Stochastické modelovanie veľkých škôd v poisťovníctve

Zuzana Rošťáková

Ústav merania

Slovenská akadémia vied

22.1.2014

Veľké škody

- škody prevyšujúce určitú hranicu
- poistné plnenia, ktoré by vo väčšom množstve mohli spôsobiť finančné ťažkosti poisťovne

  → nízky počet, ALE veľký príspevok v konečnej sume škôd

  → dôležité vedieť ich predvídať

Veľké škody

- škody prevyšujúce určitú hranicu
- poistné plnenia, ktoré by vo väčšom množstve mohli spôsobiť finančné ťažkosti poisťovne

  $\rightarrow$ nízky počet, ALE veľký príspevok v konečnej sume škôd

  $\rightarrow$ dôležité vedieť ich predvídať

Veľké škody

- škody prevyšujúce určitú hranicu
- poistné plnenia, ktoré by vo väčšom množstve mohli spôsobiť finančné ťažkosti poisťovne

  → nízky počet, ALE veľký príspevok v konečnej sume škôd

  → dôležité vedieť ich predvídať

Veľké škody

- škody prevyšujúce určitú hranicu
- poistné plnenia, ktoré by vo väčšom množstve mohli spôsobiť finančné ťažkosti poisťovne

  → nízky počet, ALE veľký príspevok v konečnej sume škôd

  → dôležité vedieť ich predvídať

Veľké škody

- škody prevyšujúce určitú hranicu
- poistné plnenia, ktoré by vo väčšom množstve mohli spôsobiť finančné ťažkosti poisťovne

  → nízky počet, ALE veľký príspevok v konečnej sume škôd

  → dôležité vedieť ich predvídať

Cieľ práce

- cieľ:

1. vytvoriť model pre popis správania sa veľkých škôd
2. na základe vytvoreného modelu odhadnúť maximálnu hodnotu budúcich škôd

- metóda POT – Peaks Over Threshold

- aplikácia poznatkov na reálne dáta
  $\longrightarrow$ SOA Group Medical Insurance Large Claims Database

Cieľ práce

- cieľ:

1. vytvoriť model pre popis správania sa veľkých škôd
2. na základe vytvoreného modelu odhadnúť maximálnu hodnotu budúcich škôd

- metóda POT – Peaks Over Threshold

- aplikácia poznatkov na reálne dáta
  $\longrightarrow$ SOA Group Medical Insurance Large Claims Database

Cieľ práce

- cieľ:

1. vytvoriť model pre popis správania sa veľkých škôd
2. na základe vytvoreného modelu odhadnúť maximálnu hodnotu budúcich škôd

- metóda POT – Peaks Over Threshold

- aplikácia poznatkov na reálne dáta
  $\longrightarrow$ SOA Group Medical Insurance Large Claims Database

Prečo práve metóda POT?

- klasické metódy založené na rozdeleniach s ťažkým chvostom (gamma, loggamma, lognormálne, ...)
  $\longrightarrow$ potrebný veľký počet dát

- metóda POT
  $\longrightarrow$ odhady budú spoľahlivé aj pri nižšom počte pozorovaní, napr. 15

Podstata metódy POT

Podstata: odhad chvostu $1 - F(u + x)$

- Zovšeobecnené rozdelenie veľkých hodnôt (Generalised extreme value distribution)
- Zovšeobecnené Paretovo rozdelenie
- funkcia priemerného prírastku

Zovšeobecnené rozdelenie veľkých hodnôt – GEV

distribučná funkcia GEV rozdelenia

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

pričom $1 + \xi x > 0$ a

$$\begin{aligned} x > -\xi^{-1}, &\quad \text{ak} \quad \xi > 0; \\ x < -\xi^{-1}, &\quad \text{ak} \quad \xi < 0; \\ x \in \mathbb{R}, &\quad \text{ak} \quad \xi = 0. \end{aligned}$$

Maximum domain of attraction

Uvažujme postupnosť náhodných premenných $\{X_i\}$ s distribučnou funkciou F.
Náhodná premenná X (resp. jej distribučná funkcia F) patrí do maximum domain of attraction rozdelenia extrémnych hodnôt H, ak existujú konštanty $a_n > 0$, $b_n \in \mathbb{R}$ také, že

$$a_n^{-1}(M_n - b_n) \xrightarrow{d} H,$$

kde $M_n = \max\{X_1, \ldots, X_n\}$.

$$X \in \mathrm{MDA}(H)$$

Zovšeobecnené Paretovo rozdelenie – GPD

distribučná funkcia

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

$$x \in D(\xi, \beta);$$

kde

$$D(\xi, \beta) = [0, \infty), \qquad \xi \geq 0;$$

$$D(\xi, \beta) = \left[0, -\frac{\beta}{\xi}\right], \qquad \xi < 0.$$

Funkcia priemerného prírastku

- $x_F = \sup \{x \in \mathbb{R} : F(x) < 1\}$
- X je náhodná premenná s pravým koncovým bodom x_F
- $u \in \mathbb{R}^+, u < x_F$ je ľubovoľné pevné

Distribučná funkcia presahu F_u pre náhodnú premennú X:

$$F_u(x) = P(X - u \leq x | X > u), \quad x \in \mathbb{R}^+.$$

Funkcia priemerného prírastku náhodnej premennej X:

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

Funkcia priemerného prírastku pre niektoré typy rozdelení pravdepodobnosti

- spojité náhodné premenné

$$\mathrm{e(u)} = \frac{\int_{\mathrm{u}}^{\infty} [1-\mathrm{F(x)}]\mathrm{dx}}{1-\mathrm{F(u)}}$$

- exponenciálne rozdelenie

$$\mathrm{e(u)} = \frac{1}{\lambda}$$

- Paretovo rozdelenie

$$\mathrm{e(u)} = \frac{\kappa + \mathrm{u}}{\alpha - 1}, \quad \alpha > 1$$

- Weibullovo rozdelenie

$$\mathrm{e(u)} = \frac{\mathrm{u}^{1-\gamma}}{\mathrm{c}\gamma}\left(1 + \mathrm{o}(1)\right), \quad \mathrm{u} \to \infty$$

Funkcia priemerného prírastku pre GPD

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

- pre GPD je vždy lineárna funkcia premennej u
- využitie pri voľbe vhodného prahu u

Metóda POT

Odhad chvostu $1 - F(u + x)$:

- náhodný výber rozsahu n z rozdelenia s distribučnou funkciou F, $F \in MDA(H_\xi)$
- $\Delta_n(u) = \{i : X_i > u\}, \qquad N_u = \mathrm{card}\Delta_n(u)$
- $1 - F(u + x) = [1 - F(u)]\,[1 - F_u(x)]$
- $1 - F(u) \approx 1 - F_n(u) = \frac{1}{n}\sum_{i=1}^{n} I_{X_i > u} = \frac{N_u}{n}$
- $F \in MDA(H_\xi)$

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

$$\Rightarrow F_u(x) \approx G_{\xi,\beta(u)}(x)$$

Metóda POT

Odhad chvostu $1 - F(u + x)$:

- náhodný výber rozsahu n z rozdelenia s distribučnou funkciou F, $F \in MDA(H_\xi)$
- $\Delta_n(u) = \{i : X_i > u\}, \qquad N_u = \mathrm{card}\Delta_n(u)$
- $1 - F(u + x) = [1 - F(u)]\,[1 - F_u(x)]$
- $1 - F(u) \approx 1 - F_n(u) = \frac{1}{n}\sum_{i=1}^{n} I_{X_i > u} = \frac{N_u}{n}$
- $F \in MDA(H_\xi)$

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

$$\Rightarrow F_u(x) \approx G_{\xi,\beta(u)}(x)$$

Metóda POT

Odhad chvostu $1 - F(u + x)$:

- náhodný výber rozsahu n z rozdelenia s distribučnou funkciou F, $F \in MDA(H_\xi)$
- $\Delta_n(u) = \{i : X_i > u\}, \qquad N_u = \mathrm{card}\Delta_n(u)$
- $1 - F(u + x) = [1 - F(u)]\,[1 - F_u(x)]$
- $1 - F(u) \approx 1 - F_n(u) = \frac{1}{n}\sum_{i=1}^{n} I_{X_i > u} = \frac{N_u}{n}$
- $F \in MDA(H_\xi)$

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

$$\Rightarrow F_u(x) \approx G_{\xi,\beta(u)}(x)$$

Metóda POT

Odhad chvostu $1 - F(u+x)$:

- náhodný výber rozsahu n z rozdelenia s distribučnou funkciou F, $F \in MDA(H_\xi)$
- $\Delta_n(u) = \{i : X_i > u\}$, $N_u = \mathrm{card}\Delta_n(u)$
- $1 - F(u+x) = [1 - F(u)]\,[1 - F_u(x)]$
- $1 - F(u) \approx 1 - F_n(u) = \frac{1}{n}\sum_{i=1}^{n} I_{X_i > u} = \frac{N_u}{n}$
- $F \in MDA(H_\xi)$

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

$$\Rightarrow F_u(x) \approx G_{\xi,\beta(u)}(x)$$

Metóda POT

Odhad chvostu $1 - F(u + x)$:

- náhodný výber rozsahu n z rozdelenia s distribučnou funkciou F, $F \in MDA(H_\xi)$
- $\Delta_n(u) = \{i : X_i > u\}, \qquad N_u = \mathrm{card}\Delta_n(u)$
- $1 - F(u + x) = [1 - F(u)]\,[1 - F_u(x)]$
- $1 - F(u) \approx 1 - F_n(u) = \frac{1}{n} \sum_{i=1}^{n} I_{X_i > u} = \frac{N_u}{n}$
- $F \in MDA(H_\xi)$

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

$$\Rightarrow F_u(x) \approx G_{\xi,\beta(u)}(x)$$

Metóda POT

Odhad chvostu $1 - F(u+x)$:

- náhodný výber rozsahu n z rozdelenia s distribučnou funkciou F, $F \in MDA(H_\xi)$
- $\Delta_n(u) = \{i : X_i > u\}$, $\qquad N_u = \mathrm{card}\Delta_n(u)$
- $1 - F(u+x) = [1 - F(u)]\,[1 - F_u(x)]$
- $1 - F(u) \approx 1 - F_n(u) = \frac{1}{n}\sum_{i=1}^{n} I_{X_i > u} = \frac{N_u}{n}$
- $F \in MDA(H_\xi)$

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

$$\Rightarrow F_u(x) \approx G_{\xi,\beta(u)}(x)$$

Metóda POT

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

Tvorba modelu pre reálne dáta

1. voľba vhodného prahu u pomocou ME–grafu a TC–grafov
2. odhad parametrov ξ, β v GPD pre zvolené prahy
3. výber najvhodnejšieho modelu na základe QQ–grafov a KS–testu

Tvorba modelu pre reálne dáta

1. voľba vhodného prahu u pomocou ME–grafu a TC–grafov
2. odhad parametrov ξ, β v GPD pre zvolené prahy
3. výber najvhodnejšieho modelu na základe QQ–grafov a KS–testu

Tvorba modelu pre reálne dáta

1. voľba vhodného prahu u pomocou ME–grafu a TC–grafov
2. odhad parametrov ξ, β v GPD pre zvolené prahy
3. výber najvhodnejšieho modelu na základe QQ–grafov a KS–testu

Tvorba modelu pre reálne dáta

1. voľba vhodného prahu u pomocou ME–grafu a TC–grafov
2. odhad parametrov ξ, β v GPD pre zvolené prahy
3. výber najvhodnejšieho modelu na základe QQ–grafov a KS–testu

1. ME–graf

- $X_1, X_2, ..., X_n$ sú nezávislé, rovnako rozdelené náhodné premenné s distribučnou funkciou F
- empirická funkcia priemerného prírastku

$$e_n(u) = \frac{1}{\mathrm{card}\Delta_n(u)} \sum_{i \in \Delta_n(u)} (X_i - u).$$

- ME–graf

$$\left\{ \left(X_{(k),n}, e_n \left(X_{(k),n} \right) \right) : k = 1, ..., n \right\}$$

$\longrightarrow$ hľadáme hodnotu u, pre ktorú je $e_n(x)$ pre $x > u$ približne lineárna funkcia

1. ME–graf pre dáta z roku 1997

1. TC–grafy

- $X \sim G_{\xi_0,\beta_0}$
- $u_0, u_1 > u_0$ sú prahy
- $\beta = \beta_0 - \xi_0 u_0$

- TC–graf 1

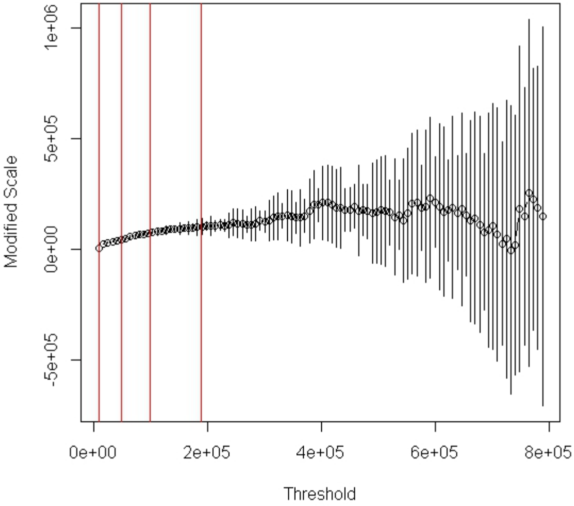
$$\{(u_1, \beta) : u_1 \leq \max\{x_1, \ldots, x_n\}\},$$

- TC–graf 2

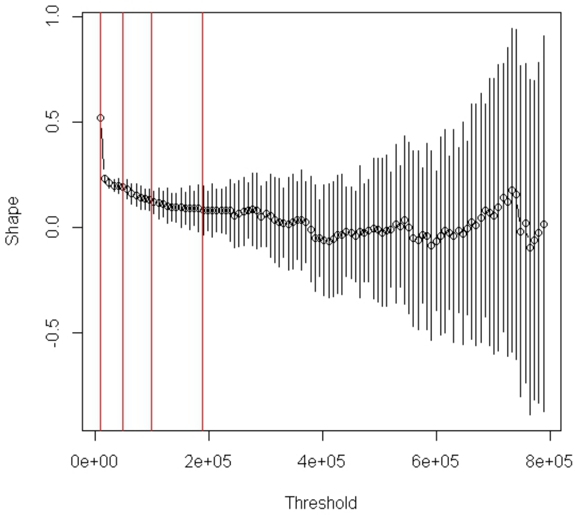
$$\{(u_1, \xi_0) : u_1 \leq \max\{x_1, \ldots, x_n\}\},$$

- hľadáme hodnotu u, od ktorej sú odhady ξ_0 a β približne konštantné

1. TC–grafy pre dáta z roku 1997



1. TC–grafy pre dáta z roku 1997



2. Odhady neznámych parametrov v GPD

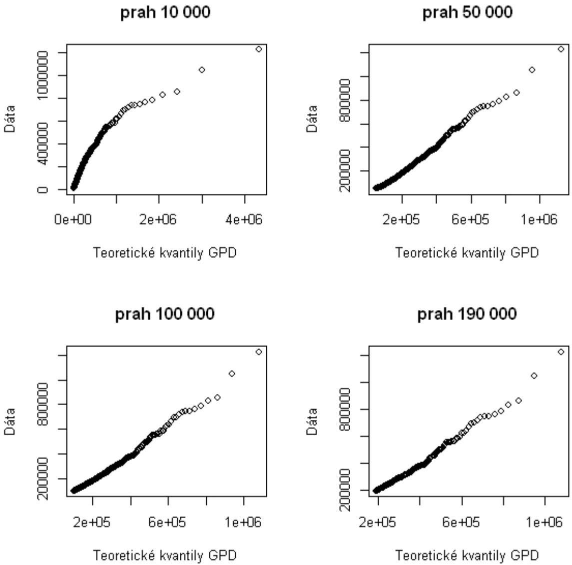
prah		MMV	s.e.	MM	s.e.
\$ 10 000	$\widehat{\xi}$	0,528	$8{,}873 \,.\, 10^{-3}$	0,541	
	$\widehat{\beta}$	$0{,}940 \,.\, 10^{4}$	$1{,}040 \,.\, 10^{2}$	$0{,}817 \,.\, 10^{4}$	
\$ 50 000	$\widehat{\xi}$	$1{,}923 \,.\, 10^{-1}$	$2{,}010 \,.\, 10^{-2}$	$4{,}048 \,.\, 10^{-1}$	0,031
	$\widehat{\beta}$	$5{,}417 \,.\, 10^{4}$	$1{,}410 \,.\, 10^{3}$	$3{,}224 \,.\, 10^{4}$	$0{,}990 \,.\, 10^{3}$
\$ 100 000	$\widehat{\xi}$	$1{,}265 \,.\, 10^{-1}$	$2{,}730 \,.\, 10^{-2}$	$2{,}797 \,.\, 10^{-1}$	$4{,}006 \,.\, 10^{-2}$
	$\widehat{\beta}$	$8{,}716 \,.\, 10^{4}$	$1{,}550 \,.\, 10^{3}$	$6{,}279 \,.\, 10^{4}$	$3{,}069 \,.\, 10^{3}$
\$ 190 000	$\widehat{\xi}$	$8{,}016 \,.\, 10^{-2}$	$4{,}498 \,.\, 10^{-2}$	$1{,}785 \,.\, 10^{-1}$	$6{,}540 \,.\, 10^{-2}$
	$\widehat{\beta}$	$1{,}199 \,.\, 10^{5}$	$2{,}135 \,.\, 10^{3}$	$9{,}851 \,.\, 10^{4}$	$8{,}424 \,.\, 10^{3}$

3. QQ–graf

$$\left\{\left(\mathrm{F}^{\leftarrow}\left(\frac{\mathrm{k}}{\mathrm{n}+1}\right), \mathrm{X}_{(\mathrm{k}),\mathrm{n}}\right) : \mathrm{k} = 1, \ldots, \mathrm{n}\right\}$$

$\longrightarrow$ približný lineárny priebeh grafu indikuje, že dáta pochádzajú z rozdelenia s distribučnou funkciou F

3. QQ–grafy pre dáta prevyšujúce zvolené prahy



3. Kolmogorov–Smirnov test

- $p_1, \ldots, p_{n_u}$ reprezentujú dáta presahujúce prah u,
- n_u je počet pozorovaní s danou vlastnosťou,

Testovali sme hypotézu:

- $H_0 : p_1, \ldots, p_{n_u}$ pochádzajú z rozdelenia s distribučnou funkciou $G_{\widehat{\xi},\widehat{\beta}}$
- $H_1 : p_1, \ldots, p_{n_u}$ nepochádzajú z rozdelenia s distribučnou funkciou $G_{\widehat{\xi},\widehat{\beta}}$.

- p–value pre zvolené prahy

\$ 50 000	\$ 100 000	\$ 190 000
$2,2.10^{-16}$	$1,912.10^{-9}$	$0,052$

Model pre dáta z roku 1997 vyššie ako $ 190 000

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

$\widehat{\xi}$	$8,016.10^{-2}$
$\widehat{\beta}$	$1,199.10^{5}$
N_u	332
n	33 325

Odhad maximálnej budúcej škody

PML = Probable Maximum Loss

- riešenie rovnice $P\left(M_n \leq PML_\varepsilon\right) = 1 - \varepsilon$
- $PML_\varepsilon = F^{\leftarrow}_{M_n}(1-\varepsilon)$
- M_n = maximálna škoda za nasledujúci časový úsek $\longrightarrow$ neznáma

Nech:

- $N, X_1, ..., X_n, ...$ sú nezávislé
- $N \sim Poi(\lambda)$ a $X_i \sim G_{\xi,\beta}$
- $M_N = \max\{X_1, ..., X_N\}$

Potom

$$P(M_N < x) = e^{-\lambda\left(1+\xi\frac{x}{\beta}\right)^{-\frac{1}{\xi}}}.$$

Odhad maximálnej budúcej škody

$$\mathrm{PML}_{\varepsilon} = \left[\left(-\frac{\lambda}{\ln(1-\varepsilon)} \right)^{\xi} - 1 \right] \frac{\beta}{\xi} + \mathrm{u}$$

Odhady pomocou vytvorených modelov

model pre dáta z roku 1997:

$$\mathrm{PML}_{0{,}05} = 1\ 716\ 680$$

$$\mathrm{PML}_{0{,}01} = 2\ 138\ 543$$

maximálna škoda v roku 1998:	\$1 342 615
v roku 1999:	\$2 568 512

Odhady pomocou vytvorených modelov

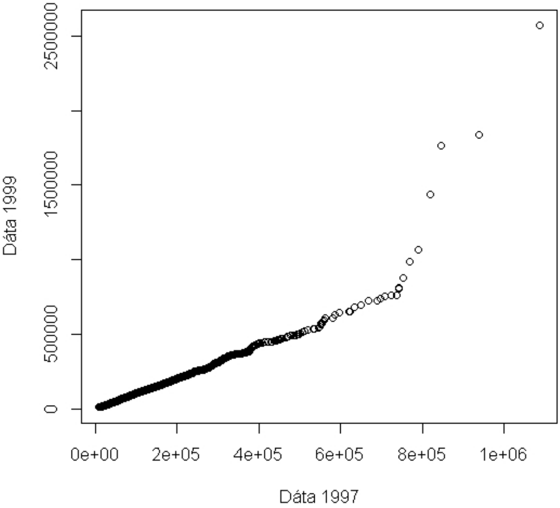
model pre dáta z roku 1997:

$$\mathrm{PML}_{0,05} = 1\ 716\ 680$$

$$\mathrm{PML}_{0,01} = 2\ 138\ 543$$

maximálna škoda v roku 1998:	\$1 342 615
v roku 1999:	\$2 568 512

Porovnanie charakteru dát



Záver

Vytvorený model je funkčný a schopný pracovať s veľkými škodami.

Literatúra

- Embrechts, P. – Klüppelberg, C. – Mikosch, T.: Modelling Extremal Events for Insurance and Finance. Springer-Verlag, 1997.
- Cebrián, A.C. – Denuit, M. – Lambert, P.: Generalized Pareto Fit to the Society of Actuaries´ Large Claims Database. In: North American Actuarial Journal, 3/2003.
- Coles, S.: An Introduction of Statistical Modeling of Extreme Values. Springer-Verlag, 2001.
- Michalíková, J.: Rozdelenia s ťažkými chvostami v neživotnom poistení. Bratislava: Univerzita Komenského. Fakulta matematiky, fyziky a informatiky, 2009. Diplomová práca. Vedúci diplomovej práce: doc. RNDr. Rastislav Potocký, CSc.
- Society of Actuaries: Medical Large Claims Experience Study. [online] Publikované: september 2004. [citované 15.3.2013] URL: <http://www.soa.org/Research/Experience-Study/Group-Health/research-medical-large-claims-experience-study.aspx>.
- Ribatet, M.A.: A User's Guide to the POT Package. [online] Verzia 1.4 (2011). Publikované: august 2006. [citované 20.3.2013] URL: <http://cran.r-project.org/web/packages/POT/vignettes/POT.pdf>.

Záver

Ďakujem za pozornosť.