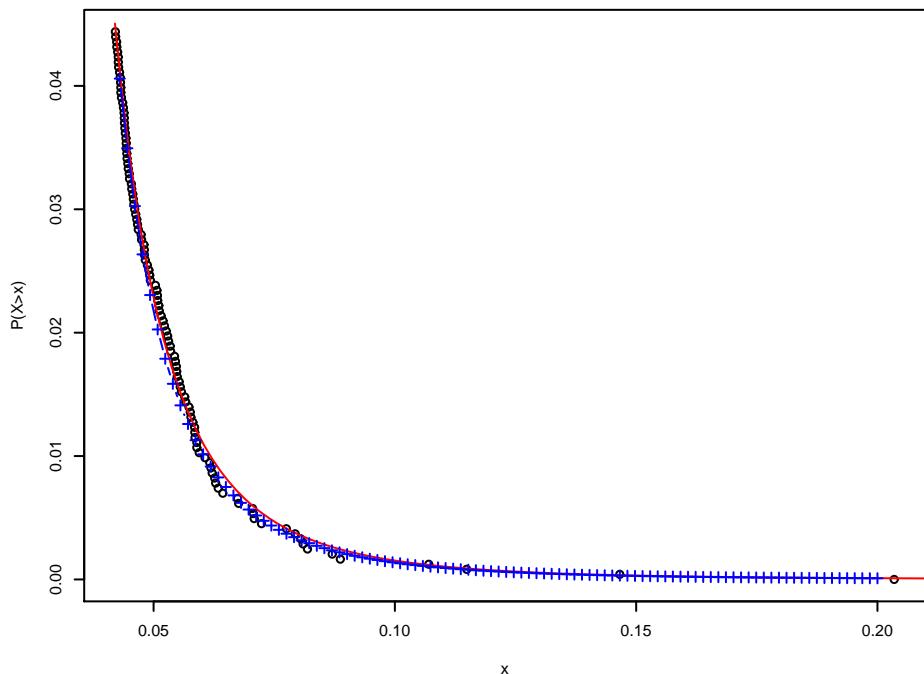
druh	$\widehat{ES}_{0.95}$	$\widehat{ES}_{0.99}$	$\widehat{ES}_{0.999}$	$\widehat{ES}_{0.9999}$
Hill	0.055	0.083	0.15	0.271
GDP	NA	0.081	0.147	0.269

Porovnání odhadu chvostu pomocí GDP a pomocí Hillova odhadu s empirickým chvostem ukazuje Obrázek 5.6.

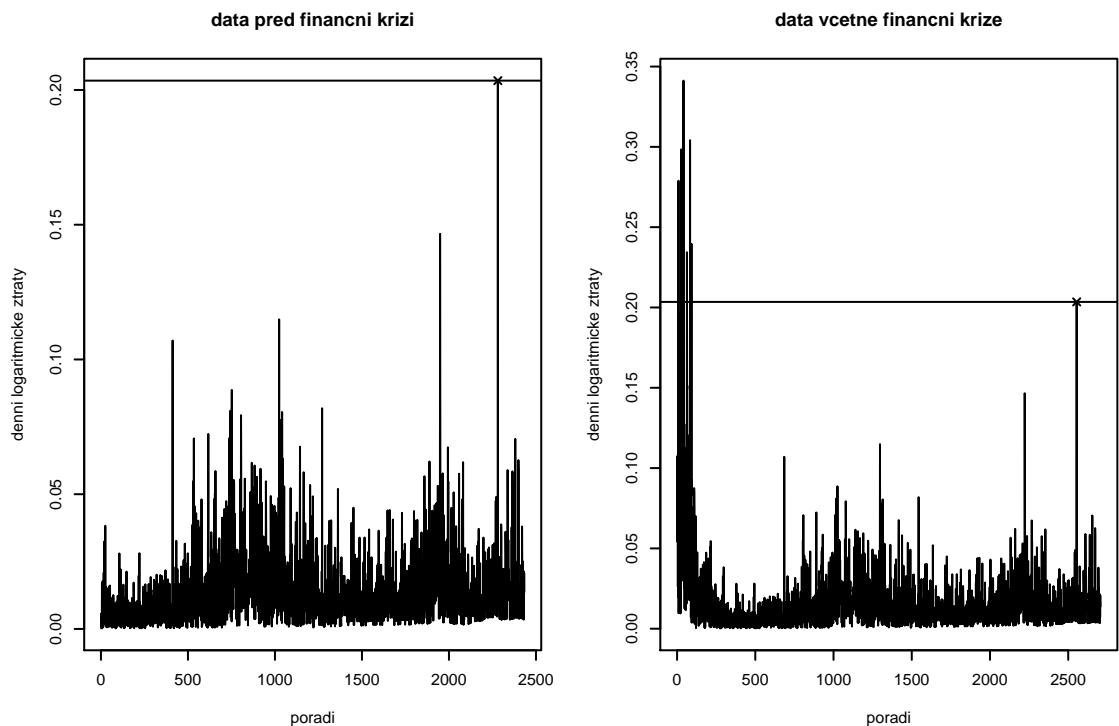
Analýza dat včetně období finanční krize

Nyní do našeho výběru dat přidáme denní očištěné logaritmické ztráty z období finanční krize. Tedy budeme analyzovat data z období od 29.5.1986 do 15.5.2009. Tento výběr obsahuje 5792 denních logaritmických návratů, z toho 2706 ztrát. Bude nás zajímat, jak data z období krize změní chvost našeho rozdělení a jak ovlivní naše předchozí odhady. Také nás bude zajímat, jak velké denní logaritmické ztráty utrpěla BAC v období finanční krize, vzhledem k námi odhadnutým míram rizika. Sílu dopadu krize na akcie BAC odhalí Obrázek 5.7, kde porovnáváme data z období před finanční krizí a data včetně finanční krize.

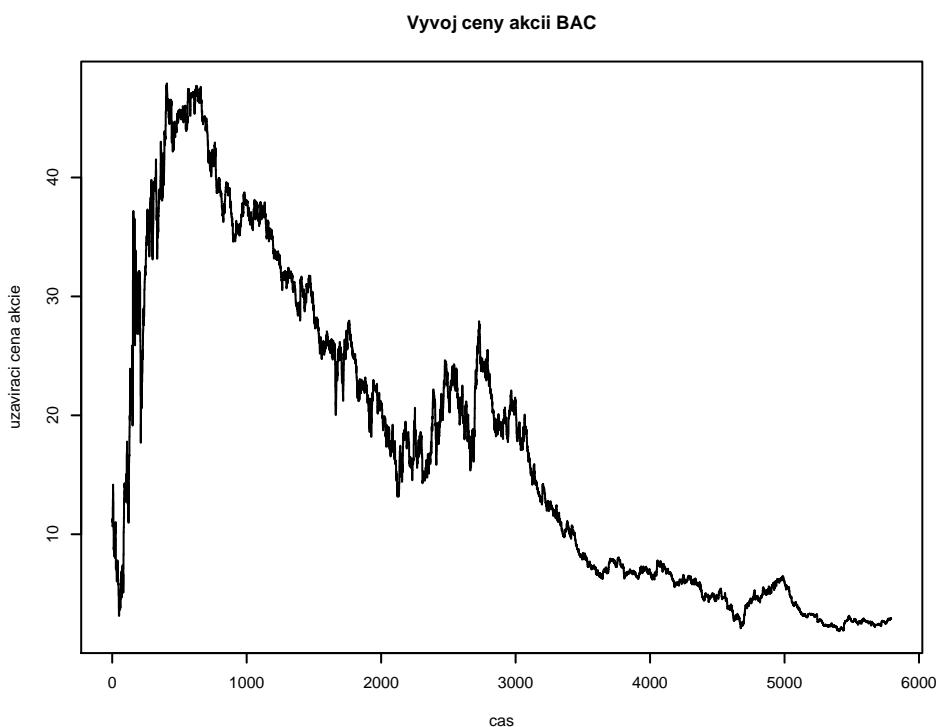
Porovnání chvostu



Obrázek 5.6: Porovnání odhadu chvostů denních logaritmických ztrát z akcií BAC. Tečkovaně je vyznačen empirický chvost, červená čára značí odhad chvostu podle Hillova odhadu, modré křížky značí odhad chvostu podle GPD.

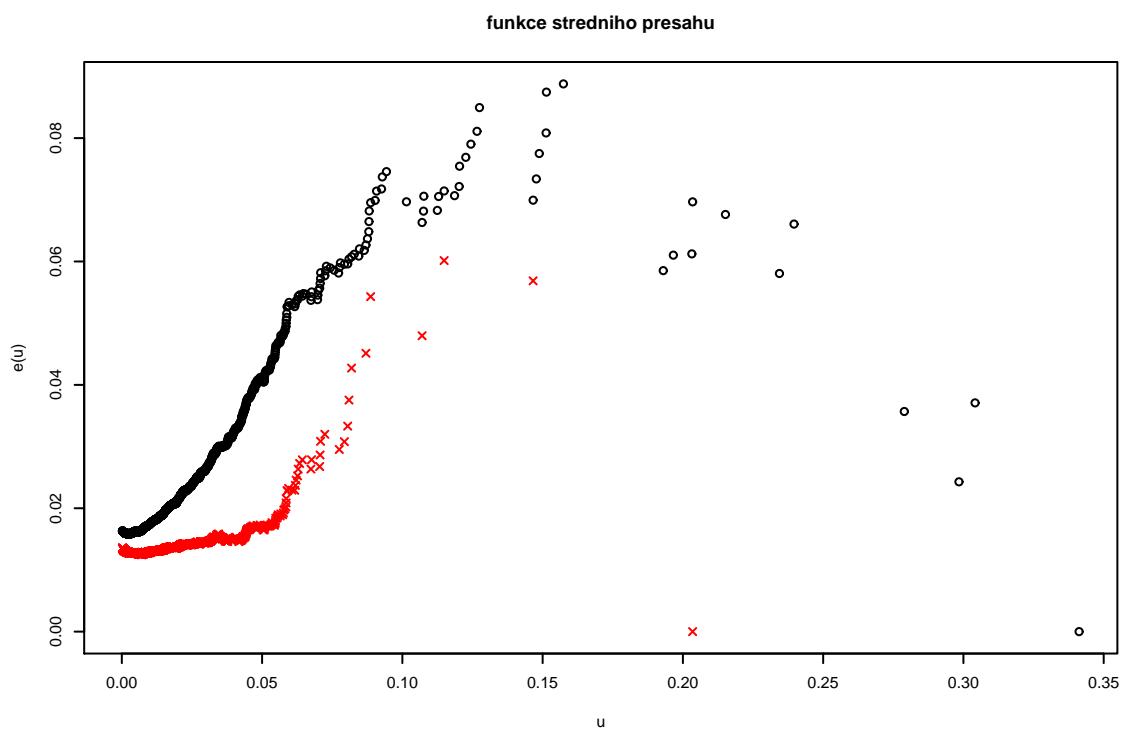


Obrázek 5.7: Poslední naměřená data jsou zobrazena vždy v levé části grafu. Na grafu dat včetně finanční krize (vpravo) je vidět velké množství drtivých ztrát v posledním roce a půl. Křížkem je na obrázcích označena nejvyšší ztráta v období před finanční krizí a úsečka ukazuje její hodnotu 0.203.



Obrázek 5.8: Uzavírací cena akcií BAC včetně období finanční krize.

Za jeden a půl roku v období krize byla celkem osmkrát překonána hodnota 0.2, což byla nejvyšší hodnota naměřená za předchozích 21 let. Navíc je hodnota 0.20 námi odhadnutý $\widehat{VaR}_{0.9999}$, což odpovídá hodnotě, která by měla být překonána pouze jednou za 10000 dní, tedy přepočteno na obchodní dny, zhruba jednou za 40 let. To ukazuje obrovský dopad krize na akcie BAC. Průměr těchto osmi největších ztrát je 0.264, což je konzistentní s našimi odhady $\widehat{ES}_{0.9999}$. Jak se změnilo rozdělení a chvost porovnáme pomocí grafu střední hodnoty přesahu - obrázek 5.9.

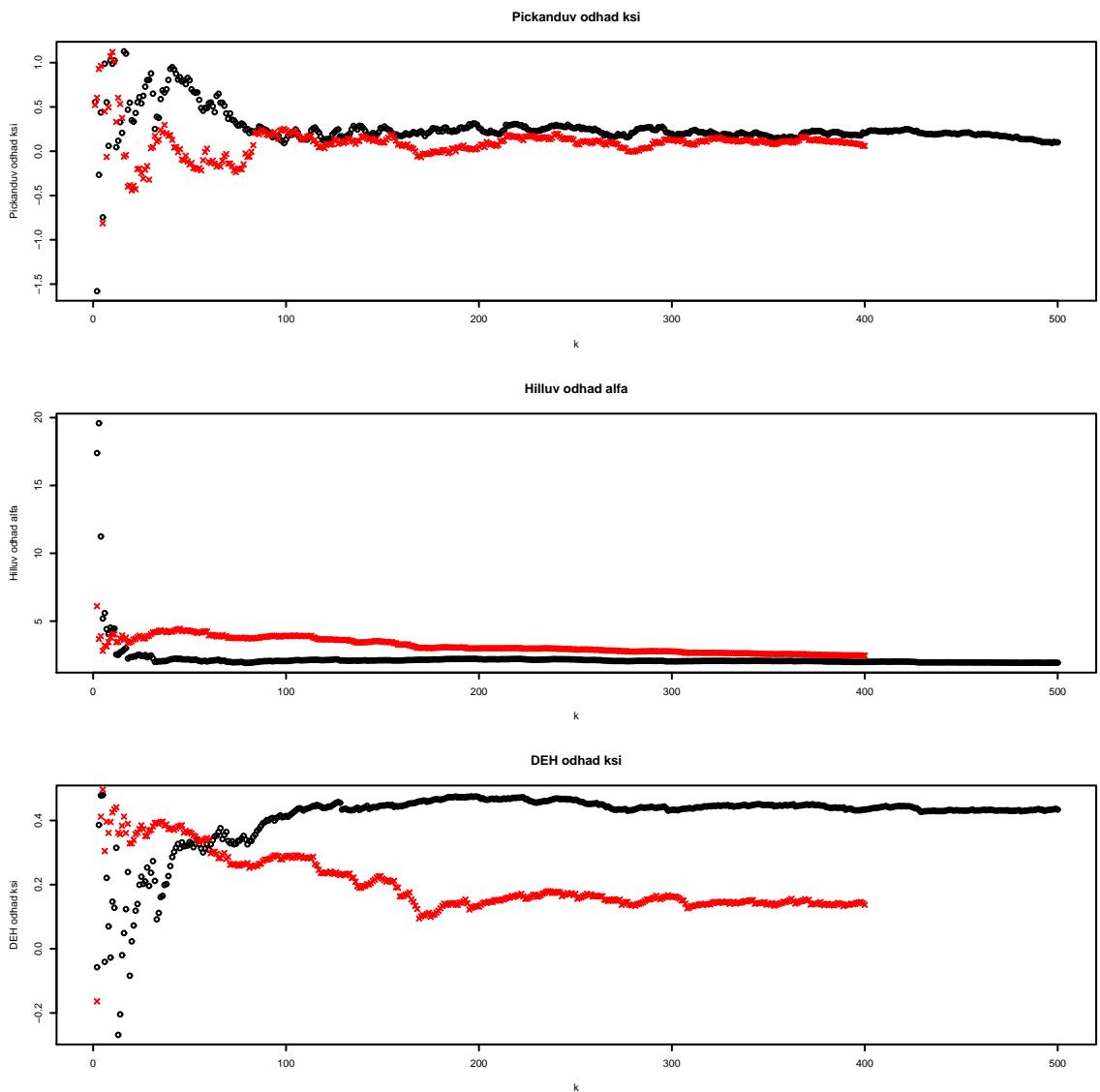


Obrázek 5.9: Červenými křížky je vyznačena empirická funkce středního přesahu z období před finanční krizí a černými kolečky je vyznačena empirická funkce středního přesahu za období včetně finanční krize.

Graf střední funkce přesahu nám říká, že se rozdelení velmi změnilo. Nová data mají opět rozdelení s těžkým chvostem a jejich $e(u)$ má podobný tvar jako $e(u)$ Paretova rozdelení. Největší směrnici má graf pro hodnoty u vyšší než 0.05, tuto hranici si také zvolíme jako vysoký práh při aplikaci odhadů pomocí GPD. Nyní provedeme odhad parametru ξ v GEV pomocí tří neparametrických metod popsaných v předchozích kapitolách. Do grafů zakreslíme body i z období před finanční krizí, abychom mohli porovnat vliv finanční krize na tihu chvostu, tedy na parametr ξ .

Především na Hillově a DEH grafu je vidět, že data včetně období finanční krize mají vyšší parametr ξ a tedy i těžší chvost. Hillův odhad, který je pro případ $\xi > 0$ nejpřesnější ukazuje hodnotu $\alpha = 2.15(\xi = 0.47)$ téměř pro všechna k . Podobnou hodnotu ξ těsně pod 0.5 ukazuje také DEH odhad. Připomeňme, že u dat před obdobím finanční krize byl $\xi = 0.26$, to tedy znamená, že rozdelení, z kterého data pocházejí má pravděpodobně nekonečný čtvrtý moment, avšak třetí moment je už konečný. Naopak u rozdelení dat včetně období finanční krize je při $\xi = 0.47$ třetí moment nekonečný, a pokud by skutečná hodnota ξ byla 0.5, nemělo by toto rozdelení ani konečný rozptyl. Pokud použijeme metodu maximální věrohodnosti na odhad parametru ξ v GPD s prahem $u = 0.05$, dostaneme 127 přesahů a odhady parametrů $\xi = 0.49$ a $\beta = 0.02$. Tedy ξ je opět těsně pod hodnotou 0.5 a výsledek je konzistentní s naším odhadem ξ pomocí Hillova odhadu. Jak se změnil chvost rozdelení přidáním dat z období finanční krize nejlépe ukáží následující tabulky, kde porovnáme odhady míry rizika spočtené pomocí Hillova odhadu dat včetně období krize a dat do období krize. P značí data z období před finanční krizí 29.5.1986-29.5.2007 a V značí data z období včetně finanční krize 29.5.1986-15.5.2009.

období dat	$\widehat{VaR}_{0.95}$	$\widehat{VaR}_{0.99}$	$\widehat{VaR}_{0.999}$	$\widehat{VaR}_{0.9999}$
P	0.041	0.062	0.112	0.201
V	0.048	0.102	0.296	0.865



Obrázek 5.10: Odhad parametru ξ pro denní logaritmické ztráty akcií BAC. Červenými křížky jsou vyznačena data před obdobím finanční krize a černými kolečky data včetně období finanční krize.

období dat	$\widehat{ES}_{0.95}$	$\widehat{ES}_{0.99}$	$\widehat{ES}_{0.999}$	$\widehat{ES}_{0.9999}$
P	0.055	0.083	0.15	0.271
V	0.09	0.19	0.554	1.618

Závěr

Pomocí metod popsaných v prvních čtyřech kapitolách této práce jsme provedli analýzu dvou souborů dat akcií BAC. Teorie extremálních rozdělení nám pomohla určit tvar pravého chvostu rozdělení těchto souborů dat. Díky tomu jsme mohli odhadnout míry rizika VaR a ES pro libovolně vysoké α . V průběhu analýzy jsme zjistili, že oba soubory dat pocházejí z rozdělení s těžkými chvosty. Soubor dat, který v sobě obsahuje data z období finanční krize, však odhalil mnohem těžší chvost, než soubor dat s daty pouze do začátku finanční krize. Tento fakt se projevil většími hodnotami VaR a ES u souboru dat včetně období finanční krize. Rozdíly u měr rizika jsou tím markatnější, cím větší volíme jejich parametr α . Platnost námi spočtených výsledků potvrzuje konzistence odhadů pomocí různých použitých metod a také konzistence odhadnutých chvostů s chvosty empirickými. Tyto výsledky lze použít například při řízení bilance aktiv a pasiv. Pro soukromého investora může ES s vhodně zvoleným α , posloužit jako odhad nejvyšší možné ztráty s danou pravděpodobností. V našem případě například odhad zohledňující data z období krize $\widehat{ES}_{0.999} = 0.554$ odpovídá denní ztrátě 43%. Tedy investor, který v současné době nakoupí akcie BAC čelí riziku, že v průměru jednou za 1000 obchodních dní ztratí 43% procent ze své investice. Strmý nárůst hodnot ES pro vysoké hodnoty α nám navíc říká, že pokud daný den obchodování jsou akcie již ve značné ztrátě, čelíme riziku, že ztráta může ještě drtivě narůst. Toto riziko můžeme částečně omezit zadáním denních stop-loss příkazů na úroveň například $\widehat{VaR}_{0.95} = 0.048$, tedy přibližně na úroveň 4,8% ztráty. Do budoucna bude zajímavé sledovat data akcií BAC po uklidnění turbulencí na akciových trzích a analyzovat, jak se chvost rozdělení denních logaritmických ztrát změní.

Literatura

- [1] Anděl J.: *Základy matematické statistiky*, Matfyzpress, Univerzita Karlova v Praze, 2007.
- [2] Embrechts P., Kluppelberg C., Mikosch T.: *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin, 1997.
- [3] McNeil A. J.: *Extreme Value Theory for Risk Managers*, ETH Zurich, 1999.