

UNIVERZITA PALACKÉHO V OLOMOUCI  
PŘÍRODOVĚDECKÁ FAKULTA

**DIPLOMOVÁ PRÁCE**

Modelování extrémních škod v neživotním  
pojištění



**Katedra matematické analýzy a aplikací matematiky**  
Vedoucí bakalářské práce: **RNDr. Ondřej Pavlačka Ph.D.**  
Vypracovala: **Alena Lattonová**  
Studijní program: N1103 Aplikovaná matematika  
Studijní obor Aplikace matematiky v ekonomii  
Forma studia: prezenční  
Rok odevzdání: 2021

## BIBLIOGRAFICKÁ IDENTIFIKACE

**Autor:** Alena Lattonová

**Název práce:** Modelování extrémních škod v neživotním pojištění

**Typ práce:** Diplomová práce

**Pracoviště:** Katedra matematické analýzy a aplikací matematiky

**Vedoucí práce:** RNDr. Ondřej Pavlačka Ph.D.

**Rok obhajoby práce:** 2021

**Abstrakt:** Tématem této diplomové práce je modelování extrémních škod v neživotním pojištění. První kapitola seznamuje se základními pojmy z oblasti pojišťovnictví, přičemž se zaměřuje na problematiku zajištění a výpočtu ceny netto zajištění v WXL-R zajištění. Druhá kapitola pojednává o teorii extrémních hodnot. Poznatky z této kapitoly jsou následně využity v praktické části práce. V té jsou na základě datového souboru sestaveny modely popisující pravděpodobnostní rozdělení výše škod, přičemž cílem je co nejpřesněji modelovat především koncové hodnoty rozdělení. Závěrem jsou následně vybrané modely využity k stanovení ceny daného zajištění.

**Klíčová slova:** Teorie extrémních hodnot, neživotním pojištění, zajištění, netto zajištění

**Počet stran:** 74

**Počet příloh:** 1

**Jazyk:** český

## BIBLIOGRAPHICAL IDENTIFICATION

**Author:** Alena Lattonová

**Title:** Modelling of large claims in non-life insurance

**Type of thesis:** Master's

**Department:** Department of Mathematical Analysis and Application of Mathematics

**Supervisor:** RNDr. Ondřej Pavlačka Ph.D.

**The year of presentation:** 2021

**Abstract:** The topic of this master thesis is modelling of large claims in non-life insurance. The first chapter introduces the basic concepts of insurance, focusing on the issue of reinsurance and the calculation of the WXL-R reinsurance price. The second chapter deals with the Extreme value theory. The knowledge from this chapter is then used in the practical part of the thesis. Based on the data set are compiled models describing the probability distribution of the amount of damages, while the aim is to describe as accurately as possible the (right) tail values of the distribution. Finally, the selected models are used to determine the price of the reinsurance.

**Key words:** Extreme value theory, non-life insurance, reinsurance, price of reinsurance

**Number of pages:** 74

**Number of appendices:** 1

**Language:** Czech

### **Prohlášení**

Prohlašuji, že jsem diplomovou práci zpracovala samostatně pod vedením pana RNDr. Ondřeje Pavlačky, Ph.D. a všechny použité zdroje jsem uvedla v seznamu literatury.

V Olomouci dne .....

.....

podpis

# Obsah

Úvod	8
<b>1 Základní pojmy z oblasti pojišťovnictví</b>	<b>9</b>
1.1 Pojištění	9
1.2 Pojistné	11
1.3 Neživotní pojištění	12
1.4 Pojistně-technické riziko	14
1.5 Zajištění	15
1.5.1 Proporcionální zajištění	17
1.5.2 Neproporcionální zajištění	18
<b>2 Teorie extrémních hodnot</b>	<b>21</b>
2.1 Metoda blokového maxima	21
2.2 Metoda excedentů přes vysoký práh	23
2.2.1 Grafické metody výběru prahu	24
2.2.2 Odhady parametrů distribuční funkce	28
2.3 Ověření modelu	34
2.3.1 Grafické ověření sestaveného modelu	34
2.3.2 Statistické ověření sestaveného modelu	37
<b>3 Modelování výše škod pro stanovení netto zajistného v WXL-R zajištění</b>	<b>39</b>
3.1 Seznámení s daty	40
3.2 Modelování výše škod na základě celého datového souboru	41
3.3 Modelování výše škod na základě koncových hodnot datového souboru	44
3.3.1 Výběr vhodného prahu	44
3.3.2 Modelování koncových hodnot zobecněným Paretovým rozdělením	48
3.3.3 Ověření kvality modelu	51
3.4 Porovnání modelů	55
3.5 Výpočet netto zajistného WXL-R zajištění	57

Závěr	72
Literatura	73

## **Poděkování**

Tímto bych ráda poděkovala svému vedoucímu diplomové práce, panu RNDr. Ondřeji Pavlačkovi, Ph.D., za trpělivost, odborné rady a čas, který mi věnoval.

# Úvod

Komerční pojištění je součástí tržní ekonomiky a s dnešní zvyšující se životní úrovní roste i jeho důležitost v běžném životě nás všech. Rozhodnutí o jeho sjednání či nesjednání může mít i závažné existenční dopady. S těmito existenčními dopady se mohou potýkat ne jen jednotlivci, ale i pojišťovny samotné.

V neživotním pojištění se lze setkat s extrémně vysokými škodami, které mohou mít i přes svou zřídkačnost velký ekonomický dopad. Extrémně vysoké škody mohou vzniknout například při katastrofických událostech, nebo naakumulováním menších škod během roku. Jednou z možností, jak se může pojišťovna proti těmto nahodilým skutečnostem bránit, je sjednat si vhodný typ zajištění. Tato práce se pokouší nastínit výhody zajištění a seznamuje s jednotlivými typy zajištění (proporcionální a neproporcionální). Dále se zaměřuje na WXL-R zajištění a stanovení ceny zajištění pro tento konkrétní typ. Cílem této práce tak je sestavení vhodného modelu, který by popisoval pravděpodobnostní rozdělení extrémně vysoké škody, neboť takovýto model je potřebný pro výpočet daného zajistného.

Tato práce rovněž seznamuje s dvěma přístupy, pomocí kterých lze na hodnoty pohlížet jako na extrémní. Jedná se o metodu blokových maxim a metodu excedentů přes zvolený práh. Na takto definované extrémní hodnoty lze následně použít modely z teorie extrémních hodnot. Obě zmíněné metody jsou v této práci popsány, přičemž větší pozornost je věnována druhé z nich. Ta bude totiž aplikována na data v praktické části této práce. Závěrem bude model založený na této metodě použit ke stanovení ceny netto zajistného.

Praktická část této práce je doprovázena vybranými výstupy a grafy ze statistického programu *R*, ve kterém byla celá tato část zpracovávána.



# Kapitola 1

## Základní pojmy z oblasti pojišťovnictví

Tato kapitola se zabývá některými základními pojmy z oblasti pojišťovnictví, přičemž převážná část informací byla čerpána z [3], [5], [7], [8] a [9].

### 1.1. Pojištění

Svět kolem nás je ovlivňován řadou nejistot a nahodilostí. Ty mohou mít kladné i negativní dopady na život člověka. V některých případech mohou člověka ohrozit i existenčně. Je proto přirozené, že se člověk snaží dopad těchto negativních důsledků eliminovat. Jednou z možností, jak snížit negativní finanční důsledky nahodilostí, je pojištění. To se zabývá pouze jevy, které mají náhodný charakter a jejichž následky vzniká finanční potřeba. Takovéto jevy, ke kterým může dojít, pak označujeme jako pojistná rizika. V případě realizace daného jevu pak mluvíme o pojistné události. V takovém případě pak pojišťovna na základě pojistné smlouvy vyplácí pojistné plnění ve formě náhrad (jednorázových nebo opakovaných formou důchodů). Ty mohou mít podobu naturální (zdravotní, právní, technická asistence) nebo peněžní (náhrada škody, vyplacení obnosu).

Pojištění lze tedy chápat jako formu ochrany proti pojistnému riziku, kdy pojistník za úplatu (jednorázovou nebo běžnou) přenesse část rizika pojištěného, které by pro něj mohlo mít závažné následky, na pojistitele. Rizika pojištěných se tak transformují na takzvané pojistně-technické riziko (nebezpečí vzniku od-

chylky mezi přijatým pojistným a vyplaceným pojistným plněním) pojistitele. Ten při poskytování pojištění spoléhá na platnost zákona velkých čísel, čili že s růstem velikosti pojistného kmene se pojistně-technické riziko (směrodatná odchylka) zmenšuje a pojistitel je tak schopen rizika nejen převzít, ale popřípadě na nich i vydělat.

Pojištění lze dle právního hlediska kvalifikovat na:

- dobrovolné pojištění: sjednává se formou pojistné smlouvy na základě rozhodnutí pojistníka, který nemá povinnost dané pojištění uzavřít;
- povinné pojištění:
  - povinné smluvní pojištění: sjednává se formou pojistné smlouvy na základě zákona, který ukládá určité skupině osob povinnost si dané pojištění sjednat (například tzv. "povinné ručení");
  - zákonné pojištění: nesjednává se pojistná smlouva, ale ze zákona plyne pro určité skupiny osob povinnost pojistné platit, to však neodráží velikost rizika (např. zákonné pojištění odpovědnosti za škodu při pracovním úrazu a nemoci z povolání).

Další z možností je pojištění klasifikovat podle toho, zda závisí na výši škody, a to na:

- pojištění obnosová: výše pojistného plnění nezávisí na výši škody, která je kryta pojistnou smlouvou, ale odpovídá sjednané pojistné částce; navíc lze ke krytí jednoho rizika sjednat neomezený počet pojištění tohoto druhu;
- pojištění škodová: výše pojistného plnění, jehož účelem je vyrovnat úbytek majetku v důsledku pojistné události ve sjednaném rozsahu, nesmí přesáhnout hodnotu škody, aby nedocházelo k obohacení pojištěného; opět lze sjednat neomezený počet pojištění tohoto druhu pro krytí jednoho rizika, ale součet pojistných plnění nesmí přesáhnout hodnotu škody.

Pojištění lze dále klasifikovat podle způsobu tvorby rezerv na:

- životní pojištění: kryje životní rizika (riziko úmrtí, riziko dožití se určitého věku, smíšené pojištění pro případ smrti nebo dožití) a jsou pro ně typické pojistné smlouvy dlouhodobějšího charakteru; dochází zde k vytváření pojistné rezervy pro krytí pojistné události, která v budoucnu nastane, ale není známo kdy;
- neživotní pojištění: kryje rizika jiné než životní; není známo zda vůbec dojde k pojistné události, ani jaká by byla případná výše škody.

## 1.2. Pojistné

Cena pojistného se kalkuluje na základě konstrukce příslušného druhu pojištění, technických podkladů pro výpočet pojistného a na základě obsahu příslušného druhu pojištění, protože každý druh pojištění může krýt jiné riziko o různé pravděpodobnosti realizace. Pojistník pak může pojistné sjednat ve formě běžného pojistného, u něhož se zaváže platit pojistné vždy na začátku sjednaného období nebo ve formě jednorázového pojistného.

Částku, kterou je pojistník smluvně vázán platit pojistiteli za krytí rizika, nazýváme brutto pojistné. To se skládá ze tří složek:

- rizikové pojistné,
- kalkulované správní náklady,
- kalkulovaný zisk.

Výpočet rizikového pojistného je založeno na principu ekvivalence, který požaduje, aby příjmy a výdaje pojišťovny byly v rovnováze. Tyto hodnoty lze ocenit pomocí očekávaných počátečních hodnot. Princip ekvivalence pak lze konkretizovat na požadavek, aby očekávaná počáteční hodnota pojistného byla rovna očekávané počáteční hodnotě pojistného plnění. V praxi by se ovšem mohlo takto kalkulované pojistné stát pro pojišťovnu značně nevýhodné, neboť skutečně realizované hodnoty se mohou od střední hodnoty zásadně lišit, proto se ve výpočtu dále ještě přihlíží na směrodatnou odchylku. Rizikové pojistné tak lze rozložit na dvě

složky, a to na netto pojistné, které je úměrné průměrné velikosti rizika, a na bezpečnostní přírážku, která kryje odchylky od průměrných pojistných plnění. Rizikové pojistné tak slouží k pokrytí výdajů na pojistná plnění a k tvorbě rezerv. Výše rizikového pojistného závisí na pravděpodobnosti výskytu rizika a případné velikosti škody při jeho realizaci. Pro výpočet pak lze použít škodní nebo úmrtnostní tabulky, které modelují skutečnost.

Kalkulované správní náklady slouží k pokrytí provozních a správních nákladů pojišťovny a nákladů spojených s provozem pojistného produktu. Do pojištění je lze zahrnout jako jednotnou správní přírážku, používanou často u neživotního pojištění, nebo jako diferencovatelnou správní přírážku, používanou převážně u životního pojištění. Tato přírážka zahrnuje počáteční jednorázové náklady a běžné správní náklady, které se většinou zahrnují do ceny pojištění jakožto procento z pojistné částky. Další diferencovatelnou správní přírážkou jsou inkasní náklady, které se většinou zahrnují do ceny pojištění jakožto procento z brutto pojistného, a náklady spojené s pravidelnou výplatou důchodů, které se většinou zahrnují do ceny pojištění jakožto procento z ročního důchodu.

Nelze opomenout, že hlavním cílem pojišťovny je dosáhnout zisku. U životního pojištění zisk plyne z poplatků a z výnosů plynoucích z investování rezerv. U neživotního pojištění si pojišťovna zisk zajistí zahrnutím položky kalkulovaný zisk do pojistného. Tuto položku si každá pojišťovna stanoví sama na základě situace na pojistném trhu.

### **1.3. Neživotní pojištění**

Pojistně-matematická teorie se staví k životnímu a neživotnímu pojištění rozdílně, z čehož plynou i rozdílné modely pro kalkulaci pojistného. Je účelné si proto ujasnit některé rozdíly mezi těmito odvětvími.

Pro neživotní pojištění jsou oproti životnímu pojištění typické smlouvy na kratší pojistnou dobu, a to většinou na jeden rok s tím, že případně dochází k plynulému prodlužování smlouvy. Pojišťovna i pojistník proto mohou smlouvu ke konci pojistného období vypovědět a cena je stanovena "spravedlivě" pro dané

období. V neživotním pojištění se ve většině případů jedná o škodové pojištění s pojistným plněním omezeným skutečným rozsahem škody, kdy její velikost a počet lze pouze odhadnout. Oproti tomu v životním pojištění je při pojistném plnění vyplácena pevně stanovená pojistná částka s tím, že u variant životního pojištění na dožití se nebo na smrt v průběhu trvání pojištění, je nastoupení jedné z variant jisté, náhodná je pouze doba, kdy k události dojde.

U životního pojištění patří mezi základní proměnné veličiny určující výši rizikového pojistného úmrtnostní tabulky a pojistně-technická úroková míra. Kalkulace pojistného v neživotním pojištění je obtížnější, protože náhodné vlivy působící na škodní průběh se vyznačují svou velkou nestabilitou. Při tvorbě kalkulačního modelu neživotního pojištění je proto nezbytné, určit proměnné veličiny ovlivňující škodní průběh, dále určit tarifní třídy, v rámci kterých je riziko přibližně stejné a následně pak odhadnout výši pojistného.

Uvedme proto možné členění neživotního pojištění s jeho nejvýznamnějšími produkty:

- Neživotní pojištění osob:
  - Úrazové pojištění,
  - Soukromé nemocenské pojištění,
  - Soukromé zdravotní pojištění;
- Majetkové pojištění:
  - Pojištění majetku obyvatelstva,
  - Pojištění podnikatelských a průmyslových rizik,
  - Pojištění zemědělských rizik;
- Pojištění finančních ztrát a záruk:
  - Pojištění pro případ přerušení provozu,
  - Pojištění úvěru,

- Pojištění cestovních kanceláří pro případ úpadku;
- Pojištění odpovědnosti:
  - Pojištění odpovědnosti za škodu způsobenou provozem vozidel,
  - Profesní odpovědnostní pojištění,
  - Obecné odpovědnostní pojištění;
- Pojištění právní ochrany:
  - Pojištění právní ochrany motorového vozidla nebo řidiče motorového vozidla,
  - Pojištění právní ochrany rodiny,
  - Pojištění právní ochrany podniků.

## 1.4. Pojistně-technické riziko

Základním úkolem komerční pojišťovny je za úplatu na sebe přenést rizika svých klientů, kteří se tak krijí před potencionálními škodami, které by pro ně mohli představovat zásadní a negativní dopad na jejich existenci. Jak již bylo uvedeno, je pojišťovna schopna tyto rizika převzít, avšak i sama pojišťovna se potýká s riziky, které by mohly zásadně ovlivnit její existenci. Jedním takovým rizikem, které vyplývá z pojišťovací činnosti, je pojistně-technické riziko pojišťovny. To je definováno jako možnost vzniku odchylky (kladné tak i záporné) od pojišťovnou předpokládaného škodního průběhu a nákladů správní režie. Aby pojišťovna eliminovala vznik negativní odchylky, je potřeba mít dostatečně velký pojistný kmen, systém rezerv a správně kalkulované pojistné, které nebude podhodnocovat skutečné riziko. Toto riziko plyne z výše a četnosti škod. Pojistně-technické riziko lze dále podle původu odchylek skutečnosti od předpokladu rozdělit na:

- náhodné pojistně-technické riziko: je spojeno s náhodnými výkyvy od očekávaného průměrného škodního průběhu, který se však v čase téměř nemění; toto riziko lze dále rozdělit z hlediska rozsahu na normální náhodné riziko,

kdy škodní průběh vykazuje běžné kolísání kolem očekávaného průměru, a na katastrofální náhodné riziko, kdy je škodní průběh ovlivňován škodami velkého rozsahu;

- pojistně-technické riziko změn: je spojeno se situací, kdy škodní průběh není z dlouhodobého hlediska konstantní, čili se v čase mění z čehož plyne rovněž změna podkladů pro výpočet pojistného;
- pojistně-technické riziko omylu: se týká situace, kdy dojde k nesprávnému odhadu škodního průběhu.

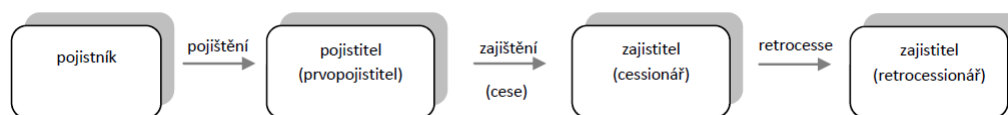
Finanční požadavky na pojišťovnu plynoucí z negativí odchylky od očekávaného škodního průběhu může pojišťovna pokrýt pomocí již výše zmíněných systému rezerv, ale i ty mohou být mnohdy nedostačující. V dnešní době se pojišťovny stále častěji potýkají s událostmi, jejich frekvence a nežádoucí finanční dopady jsou čím dál tím vyšší. Jedním z důvodů může být například i rostoucí hustota zalidnění, díky níž je v případě (přírodní) katastrofické události zasaženo více osob respektive pojištěných, z čehož plyne i větší počet pojistných plnění.

Katastrofické události, i přes svou zřídka, proto mohou mít zásadní dopad na fungování pojišťovny, a to hlavně z důvodu, že při nich často dochází k extrémně vysokým škodám. Jako ilustrativní příklad můžeme uvést přírodní katastrofy, teroristický útok či jinou katastrofu způsobenou lidským faktorem, jako je například pád letadla. Z těchto skutečností plyne potřeba co nejpřesněji modelovat výši možné škody události, při které vznikají takovéto vysoké extrémní škody. Díky tomu je pojišťovna schopna nastavit správnou cenu pojistného, popřípadě si zvolit vhodnou formu zajištění proti převzatému riziku.

## 1.5. Zajištění

Zajištění je převod části rizika pojištěného, ke kterému se uvázal pojistitel, na jiného poskytovatele pojištění označovaného jako zajistitel. Vlastním vrubem potom označujeme tu část rizika, kterou si pojistitel ponechá ke krytí. Zajistitelem může být jiná pojišťovna nebo instituce zabývající se speciálně touto

zajišťovací činností (zajišťovna). Zajistitel tedy sice na sebe přebírá část rizika pojistitele (zajistník, prvopojistitel, cedent), ale to bez jakéhokoliv právního vztahu k pojistníkovi. Zajistitel je pouze v právním vztahu s prvopojistitelem (pojišťovnou), ve kterém se prvopojistitel zavazuje k platbě zajistného ve sjednané výši. Výměnou za to se zajistitel zavazuje, že se bude v případě pojistné události pojištěného podílet na výplatě pojistných plnění dle předem sjednaných podmínek. Zajištění je teda analogické pojišťovací činnosti, jen o úroveň výše. Další úroveň, při které zajistitel převádí část převzatého rizika na další pojišťovací subjekt, se nazývá retrocese. Riziko tak lze přerozdělit mezi více pojistných subjektů, což může mít mnoho výhod. Toto schéma zajištění zachycuje obrázek 1.1.



Obrázek 1.1: Základní schéma zajištění. Zdroj: [7].

Zajišťovny jsou mnohdy nadnárodní podnikatelské subjekty, které díky svým velkým pojistným kmenům a jejich územní a produktové diverzifikaci učinili svou činnost výnosovou. I pro prvopojistitele plynou ze zajištění se výhody, jako je například vyšší finanční stabilita, která plyne ze sjednané ochrany před neočekávanou zápornou odchylkou od předpokládaného škodního průběhu. Díky zajištění tak dochází k eliminaci pojistně-technického rizika, k navýšení solventnosti prvopojistitele a k homogenizaci jeho pojistného kmene. Mezi další výhody zajištění patří možnost prvopojistitele uvést na pojistný trh nové produkty u nichž pojišťovna nemá informaci o možném škodním průběhu. Dále pojistitel nemusí díky zajištění odmítat pojistit klienty s vysokým pojistným rizikem, které by přesahovalo jeho možnosti. Riziko se totiž díky zajištění rozloží mezi více účastníků. V případě účasti nadnárodních zajišťoven je pak možné rozložit dopad rizik mezi subjekty po celém světě. Poměr, v jakém si riziko přerozdělí, závisí na sjednané formě zajištění, která může být buďto proporcionální nebo neproporcionální.



### 1.5.1. Proporcionální zajištění

Při sjednání proporcionálního zajištění se pojistné, pojistná částka a pojistné plnění dělí ve smluvně sjednaném poměru mezi prvopojistitele a zajistitele s přihlednutím na omezení dané limitem zajistitele. V praxi se nejčastěji využívají dva typy proporcionálního zajištění, a to zajištění kvótové (quota share) a excedentní (surplus).

#### Kvótové zajištění

Při kvótové zajištění se zajistitel zavazuje krýt riziko v stále stejném poměru dané sjednanou kvótou zajistitele  $q$  ( $0 < q < 1$ ), zároveň má právo na stejný poměr z pojistného. Díky jednoduchému principu má tento druh zajištění nízké nároky na administrativu. Na druhou stranu při jeho použití nemusí dojít k požadované homogenizaci pojistného kmene a do zajištění spadají i pojistné smlouvy s nízkými pojistnými částkami, které by pojistitel byl schopen pokrýt bez pomoci sám.

#### Excedentní zajištění

Prvopojistitel smluvně určí výši pojistné částky, která nepodléhá zajištění (vlastní vrub). Čímž narozdíl od kvótového zajištění dochází k lepší homogenizaci pojistného kmene a prvopojistitel si může sám zvolit částku, od které se bude zajišťovat. Zajistitel si rovněž může stanovit limit zajistitele, který představuje maximální výši, kterou je ochoten zajistitel převzít k zajištění. Tento limit se většinou udává v násobcích vlastního vrubu. Poměr, v jakém si prvopojistitel a zajistitel rozdělí pojistné a pojistné plnění, je pak stejný jako poměr mezi částí, kterou kryje zajistitel (od vlastního vrubu prvopojistitele po případný limit zajistitele) a částí, kterou kryje prvopojistitel (zbylá část). Poměr, ve kterém si rozdělují prvopojistitel a zajistitel riziko, může být tedy v tomto případě pro každou smlouvu jiný. V rámci jedné smlouvy je ale poměr, ve kterém si přerozdělí pojistné, pojistné plnění a pojistnou částku, vždy stejný.

## 1.5.2. Neproporcionální zajištění

U proporcionálního zajištění je předem stanoven poměr, podle kterého se dělí pojistná částka mezi prvopojistitele a zajistitele. Ve stejném poměru si pak přerozdělí i pojistné a pojistné plnění. U neproporcionálního zajištění není předem stanoven žádný poměr, pouze podmínky (závisí na typu sjednaného zajištění), které v případě překročení smluvně stanoveného objemu (priorita) pojistného plnění ukládají zajistiteli povinnost krýt tu část škody, která přesahuje hodnotu sjednané priority  $a$ , ale to maximálně do zajistitelem sjednané výše  $m$  (vrstva). Zajistné se v tomto případě kalkuluje na základě pravděpodobnosti, s kterou skutečná výše škody (pojistné plnění) překročí danou prioritu. K tomu lze využít historická data prvopojistitele o škodním průběhu uvažovaného pojištění. Popřípadě pokud data nejsou k dispozici, například z důvodu zavádění nového produktu na trh, je možné využít údaje o jiném dobře zdokumentovaném produktu a při kalkulaci tyto skutečnosti zohlednit.

### Zajištění nadměrku škodovosti

Zajištění nadměrku škodovosti (SL z anglického stop loss) chrání prvopojistitele před dopadem zvýšeného škodního průběhu během roku. Zajistitel k této formě zajištění většinou přistoupí pouze pod podmínkou, že bude povinen účastnit se krytí škod jen v případě, kdy bude prvopojistitel v daném roce ztrátový, tj. pojistné plnění spolu s náklady převyší přijaté pojistné. Výhodou tohoto zajištění je jeho komplexní ochrana schopnost, a to jak proti jednotlivým extrémně vysokým škodám, tak i proti naakumulování menších škod. Vzhledem k tomu, že je krytí takového rizika pro zajistitele poměrně rizikové, se tento druh zajištění sjednává méně často a za vyšší zajistné.

### Zajištění největších škod LCR (p)

Zajištění LCR(p) (z anglického largest claims reinsurance) kryje prvopojistitele před  $p$  největšími škodami, které během roku nastanou.

## Zajištění ECOMOR (p)

Při sjednání zajištění ECOMOR (p) (z francouzského excédent du cout moyen relatif) hradí prvopojistitel všechny škody pouze do výše hodnoty  $p$ -té největší škody.

## Zajištění škodního nadměrku

Zajištění škodního nadměrku (XL z anglického excess of loss) se dále ještě dělí na zajištění škodního nadměrku jednotlivých rizik (WXL-R z anglického working excess of loss per risk) a na zajištění škodního nadměrku katastrofické události (Cat-XL z anglického catastrophe excess of loss). V prvním případě se prvopojistitel zajišťuje proti škodě přesahující prioritu pojistitele a vztahující se k jedné pojistné smlouvě. V druhém případě se zajišťuje před dopadem kumulací škod vztahující se k jedné katastrofické události, kdy by součet jednotlivých, ač malých, škod mohl ve výsledku překročit prioritu prvopojistitele.

## Výpočet netto zajistného při sjednání WXL-R zajištění

Tato práce se zaměřuje především na zajištění škodního nadměrku jednotlivých rizik, proto je tomuto typu zajištění věnována větší pozornost. Pro jednoduchost budeme dále uvažovat tento typ zajištění při sjednání priority  $a$  bez sjednání kapacity vrstvy  $m$ . Toto zajištění kryje část každé škody  $X_i$ ,  $i = 1, \dots, N$  přesahující prioritu  $a$ , kde  $N$  je celkový počet vzniklých škod. Prvopojistitel při sjednání tohoto typu zajištění hradí veškeré vzniklé škody  $X_i$  pouze do hodnoty  $a$ , tj.

$$X_{P_i} = \min\{a, X_i\}, \quad i = 1, \dots, N.$$

Dle [5] lze očekávanou hodnotu jednotlivého pojistného plnění krytého prvopojistitelem  $X_P$  zapsat pomocí distribuční funkce pojistného plnění  $X$  ( $F_X(x)$ ) ve tvaru

$$E(X_P) = \int_0^a x dF_X(x) + a \cdot (1 - F_X(a)) = \int_0^a (1 - F_X(x)) dx.$$

Část škody  $X_i$  krytou zajistitelem lze zapsat ve tvaru

$$X_{Z_i} = \max\{0, X_i - a\}, \quad i = 1, \dots, N.$$

Očekávanou hodnotu jednotlivého pojistného plnění krytého zajistitelem  $X_Z$  lze dle [5] zapsat ve tvaru

$$E(X_Z) = \int_a^\infty x dF_X(x) - a \cdot (1 - F_X(a)) = \int_a^\infty (1 - F_X(x)) dx = E(X) - E(X_P).$$

Celková výše plnění zaplacená zajistitelem pak má tvar

$$S_Z = \sum_{i=1}^N X_{Z_i}.$$

Cena netto zajistného se stanovuje na základě principu ekvivalence, kdy očekávaná hodnota příjmů se musí rovnat očekávané hodnotě výdajů. Netto zajistné při sjednání WXL-R zajištění je tak rovno očekávané hodnotě zajistného plnění, tj.

$$P_Z = E(S_Z) = E(N)E(X_Z),$$

kde  $N$  a  $X_Z$  jsou náhodné, nezávislé veličiny a  $E(N)$  je očekávaný počet celkových škod. Označme dále  $E(N_Z)$  očekávaným počtem škod, na kterých se podílí zajistitel, tj. očekávaný počet škod přesahující zvolenou prioritu  $a$ . Z dat je odhad  $E(N_Z)$  vzhledem k malému počtu škod přesahující prioritu  $a$  značně nespolehlivý, proto se využívá vztahu

$$E(N_Z) = p_a E(N),$$

kde  $p_a = P(X > a) = 1 - F_X(a)$ .

Aby cena netto zajistného v WXL-R zajištění byla stanovena co nejspravedlivěji, je třeba zvolit takové modely, který by věrohodně popisovaly dané náhodné veličiny.

# Kapitola 2

## Teorie extrémních hodnot

Přejímání rizik pojišťovny na zajišťovnu má velký význam. Pojistné události s extrémně vysokým pojistným plněním mají malou pravděpodobnost, ovšem ne zanedbatelnou. Ačkoliv je jejich škodní frekvence nízká, každoročně většina finančních prostředků na celkové pojistné plnění připadá právě na tyto události. Pojišťovny si tak jako ochranu sjednávají zajištění proti extrémním škodám. Cena proporcionálního zajištění se odvíjí od sjednaného poměru, podle kterého si prvopojistitel a zajistitel rozdělí riziko. Cena neproporcionálního zajištění se stanovuje na základě pravděpodobnostního rozdělení výše a četnosti (extrémních) škod.

O problematice modelování pojednává teorie extrémních hodnot (Extreme Value Theory, EVT), které je věnovaná celá tato kapitola. Ta čerpá především z [2], [4], [11], [14] a [15]. EVT se využívá i v jiných odvětvích, např. v meteorologii, hydrologii, strojírenství, kde je potřeba modelovat extrémní hodnoty. Podle způsobu, kterým se identifikují tyto extrémní hodnoty, rozlišujeme dvě základní metody: metoda blokového maxima a metoda excedentů přes vysoký práh.

### 2.1. Metoda blokového maxima

Nechť  $X_1, \dots, X_n$  jsou nezávislé náhodné veličiny se společnou distribuční funkcí  $F(x) = P(X \leq x)$ . Tyto veličiny tvoří blok délky  $n$ , přičemž metoda blokového maxima považuje za extrémní hodnotu maximum jednotlivého bloku. Extrémní hodnota  $M_n = \max\{X_1, \dots, X_n\}$  je tedy rovněž náhodná veličina, a

to s distribuční funkcí  $F_n(x) = P(M_n \leq x)$ . Blok v praxi většinou představuje logický časový úsek, například měsíc, rok. Jeho délka by měla být zvolena tak, aby bylo možné zanedbat sezónní vlivy. Nutno dodat, že distribuční funkci  $F_n(x)$  lze modelovat za předpokladu, že funkci  $F(x)$  známe přesně. Odhad distribuční funkce  $F(x)$  je pro modelování  $F_n(x)$  nedostačující, neboť i malé odchylky od skutečného průběhu distribuce  $F(x)$  mohou vést k velkému odchylení od skutečného průběhu distribuce  $F_n(x)$ . Za předpokladu, že funkci  $F(x)$  neznáme, lze, za pomoci Fisher-Tippetovy věty, funkci  $F_n(x)$  přiřadit jedno ze tří zobecněných rozdělení extrémních hodnot (Fréchetovo, Gumbelovo, Weibullovo).

**Věta 2.1.1** (Fisher-Tippetova věta). *Nechť  $X_1, X_2, \dots, X_n$  je posloupnost identicky rozdělených náhodných veličin a  $M_n = \max\{X_1, \dots, X_n\}$ . Jestliže existují konstanty  $c_n > 0$  a  $d_n$  a nějaká nedegenerovaná distribuční funkce  $H(x)$  taková, že*

$$P(c_n^{-1}(M_n - d_n) \leq x) \rightarrow H(x),$$

*potom  $H(x)$  přísluší jednomu z následujících tří typů rozdělení:*

1. *Fréchet*

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

2. *Gumbel*

$$\Lambda_\alpha(x) = \exp(-e^{-x}), \quad \text{pro } x \in \mathbb{R}.$$

3. *Weibull*

$$\Psi_\alpha(x) = \begin{cases} 1, & \text{pro } x > 0, \\ \exp(-(-x)^{-\alpha}), & \text{pro } x \leq 0, \alpha > 0. \end{cases}$$

Fréchetovo, Gumbelovo a Weibullovo rozdělení může být nahrazeno společnou distribuční funkcí zobecněného extrémálního rozdělení (GEV):

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

Pro různý tvar parametru extrémních hodnot  $\xi$ , lze opět získat jedno z výše uvedených typů rozdělení, přičemž kladné hodnoty  $\xi$  zaručují těžké konce rozdělení. Otázkou u této metody zůstává, jakou velikost bloku nastavit. Pokud se blok zvolí příliš velký, pak metoda generuje pouze pár extrémních hodnot a jiné hodnoty, které je možné považovat za extrémní se opominou. Nedostatek pozorování pak vede k velkému rozptylu odhadu parametrů. Na druhou stranu, pokud se blok zvolí příliš malý, mohou maxima bloků zahrnovat i hodnoty, které lze považovat spíše za běžné než za extrémní. Pak zvolená distribuce dle Fisher-Tippetovy věty bude zřejmě vybrána nevhodně.

## 2.2. Metoda excedentů přes vysoký práh

Nechť  $X_1, \dots, X_n$  jsou nezávislé náhodné veličiny se společnou distribuční funkcí  $F$  a nechť  $u$  je nějaká vysoká prahová hodnota. Za extrémní hodnoty považujeme ty hodnoty  $X_i$ , které překračují zvolený práh  $u$ . Tyto hodnoty jsou stále vzájemně nezávislé náhodné veličiny. Jejich rozdělení lze v případě, kdy známe přesně funkci  $F$ , popsat pomocí podmíněné pravděpodobnosti:

$$P(X_i > y | X_i > u) = \frac{1 - F(y)}{1 - F(u)}, \text{ kde } y > u.$$

V praxi je ale  $F$  často neznámá. Touto problematikou se zabývá teorie extrémních hodnot. Ta hledá vhodnou aproximaci funkce  $F_n$ , představující rozdělení excedentů, tj.  $X - u$ , za předpokladu, že funkce  $F$  je neznámá. V případě metody excedentů přes určitý vysoký práh  $u$  je limitním rozdělením excedentů zobecněné Paretovo rozdělení (GPD) s distribuční funkcí  $H_{\xi, \sigma_u}^{(u)}(y)$ :

$$F_n(y) = P(X - u \leq y | X > u) \rightarrow H_{\xi, \sigma_u}^{(u)}(y), \quad u \rightarrow \infty, \text{ kde}$$

$$H_{\xi, \sigma_u}^{(u)}(y) = \begin{cases} 1 - \left(1 + \xi \frac{y}{\sigma_u}\right)^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp\left(\frac{-y}{\sigma_u}\right), & \xi = 0 \end{cases}, \quad y > 0.$$

$\sigma_u$  je parametr měřítka, je vždy kladný a popisuje rozptýlenost v závislosti na zvolené výšce prahu  $u$ . Proto je důležité hodnotu prahu správně nastavit.  $\xi$  je parametr tvaru či typu. Tento parametr je invariantní vzhledem k zvolené výšce prahu  $u$ . Podle hodnoty, které nabývá, se jedná o jedno ze tří rozdělení:

- $\xi > 0$ : Paretovo rozdělení;
- $\xi = 0$ : Exponencionální rozdělení;
- $\xi < 0$ : Beta rozdělení.

Metoda excedentů přes vysoký práh považuje za extrémní všechny ty hodnoty, které překročí určitý práh  $u$ . Na rozdíl od metody blokových maxim, tak jde o extrémy vzhledem k celému souboru. I v tomto případě nastává obdobná otázka, jako u metody blokových maxim a to, jakou výšku prahu zvolit. Pokud by byl práh příliš vysoký, pak je generováno pouze pár extrémních hodnot, čímž se zvyšuje variabilita odhadu parametrů distribuční funkce, která by tak pravděpodobně nedostatečně popisovala empirická data. Kdyby naopak byl práh zvolen příliš nízko, generovaly by se excedenty, které by se daly považovat spíše za běžné než za extrémní, čímž může dojít k porušení vlastnosti asymptotiky distribuce. V další části se tak budeme věnovat výběru vhodného prahu na základě grafických metod.

### 2.2.1. Grafické metody výběru prahu

Jako podkladový materiál pro volbu vhodné prahové hodnoty  $u$  lze využít následující grafy: Mean residual life plot, Threshold choice plot a L-moments plot. Následující část práce se těmito grafy zabývá podrobněji.



## Mean residual life plot

Tento typ grafu pro identifikování vhodného prahu  $u$  při modelování extrémních hodnot rozdělením GPD pracuje s jeho střední hodnotou.

Předpokládejme, že  $Y$  je náhodná veličina se zobecněným Paretovým rozdělením  $GPD(u_0, \sigma_{u_0}, \xi)$  a  $u_1$  je práh, pro který platí  $u_1 > u_0$ . Potom náhodná proměnná  $Y - u_1 | Y > u_1$  má rovněž zobecněné Pareto rozdělení, ale s parametry  $u_1, \sigma_{u_1} = \sigma_{u_0} + \xi u_1$  a  $\xi_1 = \xi$ . Za předpokladu, že  $y > 0$ , lze dojít k následujícímu vztahu:

$$\begin{aligned} P(Y - u_1 > y | Y > u_1) &= \frac{1 - H^{(u_0)}(y + u_1)}{1 - H^{(u_0)}(u_1)} = \frac{\left(1 + \xi \frac{y + u_1}{\sigma_{u_0}}\right)_+^{-1/\xi}}{\left(1 + \xi \frac{u_1}{\sigma_{u_0}}\right)_+^{-1/\xi}} = \\ &= \left(1 + \xi \frac{y}{\sigma_{u_0} + \xi u_1}\right)_+^{-1/\xi}, \end{aligned}$$

kde  $z_+ = \max(0, z)$ . Střední hodnota náhodné veličiny  $Y$  se zobecněným Paretovým rozdělením  $GPD(\mu, \sigma, \xi)$  je pro parametr tvaru  $\xi < 1$  definována výrazem:

$$E[Y] = \mu + \frac{\sigma}{1 - \xi}.$$

Pro  $\xi \geq 1$  je střední hodnota nekonečno. Za předpokladu, že excedenty proměnné  $Y$  přesahující hodnotu prahu  $u_0$  pochází ze zobecněného Pareto rozdělení, platí, že jejich střední hodnota má tvar:

$$E[Y - u_0 | Y > u_0] = \frac{\sigma_{u_0}}{1 - \xi}.$$

Pokud lze excedenty přes práh  $u_0$  aproximovat rozdělením GPD, pak GPD musí být vhodným rozdělením i pro excedenty přesahující práh  $u_1$ , kde  $u_1 > u_0$ . Tímto dojde k nárůstu škálového parametru a střední hodnota excedentů bude mít tvar:

$$E[Y - u_1 | Y > u_1] = \frac{\sigma_{u_1}}{1 - \xi} = \frac{\sigma_{u_0} + \xi u_1}{1 - \xi}.$$

Z toho plyne, že pro  $u > u_0$  je  $E[Y - u|Y > u_0]$  lineární funkcí proměnné  $u$ . Navíc teoretickou střední hodnotu excedentů přes práh  $u$ , tj.  $E[Y - u|Y > u_0]$  lze snadno odhadnout.

Mean residual life plot pak tvoří body

$$\left\{ \left( u, \frac{1}{k} \sum_{i=1}^k (y_i - u) \right) : u \leq x_{max} \right\},$$

kde  $k$  je počet pozorování překračující hodnotu prahu  $u$  a  $x_{max}$  je nejvyšší hodnota ze všech pozorování. Z grafu pak za  $u_0$  označíme takovou hodnotu, od které se mean residual life plot chová lineárně.

### Threshold choice plot

Tato metoda je založena na stabilitě odhadů parametrů rozdělení GPD. Nechť  $Y$  je náhodná veličina z  $GPD(u_0, \sigma_{u_0}, \xi)$  rozdělení a  $u_1 > u_0$ . Potom náhodná veličina  $Y|Y > u_1$  má rovněž GPD rozdělení, a to s parametry ve tvaru  $u_1$ ,  $\sigma_{u_1} = \sigma_{u_0} + \xi(u_1 - u_0)$  a  $\xi_1 = \xi$ .

Zavedeme nyní nový parametr  $\sigma^* = \sigma_u - \xi u = \sigma_{u_0} - \xi u_0$ , který je nezávislý na hodnotě  $u$ . V tomto případě lze očekávat, že odhady parametrů  $\sigma^*$  a  $\xi$  budou pro všechny prahy větší než  $u_0$  stejné. Threshold choice plot je tak reprezentován body

$$\{(u, \sigma^*) : u \leq x_{max}\} \quad a \quad \{(u, \xi) : u \leq x_{max}\}.$$

Z grafu pak za  $u_0$  označíme takovou hodnotu, od které má threshold choice plot konstantní průběh.

### L-moments plot

Další variantou výběru vhodné výšky prahu je L-moments plot, který je založen L-momentech. Ty jsou přesnější v odhadech parametrů při malém počtu dat a robustnější vůči odlehlým pozorováním než metoda maximální věrohodnosti. Navíc dokáže více způsoby charakterizovat distribuční funkci.

**Definice 2.2.1.** *Nechť  $X$  je náhodná proměnná s distribuční funkcí  $F$  a nechť  $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$  jsou její pořádkové statistiky, což je vzestupně uspořádaný náhodný výběr o velikosti  $n$ , z pravděpodobnostního rozdělení náhodné veličiny  $X$ . Pak L-momenty náhodné veličiny  $X$  definujeme:*

$$\lambda_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} EX_{r-k:r}, \quad r = 1, 2, \dots, n.$$

L-momenty  $\lambda_r$ ,  $r \geq 3$  jsou často standardizovány, proto je vhodné definovat také L-momentové poměry náhodné veličiny  $X$ :

$$\tau_r = \frac{\lambda_r}{\lambda_2}, \quad r = 3, 4, \dots, n.$$

L-momenty  $\lambda_1, \dots, \lambda_r$  společně s L-momentovými poměry představují nástroj pro popis distribuce. L-moment  $\lambda_1$  vyjadřuje míru polohy,  $\lambda_2$  míru variability,  $\tau_3$  míru šikmosti a  $\tau_4$  míru špičatosti náhodné proměnné  $X$ . L-šikmost a L-špičatost požadovaného rozdělení GPD jsou dány ve tvaru:

$$\tau_3 = \frac{1 + \xi}{3 - \xi},$$

$$\tau_4 = \frac{(1 + \xi)(2 + \xi)}{(3 - \xi)(4 - \xi)}.$$

L-špičatost lze vyjádřit pomocí L-šikmosti takto:

$$\tau_4 = \tau_3 \frac{1 + 5\tau_3}{5 + \tau_3}.$$

Tento předpis definuje teoretickou křivku v L-moment plotu. Ta je zobrazena společně s body definovanými vztahem:

$$\{(\hat{\tau}_{3,u}, \hat{\tau}_{4,u}) : u \leq x_{max}\},$$

kde  $\hat{\tau}_{3,u}$  a  $\hat{\tau}_{4,u}$  jsou odhady L-šikmosti a L-špičatosti v závislosti na prahu  $u$ . Výběr vhodného prahu je pak založen na pozici odhadnutých bodů vůči teoretické křivce.

## 2.2.2. Odhady parametrů distribuční funkce

Již bylo uvedeno, že pro modelování extrémních hodnot existují vhodné distribuce, které vychází z asymptotické teorie. Pro hodnoty přesahující vysoký práh je vhodným rozdělením zobecněné Paretovo rozdělení s parametrem měřítka  $\sigma$  a parametrem tvaru  $\xi$ . Mezi hojně používané metody odhadu parametrů patří metoda maximální věrohodnosti a metoda vážených momentů. Ačkoliv má metoda maximální věrohodnosti spoustu výhodných vlastností, tak při malém vzorku dat ( $n \leq 15$ ) se jeví, že metoda vážených momentů dosahuje lepších výsledků, co se biasu a střední kvadratické chyby týče.

Při modelování prahových hodnot se lze velmi často setkat s malým datovým vzorkem. Dále je proto uvedena i modifikovaná metoda maximální věrohodnosti, která zachovává výhody metody maximální věrohodnosti a zároveň dosahuje lepších odhadů při malém počtu pozorování. Neprve ale představíme metodu maximální věrohodnosti a metodu vážených momentů.

### Metoda maximální věrohodnosti

Metoda maximální věrohodnosti (Maximum likelihood) slouží k nalezení bodového odhadu parametru, který je ve smyslu pozorovaných dat nejvíce pravděpodobný. Mezi vlastnosti této metody patří konzistence, asymptotická normalita, asymptotická nevychýlenost, asymptotická eficeence a invariance.

Uvažujme nyní náhodný výběr  $\mathbf{X} = (X_1, \dots, X_n)$  z rozdělení s hustotou pravděpodobnosti  $f(x; \boldsymbol{\theta})$ , kde  $\boldsymbol{\theta}$  je vektor neznámých parametrů z parametrického prostoru  $\Theta \subset \mathbb{R}^k$  a  $\mathbf{x}$  je realizace náhodného výběru. Pak sdružená hustota odpovídající  $\mathbf{x}$  má tvar:

$$f(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^n f(x_i; \boldsymbol{\theta}).$$

Odhad metodou maximální věrohodnosti je založen na myšlence, že na sdruženou hustotu lze pohlížet i jako na funkci vektoru  $\boldsymbol{\theta}$  při pevných hodnotách  $\mathbf{x}$ . Takovou funkci nazýváme funkcí věrohodnostní a značíme ji  $L(\boldsymbol{\theta}|\mathbf{x})$ , resp.  $L(\mathbf{x};\boldsymbol{\theta})$

nebo  $L(\boldsymbol{\theta})$ , a platí pro ni:

$$L(\boldsymbol{\theta}|\mathbf{x}) = \prod_{i=1}^n f(x_i; \boldsymbol{\theta}).$$

Číselný vektor  $\hat{\boldsymbol{\theta}}$ , který maximalizuje věrohodnostní funkci  $L(\boldsymbol{\theta})$ , pak nazýváme maximálně věrohodným odhadem vektoru  $\boldsymbol{\theta}$ . Z praktických důvodů se někdy místo věrohodnostní funkce pracuje s jejím logaritmem, s tzv. logaritmickou věrohodnotní funkcí:

$$l(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta}) = \sum_{i=1}^n \ln f(x_i; \boldsymbol{\theta}).$$

Následující věta z [1] popisuje konvergenci maximálně věrohodného odhadu  $\hat{\boldsymbol{\theta}}$  ke skutečné hodnotě parametru, za předpokladu, že je jednorozměrný, tj.  $\boldsymbol{\theta} = \theta$ .

**Věta 2.2.1.** *Nechť jsou splněny předpoklady pro odhad jednorozměrného parametru  $\theta$ :*

- *Nechť  $\Theta$  je parametrický prostor, který obsahuje takový neprázdný otevřený interval  $\omega$  tak, že skutečná hodnota parametru  $\theta_0$  patří do  $\omega$ .*
- *Nechť  $\mathbf{X} = (X_1, \dots, X_n)'$ , kde  $X_i$  jsou nezávislé stejně rozdělené náhodné veličiny s hustotou  $f(x, \theta)$  vzhledem k nějaké  $\sigma$ -konečné míře  $\mu$ .*
- *Nechť  $M = \{x : f(x, \theta) > 0\}$  nezávisí na  $\theta$ .*
- *Nechť  $\theta_1, \theta_2 \in \Theta$ . Pak  $f(x, \theta_1) = f(x, \theta_2)$  pro skoro všechna  $x$  platí právě tehdy, je-li  $\theta_1 = \theta_2$ .*

*Nechť dále na intervalu  $\omega$  existuje  $f'(x, \theta) = \frac{\partial f(x, \theta)}{\partial \theta}$  pro skoro všechna  $x$ . Pak pro každé  $\epsilon > 0$  při  $n \rightarrow \infty$  platí, že s pravděpodobností konvergující k jedné má věrohodnostní rovnice  $\frac{\partial L(\mathbf{X}, \theta)}{\partial \theta} = 0$  takový kořen  $\hat{\theta}_n = \hat{\theta}_n(\mathbf{X})$ , že  $|\hat{\theta}_n - \theta_0| < \epsilon$ .*

Předchozí větu lze zobecnit i pro případ odhadu vektoru parametrů. Předpokládejme nyní, že  $y_1, \dots, y_k$  jsou extrémní hodnoty překračující práh  $u$ , tj. jsou to

takové hodnoty, pro které platí, že  $y_j = x_i$  pro  $x_i > u$ , kde  $i \in \{1, \dots, n\}$ . A necht' excedenty přes práh  $u$  mají tvar  $z_j = y_j - u$ , kde  $j = 1, \dots, k$  a pochází ze zobecněného Paretova rozdělení. Pak věrohodnostní funkci lze zapsat ve tvaru:

$$L(\sigma_u, \xi) = \prod_{j=1}^k \frac{dH^{(u)}(z_j; \sigma_u, \xi)}{dz_j}.$$

Pro  $\xi \neq 0$  lze logaritmus věrohodnostní funkce zapsat ve tvaru

$$l(\sigma_u, \xi) = -k \log \sigma_u - \left(1 + \frac{1}{\xi}\right) \sum_{j=1}^k \log \left(1 + \xi \frac{z_j}{\sigma_u}\right)$$

v případě, že je člen  $\left(1 + \xi \frac{z_j}{\sigma_u}\right) > 0$ , pro  $i = j, \dots, k$ . V opačném případě je  $l(\sigma_u, \xi) = -\infty$ .

Pokud je  $\xi = 0$ , pak má logaritmus věrohodnostní funkce tvar:

$$l(\sigma_u) = -k \log \sigma_u - \frac{1}{\sigma_u} \sum_{j=1}^k z_j.$$

Hledané parametry  $\sigma_u$  a  $\xi$  získáme maximalizací logaritmické věrohodnostní funkce. V tomto případě nelze lokální maximum logaritmické věrohodnostní funkce hledat analyticky a je zapotřebí použít vhodnou numerickou metodu.

### Pravděpodobnostně vážené momenty

Metoda pravděpodobnostně vážených momentů (PVM) je stejně, jako klasická metoda momentů, založena na porovnání obecných PVM s jejich empirickými protějšky. Necht'  $X$  je náhodná spojitá veličina s distribuční funkcí  $F$ . Pak pravděpodobnostně váženými momenty nazýváme veličiny

$$M_{p,r,s} = E[X^p (F(X))^r (1 - F(X))^s],$$

kde parametry  $p$ ,  $r$  a  $s$  jsou reálná čísla. Pro zobecněné Paretovo rozdělení se ukazuje za vhodné zvolit parametr  $p = 1$  a  $r = 0$ . Potom pravděpodobnostně

vážené momenty existují a za předpokladu, že  $\xi < 1$ , mají tvar:

$$\alpha_s = M_{1,0,s} = \frac{\sigma_u}{(s+1)(s+1-\xi)}.$$

Při  $s=0$  a  $s=1$  jsme schopni odvodit analytické vyjádření parametrů  $\sigma_u$  a  $\xi$ , a to takto:

$$\begin{aligned}\sigma_u &= \frac{2\alpha_0\alpha_1}{\alpha_0 - 2\alpha_1}, \\ \xi &= 2 - \frac{\alpha_0}{\alpha_0 - 2\alpha_1}.\end{aligned}$$

Odhad parametrů  $\sigma_u$  a  $\xi$  metodou pravděpodobnostně vážených momentů získáme, pokud místo momentů  $\alpha_0$  a  $\alpha_1$  použijeme jejich odhady. Takovým nevychýleným odhadem je

$$\alpha_s = \frac{1}{k} \sum_{j=1}^k \frac{(k-j)(k-j-1)\dots(k-j-s+1)}{(k-1)(k-1)\dots(k-s)} z_{j:k},$$

kde  $z_{j:k}$  je pořádková statistika náhodného výběru velikosti  $k$  z rozdělení s distribuční funkcí  $F$ . Výhodou této metody je její jednoduchost a dobré výsledky při odhadu parametrů na základě malého rozsahu výběru.

## Penalizovaná věrohodnostní funkce

Penalizování maximální věrohodnostní funkce nám umožňuje zahrnout do odhadu parametrů skrze penalizační funkci informaci, která doplňuje poznatky z dat. Mezi nejběžněji používané patří neparametrické vyhlazování, které penalizuje nerovnosti věrohodnostní funkce. Penalizovaná věrohodnostní funkce je definována jako:

$$L_p(\sigma_u, \xi) = L(\sigma_u, \xi)P(\xi),$$

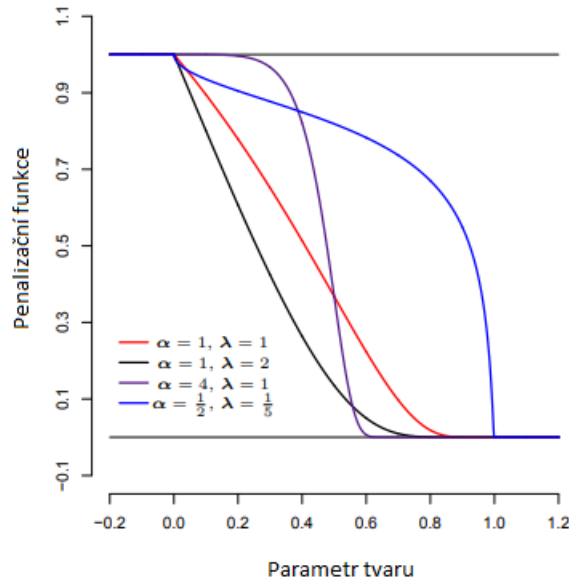
kde  $L(\sigma_u, \xi)$  je věrohodnostní funkce a  $P(\xi)$  je penalizační funkce. V této práci si představíme penalizační funkci podle Colese a Dixona uvedenou v [6], která zahrnuje do odhadu informaci, že hodnoty hledaného parametru tvaru  $\xi$  jsou menší

než jedna, a že hodnoty parametru  $\xi$  blíže k jedné, jsou méně pravděpodobné. Tato funkce má tvar:

$$P(\xi) = \begin{cases} 1, & \text{pro } \xi \leq 0, \\ \exp\left\{-\lambda\left(\frac{1}{1-\xi}-1\right)^\alpha\right\}, & \text{pro } 0 < \xi < 1, \\ 0, & \text{pro } \xi \geq 1. \end{cases}$$

Parametry  $\alpha$  a  $\lambda$  je rozumné nastavit na hodnotu jedna. Tato kombinace se na základně experimentu provedeného v [6], jeví jako nejvýhodnější. Penalizační funkce pro vybrané hodnoty  $\alpha$  a  $\lambda$  je zobrazena na obrázku 2.1.

Hodnoty  $\hat{\sigma}_u$  a  $\hat{\xi}$ , které maximalizují funkci  $L_p$ , pak představují hledané odhady pomocí penalizační věrohodnostní funkce.



Obrázek 2.1: Penalizační funkce pro vybrané hodnoty  $\alpha$  a  $\lambda$ . Zdroj: [16].

## Porovnání metod

Kvalita odhadů se často posuzuje z hlediska přesnosti odhadů extrémních kvantilů rozdělení GPD. Takovýto odhad kvantilu  $q_p$  dostaneme dosazením získa-



ných odhadů parametru  $\hat{\xi}$  a  $\hat{\sigma}$  požadovanou metodou do vzorce:

$$\hat{q}_p = \begin{cases} -\frac{\hat{\sigma}_u}{\hat{\xi}}(1 - p^{-\hat{\xi}}), & \text{pro } \hat{\xi} \neq 0, \\ -\hat{\sigma}_u \log p, & \text{pro } \hat{\xi} = 0. \end{cases}$$

V [6] byla zkoumána přesnost odhadů metodou maximální věrohodnosti a metodou pravěpodobnostně vážených momentů na základě simulací. Při těchto simulacích bylo vygenerováno tisíce vzorků různých velikostí z distribuce zobecněného extrémálního rozdělení při fixovaných hodnotách parametru polohy  $\mu$  a rozptylu  $\sigma$ , ale s různými hodnotami parametru tvaru  $\xi$ . Při výpočtech byly použity bloková maxima, ale dle [13] lze dokázat, že stejných závěrů by se dosáhlo i za použití metody excedentů přes vysoký práh.

Na základě této simulace bylo zjištěno, že při malých počtech vzorků ( $n \leq 15$ ) metoda maximální věrohodnosti, co se vchýlení a střední kvadratické chyby týče, dosahuje horších výsledků v porovnání s metodou pravěpodobnostně vážených momentů. To je patrné zejména při odhadu extrémních kvantilů  $q_p$  za předpokladu, že  $\xi$  je kladné. Příčina horšího odhadu byla vysvětlena po prozkoumání hodnot původních parametru uvažované distribuce z nichž se generovaly vzorky hodnot. Bylo zjištěno, že hlavní rozdíl mezi metodami je patrný z distribuce odhadu parametru  $\xi$ , která je v případě aplikování metody maximální věrohodnosti pozitivně vychýlená. Navíc kvantil  $q_p$  je nelineární funkcí parametru  $\xi$ , a proto i malé pozitivní vychýlení od skutečné hodnoty parametru  $\xi$  vede k podstatnému vychýlení od skutečné hodnoty kvantilu.

Dále je rovněž uvedeno, že rozdílný výkon odhadu metod může být způsoben rozdílnými předpoklady. V případě metody pravěpodobnostně vážených momentů se předpokládá, že  $\xi < 1$ , čímž se redukuje parametrický prostor odhadu z  $(-\infty, \infty)$  na  $(-\infty, 1)$ . V důsledku toho se zmenšuje i výběrový rozptyl odhadu parametru  $\xi$ . Dále, s rostoucí hodnotou  $\xi$  se zvětšuje negativní vychýlení jeho odhadu. Pokud se podíváme na definici kvantilu, jako na nelineární funkci proměnné  $\xi$ , je evidentní, že podhodnocením odhadu dochází k menší penalizaci, co se střední kvadratické chyby týče, než při nadhodnocení odhadu  $\xi$ . Jinak řečeno, metoda pravěpodobnostně vážených momentů představuje určitý kom-

promis mezi vychýlením a rozptylem odhadu parametru  $\xi$ . V důsledku toho se u distribuční funkce extrémních kvantilů  $q_p$ , která je vytvořena na základě odhadů metodou pravěpodobnostně vážených momentů, nevyskytují těžké konce, tak jako u distribuce vytvořené na základě metody věrohodnostní funkce.

Aby bylo možné obě metody objektivně porovnat, je zapotřebí omezit parametr tvaru na hodnoty  $\xi < 1$  i při odhadu metodou maximální věrohodnosti. Jedním z řešení je sestavení vhodné penalizační funkce, která by obsahovala informaci o požadované struktuře  $\xi$ . Na této myšlence je založena již zmíněná metoda penalizované věrohodnostní funkce.

## 2.3. Ověření modelu

Následující část práce se zabývá dvěma rozdílnými způsoby, jak ověřit kvalitu sestaveného modelu založeném na zobecněném Paretově rozdělení. Konkrétně se jedná o ověření sestaveného modelu pomocí grafických a statistických metod.

### 2.3.1. Grafické ověření sestaveného modelu

Grafické ověření neposkytuje jednoznačnou odpověď, zda zvolený model odpovídá datům. Graf slouží pouze jako vizuální podklad při rozhodování, který ale nedá jasnou odpověď. Konečné rozhodnutí, zda je model zvolen správně či nikoliv, záleží na pohledu, zkušenosti a znalosti hodnotitele. Ačkoliv tyto metody neposkytují jednoznačnou odpověď, lze na jejich základě učinit rozbor možné příčiny nesouladu s navrženým modelem. Navíc ověřování modelu na základě grafických metod je citlivější než ověřování pomocí testů, kde schopnost detekovat nevhodný model závisí na síle testu. Proto je vhodné při analýze přihlížet k závěru obou metod. Dále jsou uvedeny některé grafické metody, které lze při analýze použít.

#### Probability plot

Tato metoda porovnává hodnoty empirické distribuční funkce s hodnotami teoretické distribuční funkce. Necht'  $z_{1:k} \leq z_{2:k} \leq \dots \leq z_{k:k}$  je  $k$  vzestupně

uspořádaných excedentů  $z_1, \dots, z_k$  přes práh  $u$ .  $z_1, \dots, z_k$  jsou tedy takové hodnoty, které lze definovat vztahem  $z_j = x_j - u$ , kde  $j \in \{1, \dots, k\}$  je index odpovídající hodnotě  $x_i$ , která splňuje, že  $x_i > u$  a kde  $x_1, \dots, x_n$  jsou napozorovaná data. Empirickou distribuční funkci lze definovat vztahem:

$$\tilde{H}^{(u)}(z_{j:k}) = \frac{1}{k} \sum_{i=1}^k \mathbb{I}(z_{i:k} \leq z_{j:k}) = \frac{j}{k}$$

a teoretickou distribuční funkci lze sestavit dosazením odhadnutých parametrů do vzorce pro rozdělení GPD, čímž získáme:

$$\hat{H}^{(u)}(z_{j:k}) = 1 - \left(1 + \hat{\xi} \frac{z_{j:k}}{\hat{\sigma}_u}\right)^{-\frac{1}{\hat{\xi}}}.$$

Probability plot pak tvoří body

$$\{(\tilde{H}^{(u)}(z_{j:k}), \hat{H}^{(u)}(z_{j:k})); j = 1, \dots, k\}.$$

Pokud je navržený model správný, pak jsou hodnoty  $\tilde{H}^{(u)}(z_{j:k})$  a  $\hat{H}^{(u)}(z_{j:k})$  pro každé  $j$  téměř shodné a body by tak měly ležet na přímce, blízké ose prvního kvadrantu. Jakékoliv odchýlení poukazuje na nedostatky modelu. Nevýhodou této metody je, že s rostoucí hodnotou  $j$  se hodnoty empirické i teoretické funkce blíží jedné. O tom jak model sedí na velké hodnoty, tak máme méně informací.

### Quantile-Quantile plot

Quantile-Quantile plot (Q-Q plot) představuje alternativu k probability plotu, která se snaží odstranit jeho nedostatky. Tento graf vykresluje proti sobě hodnoty kvantilů teoretického a empirického rozdělení. Q-Q plot pak představují body

$$\left\{ \left( \left( \hat{H}^{(u)} \right)^{-1} \left( \frac{j}{k} \right), z_{j:k} \right); j = 1, \dots, k \right\},$$

kde

$$\left( \hat{H}^{(u)} \right)^{-1} \left( \frac{j}{k} \right) = -\frac{\hat{\sigma}_u}{\hat{\xi}} \left( 1 - \left( 1 - \frac{j}{k} \right)^{-\hat{\xi}} \right).$$

Tak jako u probability plotu se body v případě nevhodného modelu odchylují od osy prvního kvadrantu.

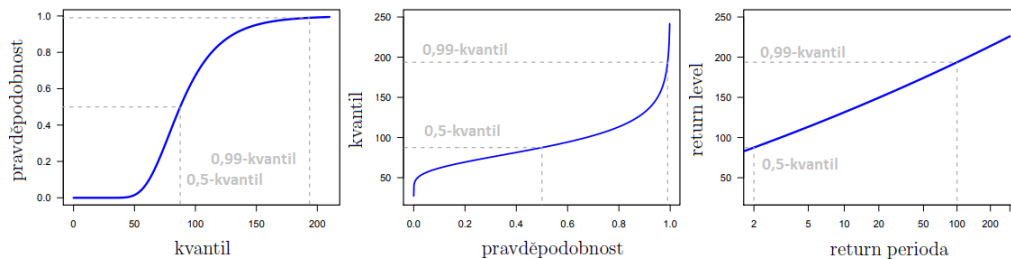
## Return level plot

Return level plot se obzvláště hodí pro ověření vhodnosti modelu extrémních hodnot. Konec distribuční funkce v takovém případě bývá nahuštěný. Return level plot je tvořen křivkou definovanou jako:

$$\{(\log y_p, \hat{q}_p); 0 < p < 1\},$$

kde  $\hat{q}_p$  je odhad kvantilu  $q_p$  a  $y_p$  je definován jako  $y_p = -\log(1 - p)$ . Do grafu se proti  $y_p$  rovněž vykreslují i hodnoty empirických kvantilů reprezentující data. Ty jsou reprezentovány body. Jakákoliv odchylka bodů od křivky poukazuje na nedostatky modelu. Tento graf lze často vidět společně s vykresleným konfidenčním intervalem pro zvýšení informativnosti.

Z tohoto grafu lze rovněž vyčíst, jaká je pravděpodobnost, že daná hodnota bude překročena. Pokud by vodorovná osa představovala return period, tj. stručně řečeno průměrnou dobu, za kterou dojde k opakování události. A svislá osa by představovala return level (totéž jako kvantil), což označuje maximální hodnotu, ke které během return periody dochází. Pak lze popsat vztah mezi distribuční funkcí, kvantilovou funkcí a return level plotem pomocí schématu na obrázku 2.2.



Obrázek 2.2: Vztah mezi distribuční funkcí, kvantilovou funkcí a return level plotem. Zdroj: [10].

## Density plot

Poslední graf, o kterém tato část pojednává, je density plot, který narozdíl od již dříve zmíněných neporovnává teoretickou distribuci s empirickou. V tomto případě graf porovnává hustotu navrženého modelu s normovaným histogramem sestaveného z napozorovaných hodnot. Tato metoda se ovšem jeví jako méně informativní, neboť je závislá na velikosti zvolených intervalů histogramu.

### 2.3.2. Statistické ověření sestaveného modelu

Při statistickém ověření modelu se používají testové statistiky, které jsou založeny na porovnání empirické a distribuční funkce. Narozdíl od grafické analýzy se lze, na základě hodnoty testové statistiky jednoznačně rozhodnout na zvolené hladině významnosti  $\alpha$ , zda je sestavený model správný. V této práci si blíže představíme tři testové statistiky.

#### Kolmogorovův-Smirnovův test

Tato metoda patří mezi neparametrické testy a je založena na měření maximální vzdálenosti empirické a teoretické distribuční funkce. Definujme nyní:

$$D^+ = \max_{1 \leq j \leq k} \left\{ \frac{j}{k} - \hat{H}^{(u)}(z_{j:k}) \right\},$$
$$D^- = \max_{1 \leq j \leq k} \left\{ \hat{H}^{(u)}(z_{j:k}) - \frac{j-1}{k} \right\},$$
$$D = \max\{D^+, D^-\},$$

kde  $\hat{H}^{(u)}$  je teoretická distribuční funkce,  $z_{1:k} \leq z_{2:k} \leq \dots \leq z_{k:k}$  je  $k$  vzestupně uspořádaných extrémních hodnot přes práh  $u$  a  $D$  je testová statistika. Nulovou hypotézu o shodě empirické a teoretické distribuční funkce zamítáme na hladině významnosti  $\alpha$ , pokud je hodnota testové statistiky  $D$  větší než příslušná tabulovaná kritická hodnota.

### Cramer-von Mises test

Tento test je založen na testové statistice ve tvaru:

$$W^2 = \frac{1}{12k} + \sum_{j=1}^k \left( \hat{H}^{(u)}(z_{j:k}) - \frac{2j-1}{2k} \right)^2,$$

která se rovněž porovnává s příslušnou tabelovanou hodnotou.

### Anderson-Darlingův test

Tento test dává větší váhu koncům distribuční funkce než již zmíněný Kolmogorovův-Smirnovův test a je definován předpisem:

$$A^2 = -\frac{1}{k} \sum_{j=1}^k (2j-1) (\log \hat{H}^{(u)}(z_{j:k}) + \log \hat{H}^{(u)}(z_{(k+1-j):k})) - k.$$

Kritické hodnoty u předešlých metod jsou tabelovány. Tato metoda využívá k výpočtům kritických hodnot distribuční funkci, díky čemuž je test více citlivý. Na druhou stranu je však nutné pro každou testovanou distribuci kritické hodnoty modifikovat.

### Porovnání testových statistik

Z historického hlediska se nejvíce používal Kolmogorovův-Smirnovův test, ale vzhledem k jeho malé síle testu se nyní v praxi nejvíce používá citlivější Anderson-Darlingův test, popřípadě Cramer-von Misesův test.

## Kapitola 3

# Modelování výše škod pro stanovení netto zajistného v WXL-R zajištění

Čtenář již byl seznámen s teorií dané problematiky. Tato kapitola se pokouší na základě reálných dat nabyté znalosti využít při sestavení vhodného modelu potřebného pro výpočet ceny netto zajistného ve WXL-R zajištění. Čerpáno je především z [5] a [14]. Součástí kapitoly jsou vybrané příkazy ze statistického programu *R*, jenž byly během této analýzy použity. Celý zdrojový kód s výpočty z této kapitoly je pak dostupný v příloženém CD.

V úvodu této kapitoly je čtenář nejprve seznámen s daty. K tomu je využita popisná statistika. Dále jsou představeny dva pohledy, jakými lze daná data modelovat, přičemž cílem je najít model, který by dobře popisoval především pravděpodobnostní rozdělení vysokých hodnot napozorovaných dat, neboť takové hodnoty jsou důležité při kalkulaci ceny WXL-R zajistného.

Jako první je představen přístup, při kterém je model sestavený na základě všech dostupných hodnot dané proměnné. Jako druhý je představen přístup, který se pokouší najít vhodný práh, přes který se napozorované hodnoty modelují zobecněným Paretovým rozdělením. K odhadu parametrů tohoto rozdělení je využita metoda maximální věrohodnosti, metoda penalizované maximální věrohodnosti a metoda pravděpodobnostně vážených momentů. V závěru je zhodnoceno, který z přístupů je vhodnější při modelování extrémních (vysokých) hodnot. Oba

modely odpovídající těmto přístupům jsou následně využity ke kalkulaci ceny netto zajistného v WXL-R zajištění.

### 3.1. Seznámení s daty

Pro praktickou část této práce jsme zvolili datovou sadu, která popisuje celkovou výši nároku zákazníka při pojištění vozidla ve státech Iowa, Kansas, Missouri, Nebraska a Oklahoma za rok 2011. Datový set zohledňuje i jiné faktory, přičemž celkově má 9 134 pozorování a 26 sloupců:

*Customer, Country, State.Code, State, Claim.Amount, Response, Coverage, Education, Effective.To.Date, EmploymentStatus, Gender, Income, Location.Code, Marital.Status, Monthly.Premium.Auto, Number.of.Policies, Months.Since.Last.Claim, Months.Since.Policy.Inception, Policy.Type, Policy, Number.of.Open.Complaints, Claim.Reason, Sales.Channel, Total.Claim.Amount, Vehicle.Class, Vehicle.Size.* Tato data jsou volně dostupná na internetové stránce [12]. Vzhledem k cíli této práce byl k analýze použit pouze sloupec *Total.Claim.Amount* (dále v textu označeno zkráceně jako *claims*, data nebo datový soubor).

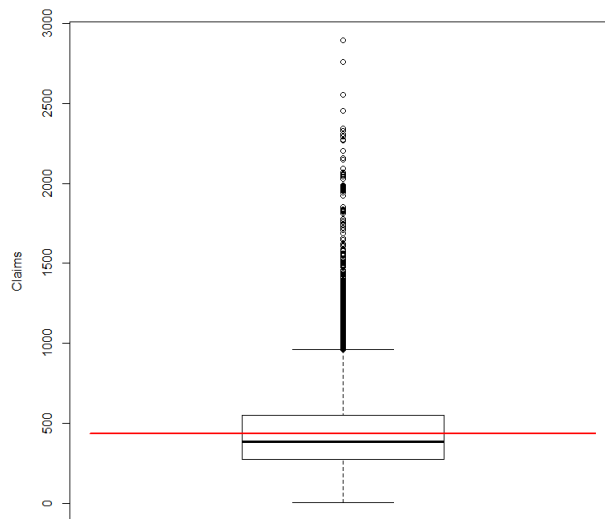
Hodnoty *claims* se pohybují od hodnoty 0,099 do 2 893,240 s celkovou hodnotou nároků ve výši 3 964 967,05, přičemž horní kvartil je roven hodnotě 547,515. Pro pozorování jsou tedy typičtější nižší hodnoty, v porovnání s rozsahem souboru hodnot. Dále je v tabulce 3.1 uveden výstup z programu *R* s přehledem základních charakteristik (minimum, dolní kvartil, medián, průměr, horní kvartil, maximum, směrodatná odchylka, šikmost). Data jsou rovněž pro lepší představu

Tabulka 3.1: Přehled základních charakteristik: minimum (Min.), dolní kvartil (Q1), medián, průměr, horní kvartil (Q3), maximum (Max.), směrodatná odchylka (sd), šikmost.

Min.	Q1	Medián	Průměr	Q3	Max.	sd	Šikmost
0,099	272,26	383,95	434,09	547,515	2 893,24	290,5	1,715



vizualizována graficky. Na obrázku 3.1 je k nahlédnutí boxplot, kde červená horizontální linka představuje průměr a černá, tučně zvýrazněná medián. Z boxplotu je rovněž patrné, že data obsahují odlehlé hodnoty. Na základě těchto dat se pokusíme stanovit cenu netto zajistného v WXL-R zajištění s prioritou  $a = 2\,000$ .



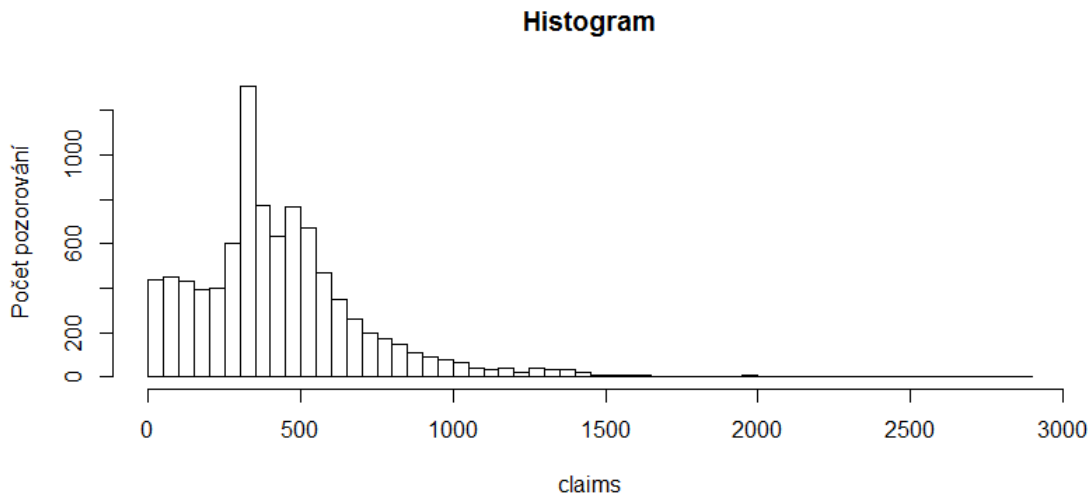
Obrázek 3.1: Boxplot pro hodnoty *claims*.

## 3.2. Modelování výše škod na základě celého datového souboru

V předešlé části jsme si představili základní charakteristiky datového souboru. Nyní bude naším cílem najít vhodný model, který by věrohodně popisoval (vysoké) extrémní hodnoty datového souboru. Vytvoříme proto nejprve model, který by dobře reprezentoval celková data s tím, že náš hlavní zájem je zaměřen na to, jak dobře aproximuje koncové hodnoty.

Ve statistice se nejčastěji setkáme s normálním rozdělením, ke kterému by jsme se rádi přiblížili. V tomto případě se ale, dle histogramu 3.2, nejedná o data z to-

hoto rozdělení. Ani při logaritmování dat histogram nevykazoval tvar pocházející z normálního rozdělení. Data se proto pokusíme modelovat gamma rozdělením. V programu *R* jsme pro odhad parametrů gamma rozdělení metodou maximální

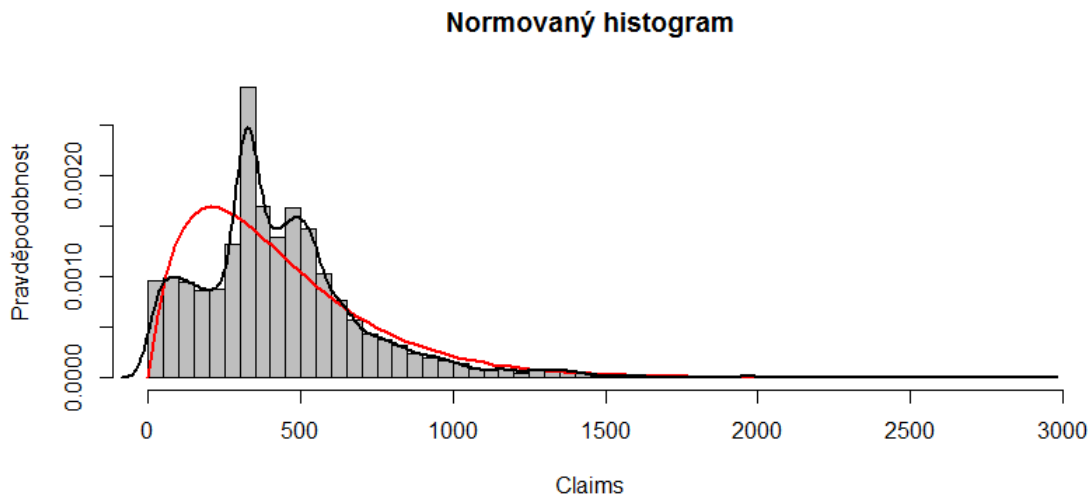


Obrázek 3.2: Histogram pro hodnoty *claims*.

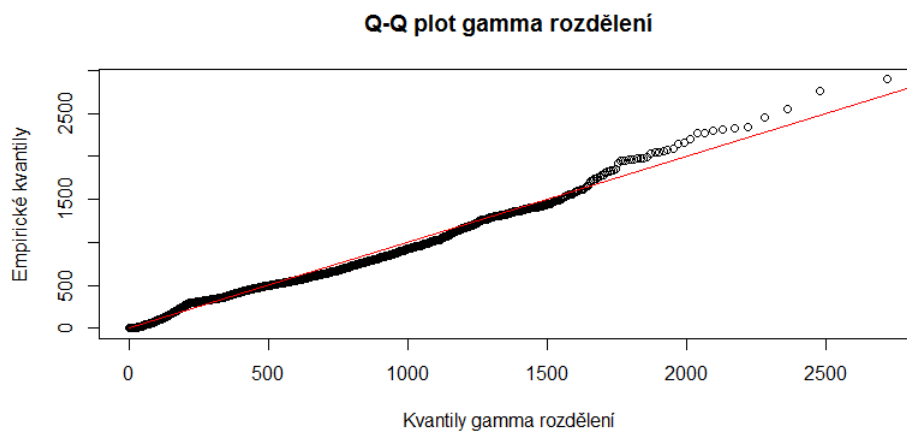
věrohodnosti použili funkci *fitdistr()* z knihovny *MASS*. Na základě tohoto odhadu parametrů byl vytvořen obrázek 3.3, kde je vyzobrazen histogram upravený tak, aby plocha pod křivkou byla rovna jedné. Histogram lemuje černá křivka, která odpovídá empirickému rozdělení. Společně s touto křivkou je v grafu zobrazena i křivka červená, reprezentující hustotu gamma rozdělení s odhadnutým parametrem měřítka rovný hodnotě 226,3495 a s odhadnutým parametrem tvaru rovný hodnotě 1,9178. Červená křivka odpovídající hustotě gamma rozdělení při koncových hodnotách zhruba kopíruje černou empirickou křivku, čili z tohoto pohledu nelze říci, že by daný model vyloženě nevyhovoval.

Pro grafickou diagnostiku správnosti modelu je lepší použít Q-Q plot, který pro daná data najdeme na obrázku 3.4. Na něm je patrné, že až na koncové hodnoty, body kopírují červeně označenou osu prvního kvadrantu. Zdá se tedy, že model celkově sedí na data uspokojivě. Lze dokonce říci, že přibližně do hodnoty 1 600 model aproximuje data dobře. Avšak při krajních vysokých hodnotách mo-

del selhává. Pokud je naším cílem modelovat data pro výpočet ceny zajištění, kde hlavním zájmem je co nejlépe odhadnout nastání vysokých hodnot, je evidentní, že tento model není zcela ideální. Jako řešení se jeví, zaměřit se zvláště na tu část dat, která nás především zajímá, a tak dosáhnout přesnějšího popisu chování krajních vysokých hodnot.



Obrázek 3.3: Normovaný histogram hodnot *claims* společně s hustotou gamma rozdělení  $\Gamma(1, 9178; 226, 3495)$ .



Obrázek 3.4: Q-Q plot z gamma rozdělení  $\Gamma(1, 9178; 226, 3495)$ .

## 3.3. Modelování výše škod na základě koncových hodnot datového souboru

Připomeňme si, že v našem zájmu je, co nejlépe modelovat především (vysoké) extrémní hodnoty. V této části práce už nebudeme modelovat celá data, jako v předešlém případě, ale pokusíme se najít vhodnou prahovou hodnotu  $u$ , která by definovala extrémní hodnoty. Hodnoty, které přesáhnou práh  $u$  budeme následně modelovat zobecněným Paretovým rozdělením.

### 3.3.1. Výběr vhodného prahu

Pro výběr vhodného prahu využijeme následující grafy z teoretické části práce: mean residual plot, threshold choice plot a L-moments plot. Detailněji si jednotlivé grafy aplikované na data *claims* rozebereme, přičemž na konci této části se pokusíme na základě všech poznatků z těchto grafů vybrat neoptimalnější práh.

#### Mean residual plot

Tak jako v teoretické části si jako první představíme mean residual plot, který jsme v programu *R* vykreslili pomocí příkazu `mrl.plot(claims, conf = 0.95)`. Jedná se o grafickou metodu výběru prahu, kdy vodorovná osa představuje zvolenou prahovou hodnotu  $u$  a svislá osa představuje výběrový průměr z kladných hodnot, které popisují o kolik dané pozorování přesáhlo daný práh  $u$  (Mean excess). Tento graf je zobrazen na obrázku 3.5 společně s 95% konfidenčním intervalem (červená křivka). Za optimální práh  $u$  se považuje taková hodnota, od které je závislost v grafu přibližně lineární. Dále je třeba brát v potaz počet pozorování, které přesahují práh, neboť s klesajícím počtem pozorování se zvyšuje šířka konfidenčního intervalu. Čtenáře upozorňujeme, že volba prahu na základě grafických podkladů je subjektivní.

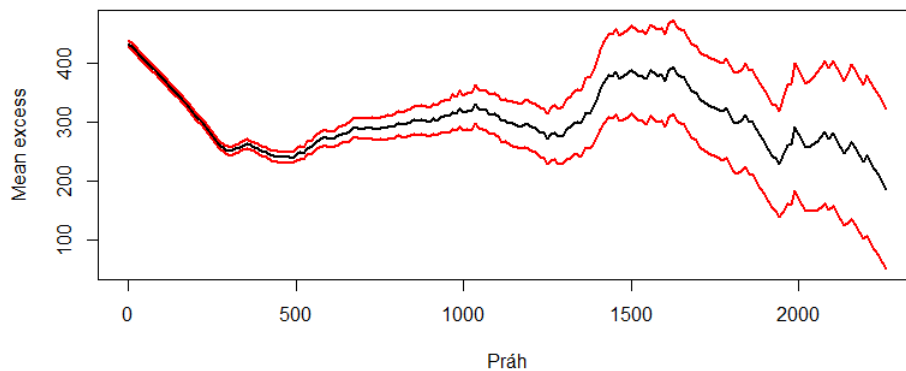
Tento typ grafu lze využít i k samotnému odhadu parametru tvaru zobecněného Paretova rozdělení, a to na základě metody vážených nejmenších čtverců, kde váhy jsou normované a odpovídají poměru počtu pozorování přesahující daný

práh ku celkovému počtu pozorování. Odhad parametru tvaru pak odpovídá směrnici přímký, která by proložila data přes zvolený práh.

Graf na obrázku 3.5 vykazuje lineární chování přibližně do hodnoty  $u = 400$ , pak se směrnice pomyslné přímký zcela mění. Od hodnoty prahu 500 se zdají být hodnoty mean excesů rovněž lineárně závislé na hodnotě prahu, a to až do hodnoty  $u = 1\ 000$ , od které trend začne mírně klesat, a to přibližně až do hodnoty  $u = 1\ 300$ . Od hodnoty prahu  $u = 1\ 500$  hodnoty mean excesů vykazují lineárně klesající průběh. Počet hodnot přesahující zmíněné prahu jsou uvedeny v tabulce 3.2.

Tabulka 3.2: Počet excedentů přes zvolený práh  $u$ .

Práh $u$	400	500	1 000	1 300	1 500
Počet excedentů	4 349	2 945	403	167	66

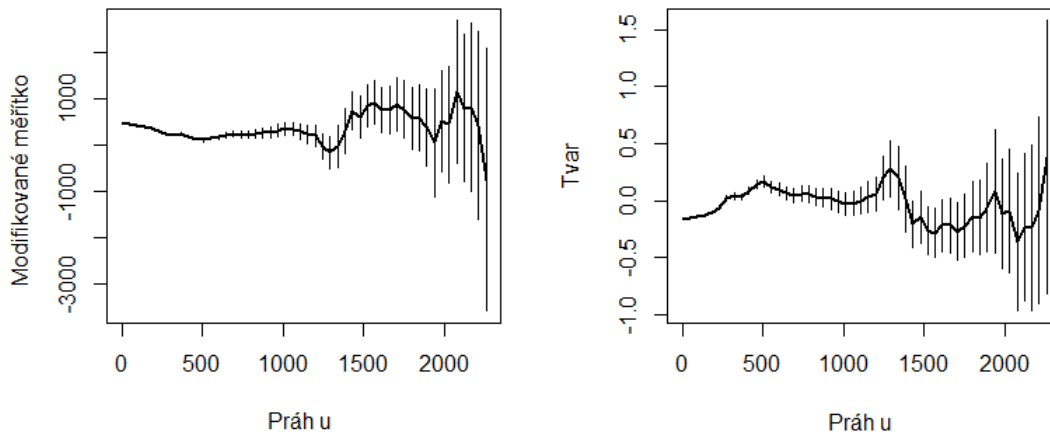


Obrázek 3.5: Mean residual plot pro hodnoty *claims*.

### Threshold choice plot

Dále si data vizualizujeme pomocí threshold choice plotu. K tomu lze využít příkaz `tcplot(claims)`, jehož výstupem jsou grafy na obrázku 3.6. Ty zobrazují empirické odhady modifikovaného parametru měřítka a tvaru, v závislosti

na volbě hodnoty prahu, společně s 95% konfidenčními intervaly. Při výběru vhodného prahu spoléháme na stabilitu parametrů. Zobecněné Paretovo rozdělení je lineární funkcí prahové hodnoty  $u$ , proto by měl parametr měřítka vykazovat lineární průběh pro hodnoty, které jsou větší než optimální práh. Graf pro parametr tvaru by měl vykazovat pro hodnoty přes optimální práh konstantní průběh. Při výběru prahu je rovněž nutné zohlednit i variabilitu odhadu parametru. Z obrázku 3.6 je patré, že průběh křivek je přibližně protichůdný, aneb



Obrázek 3.6: Threshold choice plot pro hodnoty *claims*.

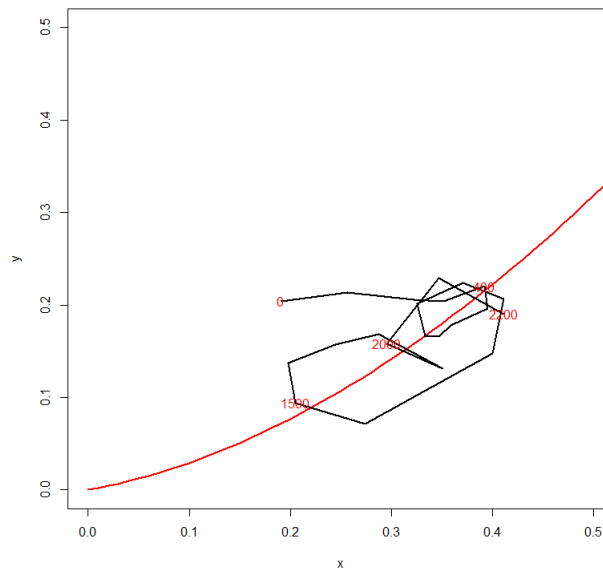
když jedna klesá, druhá roste a naopak. V obou případech tak dochází k změně trendu křivek přibližně při těchto následujících hodnotách prahu, tj.  $u = 1\,300$ ,  $u = 1\,500$ ,  $u = 1\,900$  a  $u = 2\,100$ . Povšimněme si rovněž, že průběh křivky modifikovaného měřítka je přibližně konstantní asi do hodnoty  $u = 1\,300$ , pak je průběh křivky znatelně kolisavější. S touto hodnotou prahu rovněž poměrně rychle narůstá rozptyl modifikovaného měřítka, který souvisí s počtem hodnot přesahující daný práh. Tyto počty pro zmíněné prahy jsou uvedeny v tabulce 3.3.

Tabulka 3.3: Počet excedentů přes zvolený práh  $u$ .

Práh $u$	1 300	1 500	1 900	2 100
Počet excedentů	167	66	30	13

### L-moments plot

L-moments plot aplikovaný na data *claims* je zobrazen na obrázku 3.7. Graf zobrazuje červenou křivkou teoretický vztah mezi L-šikmostí a L-špičatostí, tj.  $\tau_4 = \tau_3(1 + 5\tau_3)/(5 + \tau_3)$ , společně s černou křivkou, která představuje vztah mezi empirickými odhady L-šikmostí a L-špičatostí, které by ideálně měly červenou křivku kopírovat. Horní konec černé křivky na obrázku 3.7 představuje empirické



Obrázek 3.7: L-moments plot pro hodnoty *claims*.

odhady pro nejnižší práh a je označen číslem nula. S rostoucím prahem se křivka blíží k teoretické červené křivce, kterou přibližně při hodnotě prahu  $u = 400$  protne. Dále se jí ještě přiblíží například při hodnotách blízké prahu  $u = 1\,500$  a  $u = 2\,000$ . S hodnotou prahu  $u = 2\,200$  se černá křivka začíná zcela vzdalovat od

červené křivky k ose  $\tau_3$ . Optimální prahová hodnota by tak ideálně neměla o moc překračovat tuto zmíněnou hodnotu prahu. Počet hodnot přesahující zmíněné prahy jsou uvedeny v tabulce 3.4.

Tabulka 3.4: Počet excedentů přes zvolený práh  $u$ .

Práh $u$	400	1 500	2 000	2 200
Počet excedentů	4349	66	19	11

### Výběr prahu

Jak jsme se již zmínili, nelze na základě grafických podkladů zvolit vhodný práh jednoznačně. Výběr je subjektivní a záleží na pohledu analytika. Při výběru je třeba brát v potaz, že zvolený práh by neměl být příliš vysoký ani příliš nízký, neboť v prvním případě hrozí málo hodnot přesahující práh a v druhém případě hrozí, že by se analýza zakládala na hodnotách, které nejsou extrémní, ale spíše běžné. Dále je třeba při výběru prahu přihlížet i k následnému využití modelu. Cílem této práce je sestavit vhodný model pro výpočet ceny netto zajistného. V takovém případě se doporučuje zvolit práh podstatně menší než je hodnota zvolené priority  $a$ . Pokud vezmeme v úvahu všechny poznatky z výše uvedených grafů, přikláníme se k výběru prahu  $u = 1\,500$ .

### 3.3.2. Modelování koncových hodnot zobecněným Pareto-vým rozdělením

V předešlé části jsme se zabývali výběrem vhodného prahu  $u$ , který by nám definoval hodnotu, od které lze napozorovaná data považovat za extrémní (z hlediska vysokých hodnot). Výši přesahu těchto extrémních hodnot přes práh  $u$  se nyní pokusíme modelovat pomocí zobecněného Paretova rozdělení. V teoretické části jsme si pro odhad parametrů tohoto rozdělení (tvaru, měříta) představili tyto metody: metoda maximální věrohodnosti, metoda pravděpodobnostně vážených momentů a metoda penalizované maximální věrohodnosti. Každá metoda může



mít trochu odlišné výsledky, proto na datech vyzkoušíme všechny tři metody odhadů. Zároveň pro demonstraci budeme analyzovat i prahy, které jsme nevybraly jako nejoptimálnější, ale uvažovali jsme nad nimi, tj.  $u = 1\ 300$ ,  $u = 1\ 500$  a  $u = 1\ 700$ . Kvalitu modelů pro dané prahy a metody odhadů budeme následně ověřovat graficky.

### Odhad parametrů zobecněného Paretova rozdělení

V programu *R* je pro odhad parametrů zobecněného Paretova rozdělení dostupná funkce *fitgpd()*. Tuto funkci aplikujeme na data *claims* příkazem

$$\text{fitgpd}(\text{claims}, \text{threshold}, \text{method})\$param,$$

kde parametr *threshold* nahradíme hodnotou požadovaného prahu a místo parametru *method* udáme v uvozovkách anglicky danou zkratku metody odhadu parametrů, kterou požadujeme. Pro metodu maximální věrohodnosti je to "mle". Pro metodu pravděpodobnostně vážených momentů je to "pwmb" a pro metodu penalizované maximální věrohodnosti je to "mple". Pro takto získané odhady parametrů lze použitím příkazů

$$\text{gpd.fishape}(\text{fitted}, \text{conf}) \quad \text{a} \quad \text{gpd.fiscale}(\text{fitted}, \text{conf})$$

získat jejich konfidenční intervaly. První z příkazů odpovídá konfidenčnímu intervalu odhadu parametru tvaru, druhý pak konfidenčnímu intervalu odhadu měřítka. Parametr *fitted* odpovídá výstupu již zmíněné funkce

$$\text{fitgpd}(\text{claims}, \text{threshold}, \text{method}),$$

pro dané nastavení prahu a metody, a parametr *conf* odpovídá hodnotě zvolené pravděpodobnosti spolehlivosti daného intervalu. V této práci jsme zvolili 95% konfidenční interval pro všechny kombinace prahů a metod, tj.  $\text{conf} = 0.95$ .

Odhady parametru tvaru  $\xi$  společně s 95% konfidenčním intervalem jsou k nahlédnutí v tabulce 3.5. Ty vykazují záporné hodnoty odhadu pro prahy  $u = 1\ 500$  a  $u = 1\ 700$  při aplikaci všech třech již zmíněných metod odhadu, tj.

Tabulka 3.5: Odhad parametru tvaru  $\xi$  pro zvolenou metodu odhadu a zvolený práh  $u$ .

Metoda	Práh $u$	Odhad $\xi$	95% konfidenční interval
<b>mle</b>	<b>1 300</b>	0,2779	(0,0294 ; 0,5264)
<b>mle</b>	<b>1 500</b>	-0,2762	(-0,4775 ; -0,0748)
<b>mle</b>	<b>1 700</b>	-0,2696	(-0,5224 ; -0,0169)
<b>mple</b>	<b>1 300</b>	0,2508	(0,0192 ; 0,4823)
<b>mple</b>	<b>1 500</b>	-0,2758	(-0,4773 ; -0,0742)
<b>mple</b>	<b>1 700</b>	-0,2697	(-0,5227 ; -0,0167)
<b>pwmb</b>	<b>1 300</b>	0,2625	(0,0672 ; 0,4578)
<b>pwmb</b>	<b>1 500</b>	-0,2776	(-0,5937 ; 0,0384)
<b>pwmb</b>	<b>1 700</b>	-0,3541	(-0,7616 ; 0,0533)

metoda maximální věrohodnosti, metoda penalizované maximální věrohodnosti a metoda pravděpodobnostně vážených momentů. Jedná se tak o Beta rozdělení. Pokud se zaměříme pouze na odhady parametru tvaru metodou maximální věrohodnosti a metodou penalizované maximální věrohodnosti, zjistíme, že jsou prakticky totožné. Jak již bylo uvedeno v teoretické části, penalizační funkce v případě záporné hodnoty parametru tvaru nabývá hodnoty jedna. V takovém případě se pak penalizovaná věrohodnotní a věrohodnostní funkce rovnají. Drobné rozdíly v odhadech parametrů jsou zde pak patrně způsobeny různou tolerancí numerických algoritmů maximalizujících tyto funkce v programu *R*. Tyto hodnoty můžeme dále porovnat s odhady parametru tvaru metodou pravděpodobnostně vážených momentů. Rozdíly v odhadech jsou při hodnotě prahu  $u = 1\ 500$  nepatrné, avšak při hodnotě prahu  $u = 1\ 700$  už můžeme pozorovat rozdíly větší. Při prahové hodnotě  $u = 1\ 300$  již můžeme pozorovat rozdíly v odhadech parametru tvaru u všech třech zvolených metod. Při této prahové hodnotě je odhad parametru tvaru kladný a jedná se tak o Paretovo rozdělení.

Odhady parametrů měřítka společně s 95% konfidenčním intervalem pro zvolené prahové hodnoty  $u = 1\ 300$ ,  $u = 1\ 500$  a  $u = 1\ 700$  jsou uvedeny v tabulce 3.6. Z ní je rovněž patrné, že odhady parametru měřítka jsou při zvolení metody maximální věrohodnosti a metody penalizované maximální věrohodnosti

Tabulka 3.6: Odhad parametru měřítka  $\sigma$  pro zvolenou metodu odhadu a zvolený práh  $u$ .

Metoda	Práh $u$	Odhad $\sigma$	95% konfidenční interval
<b>mle</b>	<b>1 300</b>	207,2639	(147,5293 ; 266,9986)
<b>mle</b>	<b>1 500</b>	496,4164	(343,9300 ; 648,9027)
<b>mle</b>	<b>1 700</b>	434,2434	(269,0335 ; 599,4532)
<b>mple</b>	<b>1 300</b>	212,2628	(152,6716 ; 271,8541)
<b>mple</b>	<b>1 500</b>	496,1168	(343,7666 ; 648,4669)
<b>mple</b>	<b>1 700</b>	434,3915	(269,0368 ; 599,7462)
<b>pwmb</b>	<b>1 300</b>	205,7064	(155,8949 ; 255,5180)
<b>pwmb</b>	<b>1 500</b>	496,9204	(313,8921 ; 679,9487)
<b>pwmb</b>	<b>1 700</b>	463,9393	(254,1049 ; 673,7738)

pro prahovou hodnoty  $u = 1\ 500$  a  $u = 1\ 700$  téměř totožné. Větší hodnoty při těchto prazích pak dosahují odhady parametru měřítka při aplikování metody pravděpodobnostně vážených momentů.

Dodejme, že při malém vzorku dat ( $n \leq 15$ ) vykazuje odhad metodou pravděpodobnostně vážených momentů lepších výsledků, co se vychýlení a střední kvadratické chyby týče, než metoda maximální věrohodnosti. Jelikož žádný z vybraných prahů negeneruje takto malý počet hodnot, přikláníme se k odhadu parametrů založeném na metodě maximální věrohodnosti.

### 3.3.3. Ověření kvality modelu

Kvalitu odhadnutých parametrů pro jednotlivé prahové hodnoty a metody odhadu lze ověřit pomocí statistických testů nebo vizuálně pomocí grafů. Oba přístupy si dále detailněji představíme.

#### Statistické ověření modelu

V této části práce se zaměříme na statistické ověření kvality modelů. Pro tento účel využijeme Anderson-Darlingův, Kolmogorov-Smirnovův a Cramer-von Misesův test. V programu *R* jsou pro tyto testy implementovány funkce *ad.test()*, *ks.test()* a *cvm.test()*. Vstupy těchto funkcí jsou obdobné pro všechny

zmíněné testy. Detailněji si proto představíme pouze první z funkcí, odpovídající Anderson-Darlingovu testu, kdy jsme pro výpočet p-hodnoty použili příkaz

$$ad.test(claims[claims > u] - u, pgpd, shape, scale)$p.value,$$

kde *claims* jsou analyzovaná data, parametr *u* odpovídá hodnotě zvoleného prahu a *pgpd* je argument funkce, který odpovídá zobecněnému Paretovu rozdělení. Parametr *shape* a *scale* odpovídá hodnotě odhadu parametru tvaru a měřítka pro daný práh *u* a danou metodu odhadu parametrů uvedených v tabulkách 3.5 a 3.6. Přehled p-hodnot pro jednotlivé testy, při daných prahových hodnotách a metodě odhadu parametrů jsou uvedeny v tabulce 3.7.

Ve všech případech je p-hodnota  $> 0,05$ . Nulovou hypotézu, že dané excedenty pochází ze zobecněného Paretova rozdělení s danými parametry, nelze na hladině významnosti 0,05 zamítnout u žádného z uvedených případů.

Tabulka 3.7: P-hodnoty při použití Anderson-Darlingova testu (ad test), Kolmogorov-Smirnovova testu (ks test) a Cramer-von Misesova testu (cvm test) pro testování zobecněného Paretova rozdělení s parametry dané hodnotou prahu *u* a metodou odhadu parametrů.

Metoda	Práh <i>u</i>	ad test	ks test	cvm test
<b>mle</b>	<b>1 300</b>	0,1705	0,3101	0,2546
<b>mle</b>	<b>1 500</b>	0,9473	0,7769	0,8807
<b>mle</b>	<b>1 700</b>	0,9035	0,5659	0,7948
<b>mple</b>	<b>1 300</b>	0,1387	0,2434	0,2075
<b>mple</b>	<b>1 500</b>	0,9472	0,7747	0,8802
<b>mple</b>	<b>1 700</b>	0,9036	0,5672	0,7951
<b>pwmb</b>	<b>1 300</b>	0,1584	0,3550	0,2544
<b>pwmb</b>	<b>1 500</b>	0,9474	0,7786	0,8814
<b>pwmb</b>	<b>1 700</b>	0,8882	0,7339	0,8380

## Grafické ověření modelu

Již jsme ověřili správnost modelů dle statistických testů. V této části ověříme modely vizuálně. Pro jednotlivé kombinace prahových hodnot a metod odhadů

parametrů byly k ověření vybrány tyto grafické metody: probability plot, Q-Q plot, return level plot a density plot. Jejich výstupy jsou vzhledem k rozsáhlému počtu zobrazeny na konci této kapitoly.

Probability plot zobrazuje empirické hodnoty kumulativní distribuční funkce extrémních hodnot přesahující daný práh  $u$  vůči hodnotám kumulativní distribuční funkce teoretického zobecněného Paretova rozdělení. V ideálním případě jsou hodnoty schodné a leží na ose prvního kvadrantu. Tento typ grafu pro hodnoty přes práh  $u = 1\,300$  při odhadu parametrů zobecněného Paretova rozdělení metodou maximální věrohodnosti, penalizované maximální věrohodnosti a pravděpodobnostně vážených momentů je zobrazen na obrázku 3.9. Pro hodnoty přes práh  $u = 1\,500$  ( $u = 1\,700$ ) jsou grafy zobrazeny na obrázku 3.10 (3.11). Jelikož jsou odhady parametrů tvaru a měřítka pro tyto prahové hodnoty při použití odhadu metodou maximální věrohodnosti a penalizované maximální věrohodnosti prakticky totožné, budeme dále uvádět grafické výstupy pouze pro metodu maximální věrohodnosti.

Q-Q plot zobrazuje uspořádané empirické hodnoty zkoumaného datového souboru *claims* přesahující danou hodnotu  $u$  vůči kvantilům teoretického zobecněného Paretova rozdělení. Při vhodně zvoleném rozdělení body v grafu rovněž kopírují osu prvního kvadrantu jako u předešlého typu grafu. Rozdíly mezi modely při aplikování jednotlivých metod odhadů parametrů při hodnotě prahu  $u = 1\,300$ ,  $u = 1\,500$  a  $u = 1\,700$  jsou patrné pomocí Q-Q plotu na obrázcích 3.12, 3.13 a 3.14. Na základě těchto grafů lze usuzovat, že až na modely s prahovou hodnotou  $u = 1\,300$ , jsou modely zvolené vhodně, neboť body v Q-Q plotu kopírují červenou teoretickou křivku.

Return level plot pro napozorované datové hodnoty *claims* přesahující prahy  $u = 1\,300$ ,  $u = 1\,500$  a  $u = 1\,700$  při použití jednotlivých metod odhadů parametrů jsou zobrazeny na obrázcích 3.15, 3.16 a 3.17. V tomto případě rovněž při vhodně zvoleném modelu body v grafu kopírují červenou křivku. Tomu se nejvíce blíží modely při prahové hodnotě  $u = 1\,500$  a  $u = 1\,700$ .

Density plot zobrazuje normovaný histogram datových hodnot přesahujících

daný práh  $u$ . Histogram je vyzobrazen společně s černou křivkou, která jeho tvar kopíruje a odpovídá tak empirické hustotě. Hustota teoretického zobecněného Paretova rozdělení je zobrazena červeně. Tento typ grafu pro zvolené prahové hodnoty  $u = 1\,300$ ,  $u = 1\,500$  a  $u = 1\,700$  při použití jednotlivých metod odhadů parametrů jsou zobrazeny na obrázcích 3.18, 3.19 a 3.20. I v tomto případě se jeví jako nejlepší volba prahů  $u = 1\,500$  a  $u = 1\,700$ , neboť se empirické a teoretické křivky kopírují.

S ohledem na grafickou analýzu jsme se rozhodli pro volbu prahu  $u = 1\,500$ . Při odhadu parametrů na základě hodnot přesahující tento práh se model jeví jako dostačující a stejně dobrý jako model s prahovou hodnotou  $u = 1\,700$ . Zvolený práh přesahuje více hodnot (66), tudíž k odhadu parametrů bylo využito více informací z dat a měl by tak být přesnější. Zároveň jsme model nezakládali na všech dostupných datových hodnotách *claims*, jako na začátku této kapitoly při modelování dat gamma rozdělením, kde toto rozdělení dle Q-Q plotu nedostatečně modelovalo koncové hodnoty. Více dat v tomto případě zákonitě nezaručuje lepší model, neboť extrémní (vysoké) hodnoty se řídí jiným rozdělením než celek. Jako výhodnější se proto jeví, pokusit se najít práh, přes který lze hodnoty považovat za extrémní. Toto tvrzení dále ještě ověříme v následující části.

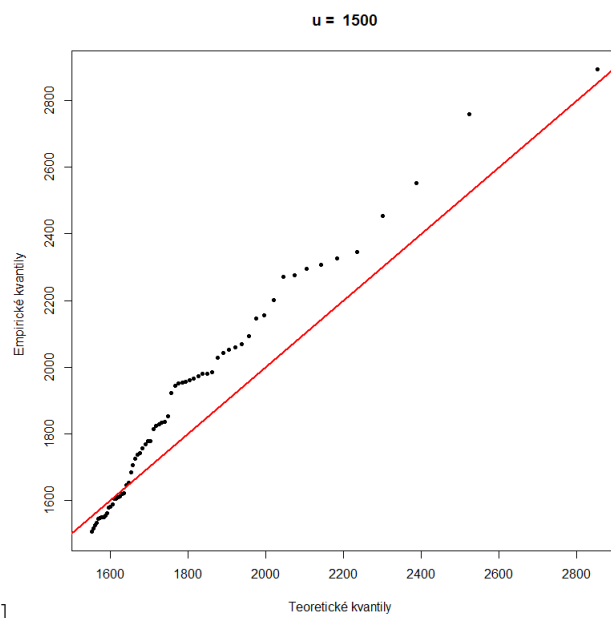
Při volbě vhodného prahu jsme přihlíželi k případnému využití modelu. Cílem této diplomové práce je modelovat data *claims* pro výpočet ceny netto zajištěného v WXL-R zajištění s prioritou  $a = 2\,000$ . V tomto případě se dle [5] doporučuje zvolit prahovou hodnotu  $u$  podstatně menší než je priorita  $a$ . Celkově se proto přikláníme k volbě prahu  $u = 1\,500$ . Vzhledem k poměrně velkému počtu hodnot přesahující tento práh ( $n > 15$ ) jsme zvolili parametry zobecněného Paretova rozdělení odhadnuté metodou maximální věrohodnosti, tj.  $\hat{\xi} = -0,2762$ ,  $\hat{\sigma} = 496,4164$ . Tento model bude následně aplikován při výpočtu ceny daného zajištění.

### 3.4. Porovnání modelů

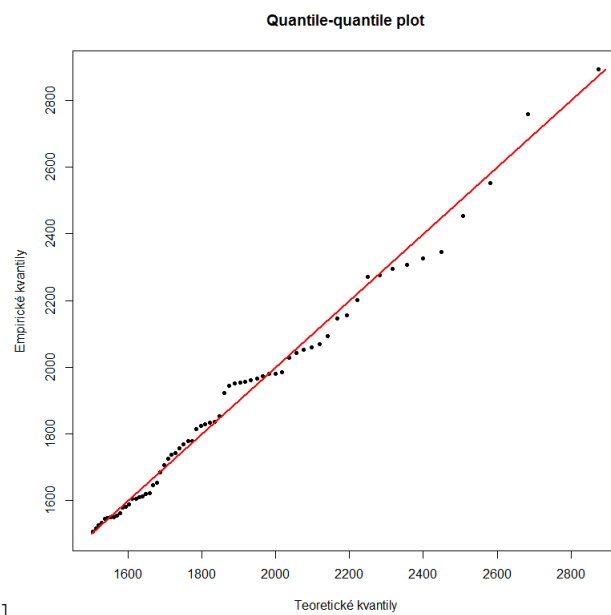
Na základě datového souboru *claims* jsme navrhli dva modely, které nyní porovnáme. U prvního jsme na základě celého datového souboru zvolili model založený na gamma rozdělení. V druhém případě jsem zvolili model založený na zobecněném Paretově rozdělení, a to pro excedenty generované prahovou hodnotu  $u = 1\,500$ . Na obrázku 3.8 jsou k porovnání Q-Q ploty těchto zmíněných modelů pro data převyšující hodnotu 1 500. Je evidentní, že pro tyto koncové hodnoty je vhodnější model druhý, založený na zobecněném Paretově rozdělení, neboť body jsou blíže červené teoretické křivce, reprezentované osou prvního kvadrantu. Toto tvrzení jsme ověřili Kolmogorov-Smirnovovým testem. Výsledné p-hodnoty jsou uvedeny v tabulce 3.8. Na hladině 0,05 zamítáme nulovou hypotézu, že dané koncové hodnoty pochází z gamma rozdělení. Pro zobecněné Paretovo rozdělení nulovou hypotézu na dané hladině zamítnout nelze. Výpočet ceny netto zajistného v této práci tak budeme provádět na základě modelu založeném na zobecněném Paretově rozdělení s parametrem tvaru  $\xi = -0,2762$  a měřítka  $\sigma = 496,4164$ . Pro porovnání rovněž uvedeme výpočet ceny netto zajistného založeném na modelu gamma rozdělení s parametrem tvaru rovný hodnotě 1,9178 a parametrem měřítka s hodnotou 226,3495.

Tabulka 3.8: P-hodnoty při použití Kolmogorov-Smirnovova testu na modely založené na gamma ( $\Gamma$ ) a zobecněném Paretově rozdělení (GPD), a data přesahující hodnotu 1 500.

Rozdělení	p-hodnota
$\Gamma(1,9178; 226,3495)$	0,0000
$GPD(496,4164; -0,2762)$	0,7769



[a]



[b]

Obrázek 3.8: Q-Q plot pro data *claims* přesahující hodnotu 1 500 pro model založený na a) gamma rozdělení b) zobecněném Paretově rozdělení.



### 3.5. Výpočet netto zajistného WXL-R zajištění

Z teoretické části víme, že netto zajistné WXL-R zajištění s prioritou  $a$  lze spočítat ze vztahu

$$P_Z = E(S_Z) = E(N)E(X_Z).$$

Očekávanou část škody  $X$  hrazenou zajistitelem lze vyjádřit ve tvaru

$$E(X_Z) = \int_a^\infty (1 - F_X(x))dx,$$

kde  $F_X(x)$  je distribuční funkce pojistného plnění (škody)  $X$ .

Jako první uvedeme výpočet ceny netto zajistného při použití modelu popisující výši škody  $X$  rozdělením  $\Gamma(1, 9178; 226, 3495)$ . Očekávanou výši jednotlivého pojistného plnění hrazeného zajistitelem pak lze vypočítat následovně:

$$E(X_Z) = \int_a^\infty (1 - F_X^G(x))dx = 0,302736,$$

kde  $F_X^G(x)$  odpovídá distribuční funkci daného gamma rozdělení.

Vzhledem k tomu, že známe pouze počet pojistných událostí za rok 2011 a nemáme informaci o celkovém počtu sjednaných pojistných smluv, tak budeme dále předpokládat, že náhodná veličina  $N$  pochází z Poissonova rozdělení s parametrem  $\lambda = 9\,134$ , tj.  $N \sim Po(\lambda = 9\,134)$  s očekávanou hodnotou  $E(N) = 9\,134$ . Cena netto zajistného při sjednání WXL-R zajištění s prioritou  $a = 2\,000$  pro krytí škody vzniklé na vozidle byla na základě modelu  $\Gamma(1, 9178; 226, 3495)$  stanovena následovně:

$$P_Z = E(N)E(X_Z) = 9\,134 \cdot 0,302736 = 2\,765,191.$$

Jako další uvedeme analogicky provedený výpočet ceny netto zajistného v WXL-R zajištění se zvolenou hodnotou priority  $a = 2\,000$ , kde jsme na základě dat *claims* zvolili zobecněné Paretovo rozdělení s parametry  $\xi = -0,2762$  a  $\sigma = 496,4164$  jako vhodné rozdělení pro modelování výše přesahu škody  $X$  přes hodnotu  $u = 1\,500$ , tj.  $X_u$ . Distribuční funkci  $X_u$  lze zapsat ve tvaru

$$F_X^{(1\,500)}(x) = 1 - \left(1 - 0,2762 \frac{x}{496,4164}\right)^{1/0,2762}.$$

Z teoretické části dále víme, že pokud  $Y$  je náhodná veličina z rozdělení GPD s parametry  $u_0, \sigma_{u_0}, \xi$ , potom náhodná veličina  $Y|Y > u_1$ , kde  $u_1 > u_0$ , má rovněž GPD rozdělení, a to s parametry ve tvaru  $u_1, \sigma_{u_1} = \sigma_{u_0} + \xi(u_1 - u_0)$  a  $\xi_1 = \xi$ . Díky tomuto vztahu lze zapsat distribuční funkci výše přesahu škody  $X$  přes prioritu  $a$  ( $X_a$ ) ve tvaru

$$F_X^{(2\ 000)}(x) = 1 - \left(1 - 0,2762 \frac{x}{358,3164}\right)^{1/0,2762}.$$

Označme nyní  $E(X_Z^a)$  očekávanou hodnotou jednotlivého pojistného plnění přesahující hodnotu  $a = 2\ 000$ . Úpravou a dosazením do výrazu pro výpočet  $E(X_Z^a)$  dostaneme následující:

$$\begin{aligned} E(X_Z^a) &= \int_a^\infty \left(1 - F_X^{(2\ 000)}(x - a)\right) dx = \\ &= \int_{2\ 000}^\infty \left(1 - \left(1 - \left(1 - 0,2762 \frac{x - 2\ 000}{358,3164}\right)^{1/0,2762}\right)\right) dx = \\ &= 280,7682. \end{aligned}$$

Očekávaný počet škod na kterých se podílí zajistitel, tj. počet škod přesahující prioritu  $a$ ,  $E(N_Z)$  lze dle teoretické části odvodit z očekávaného celkového počtu pojistných událostí  $E(N)$  ze vztahu:

$$E(N_Z) = p_a E(N),$$

kde  $p_a = P(X > a) = 1 - F_X(a)$ .

$E(N_Z)$  lze dále pomocí distribuční funkce přesahu škod  $X$  přes práh  $u$  ( $X_u$ )  $F_X^u(x)$  přepsat na tvar:

$$E(N_Z) = p_a^u E(N_u),$$

kde  $p_a^u = P(X_u > a - u) = 1 - F_X^u(a - u)$  a  $E(N_u)$  je očekávaný počet škod přesahující práh  $u$ . K tomuto výpočtu využijeme distribuční funkci výše přesahu škody  $X$  přes práh  $u = 1\ 500$ , tj.  $F_X^{(1\ 500)}(x)$ . Úpravou a dosazením pak získáme následující:

$$p_a^u = P(X_u > 2\ 000 - 1\ 500) = 1 - F_X^{(1\ 500)}(2\ 000 - 1\ 500) = 0,3071868.$$

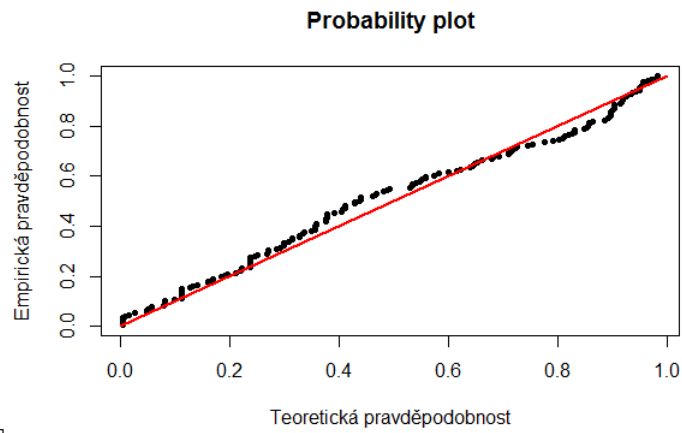
Dále v této práci budeme předpokládat, že náhodná veličina  $N_u$  pochází z Poissonova rozdělení s parametrem  $\lambda = 66$ , tj.  $N_u \sim Po(\lambda = 66)$  s očekávanou hodnotou  $E(N_u) = 66$ . Očekávaný počet škod, na kterých se podílí zajistitel,  $E(N_Z)$ , tak lze vypočítat následovně:

$$E(N_Z) = p_a^u E(N_u) = 0,3071868 \cdot 66 = 20,27433.$$

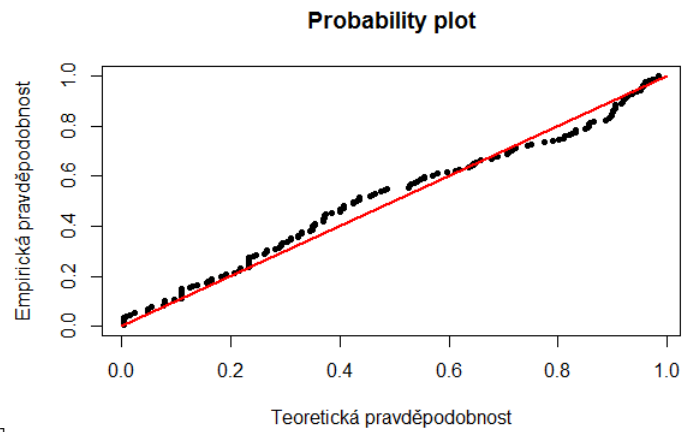
Cena netto pojistného při sjednání WXL-R zajištění s prioritou  $a = 2\,000$  pro krytí škody vzniklé na vozidle byla na základě modelu, popisující pravděpodobnostní rozdělení výše přesahu škody přes hodnotou  $1\,500$  zobecněným Paretoovým rozdělením s parametrem tvaru  $\xi = -0,2762$  a parametrem měřítka  $\sigma = 496,4164$ , stanovena následovně:

$$P_Z = E(N_Z)E(X_Z^a) = 20,27433 \cdot 280,7682 = 5\,692,386.$$

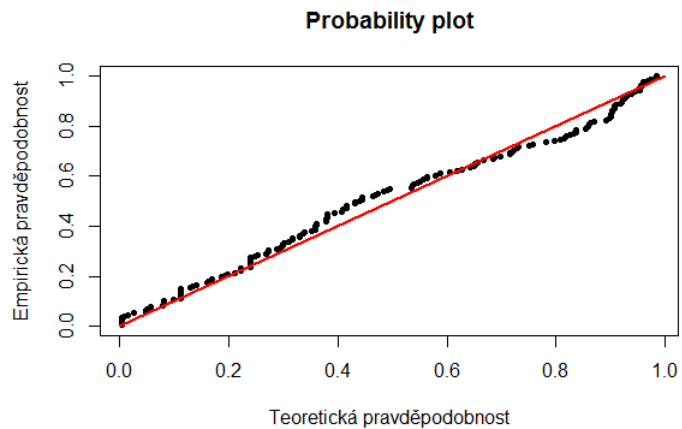
Tato cena odpovídá očekávané celkové výši škod hrazené zajistitelem na základě modelu zobecněného Paretova rozdělení. Pro porovnání, dle dat by součet částí škod přesahujících hodnotu  $a = 2\,000$  (19) byl roven číslu  $5\,324,032$ .



[a]

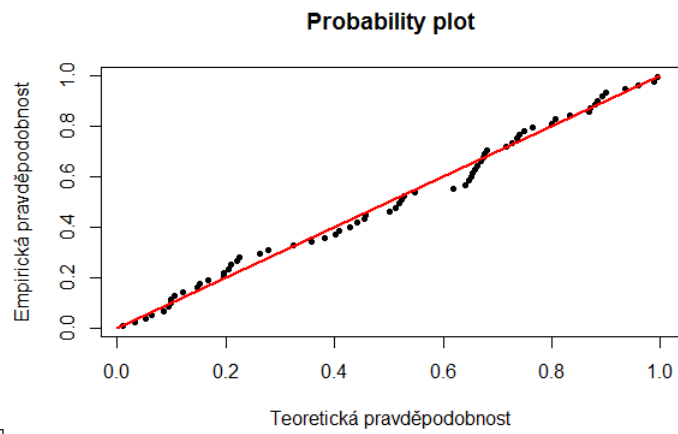


[b]

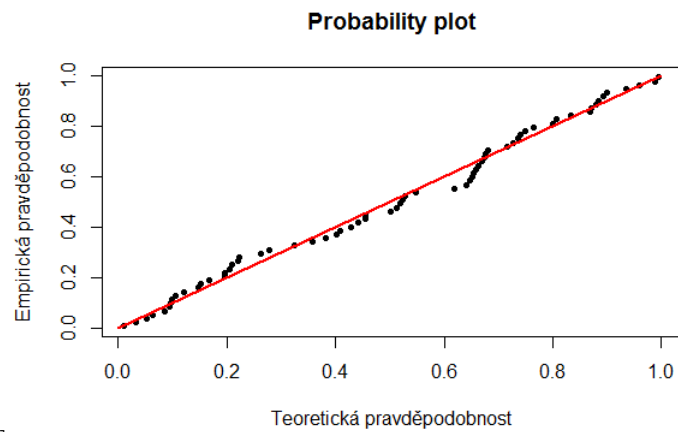


[c]

Obrázek 3.9: Probability plot pro data při prahové hodnotě  $u = 1\,300$  a odhadu parametrů metodou a) maximální věrohodnosti b) penalizované maximální věrohodnosti c) pravděpodobnostně vážených momentů.

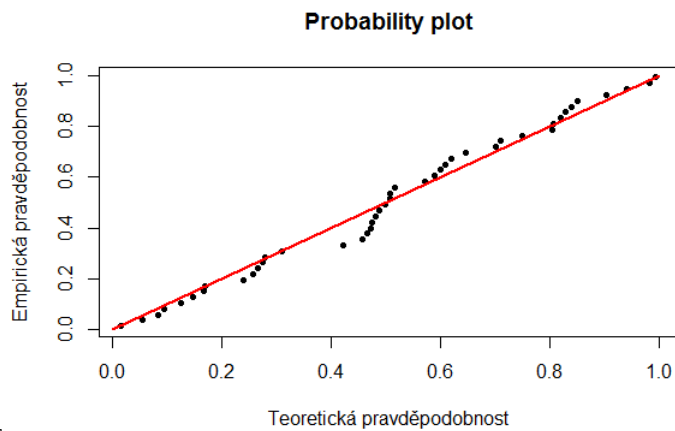


[a]

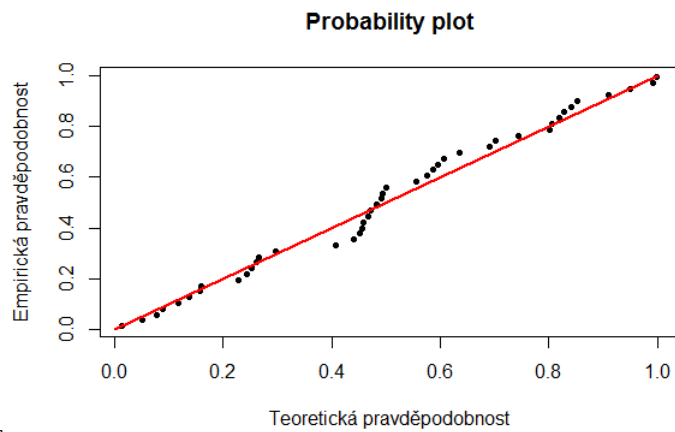


[b]

Obrázek 3.10: Probability plot pro data při prahové hodnotě  $u = 1\,500$  a odhadu parametrů metodou a) maximální věrohodnosti b) pravděpodobnostně vážených momentů.

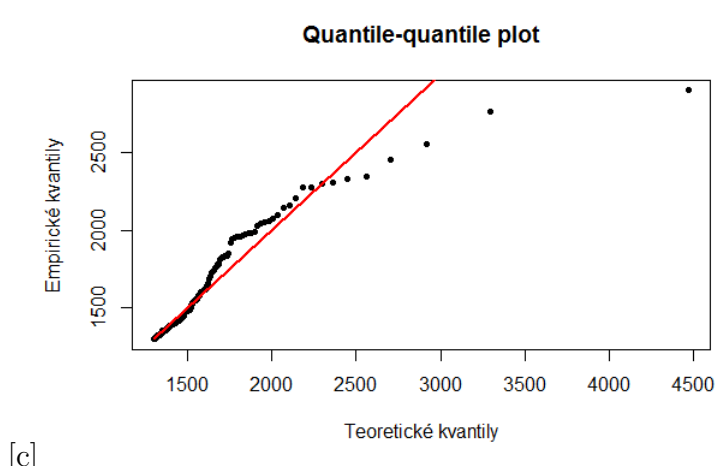
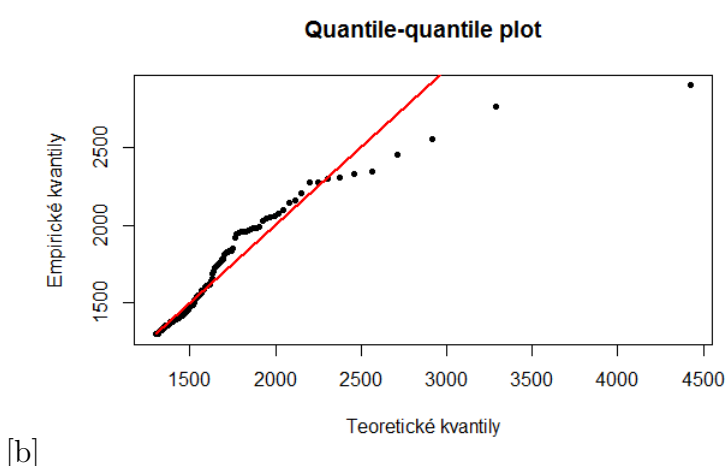
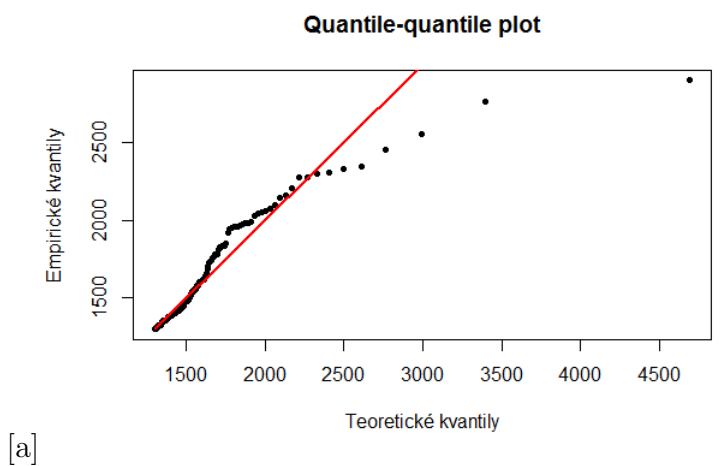


[a]

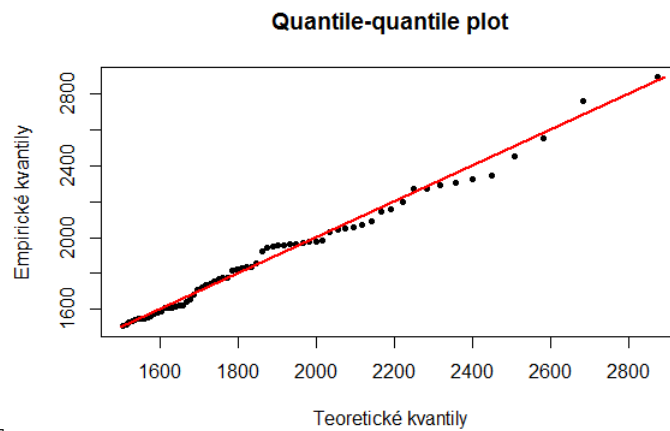


[b]

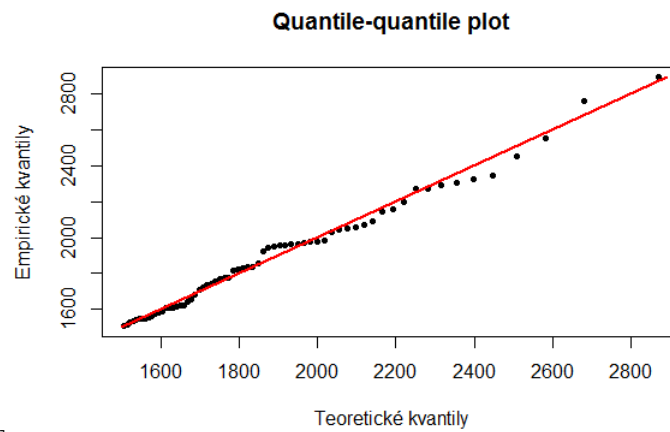
Obrázek 3.11: Probability plot pro data při prahové hodnotě  $u = 1\,700$  a odhadu parametrů metodou a) maximální věrohodnosti b) pravděpodobnostně vážených momentů.



Obrázek 3.12: Quantile plot pro data při prahové hodnotě  $u = 1\,300$  a odhadu parametrů metodou a) maximální věrohodnosti b) penalizované maximální věrohodnosti c) pravděpodobnostně vážených momentů.



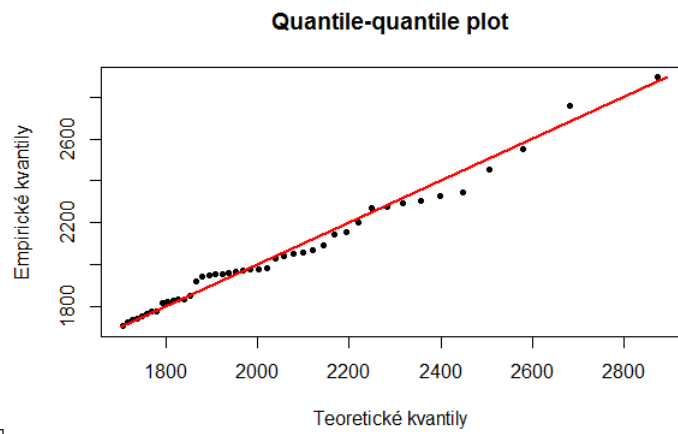
[a]



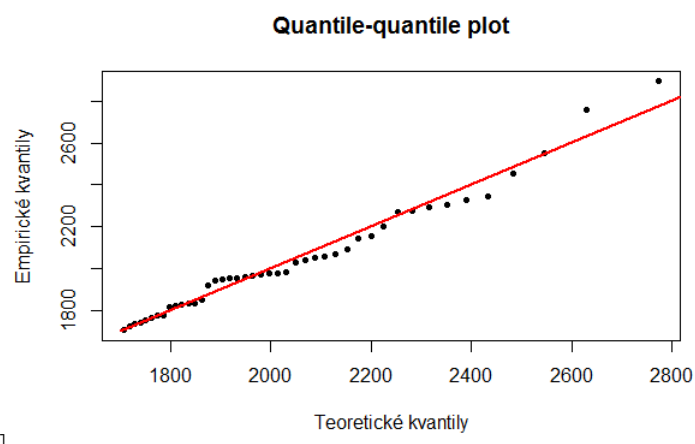
[b]

Obrázek 3.13: Quantile plot pro data při prahové hodnotě  $u = 1\,500$  a odhadu parametrů metodou a) maximální věrohodnosti b) pravděpodobnostně vážených momentů.



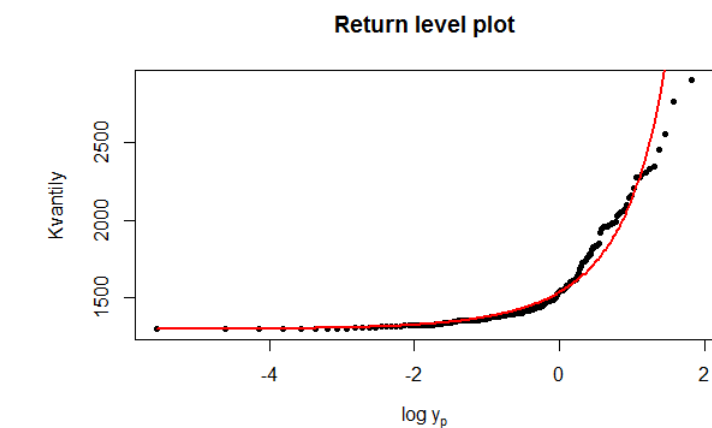


[a]

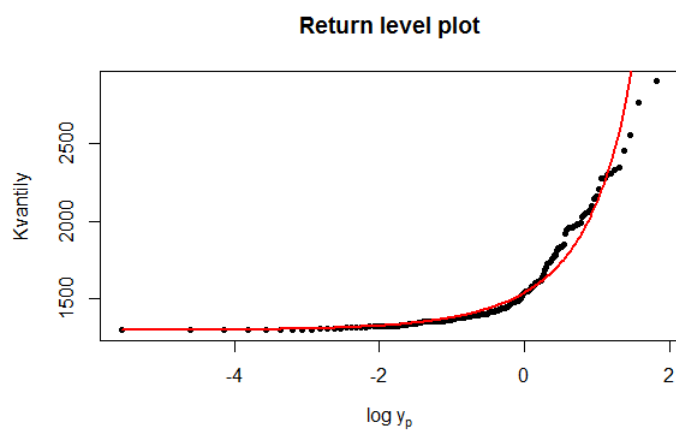


[b]

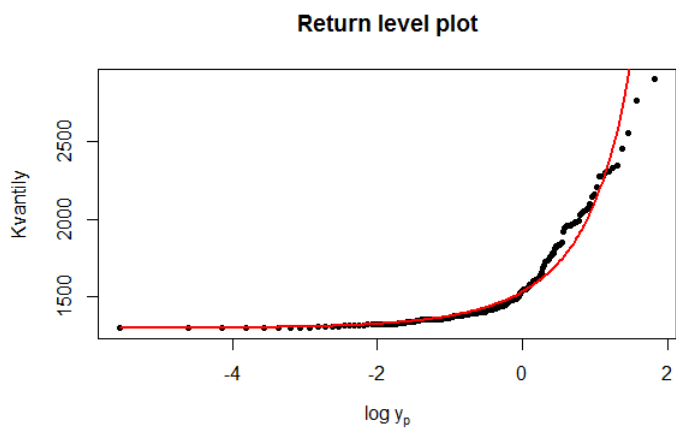
Obrázek 3.14: Quantile plot pro data při prahové hodnotě  $u = 1\,700$  a odhadu parametrů metodou a) maximální věrohodnosti b) pravděpodobnostně vážených momentů.



[a]

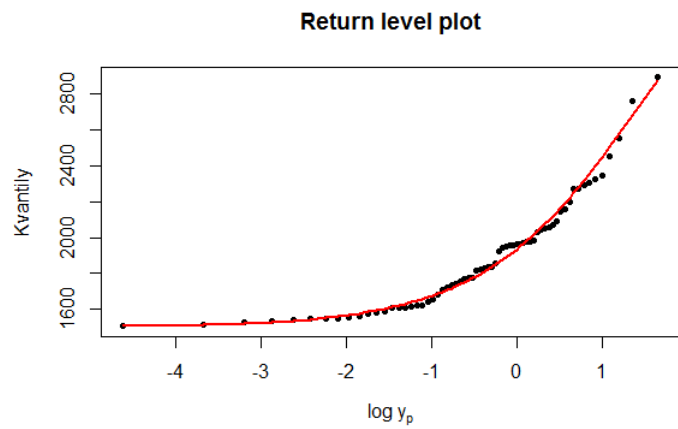


[b]

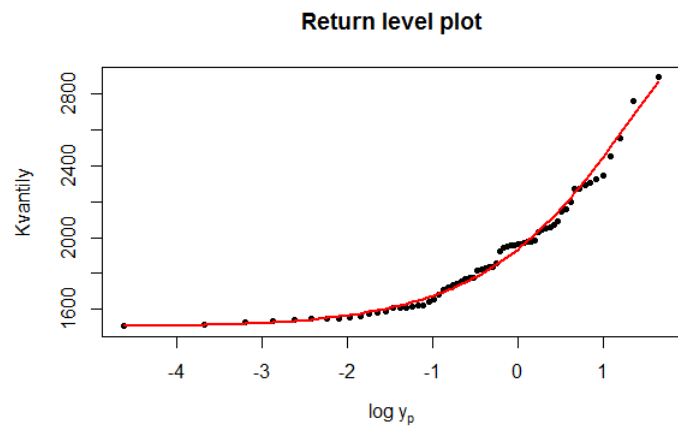


[c]

Obrázek 3.15: Return level plot pro data při prahové hodnotě  $u = 1\,300$  a odhadu parametrů metodou a) maximální věrohodnosti b) penalizované maximální věrohodnosti c) pravděpodobnostně vážených momentů.

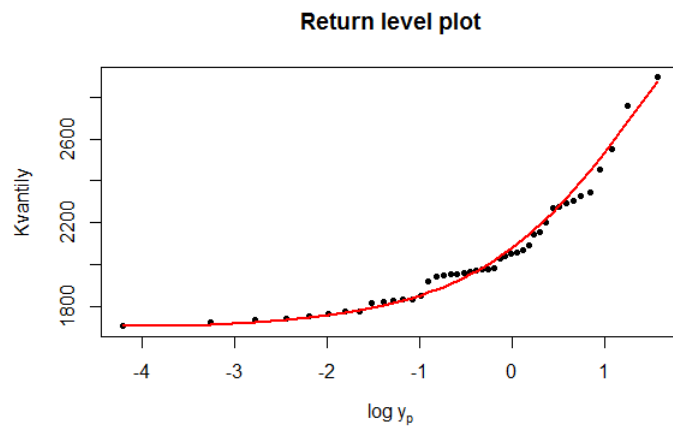


[a]

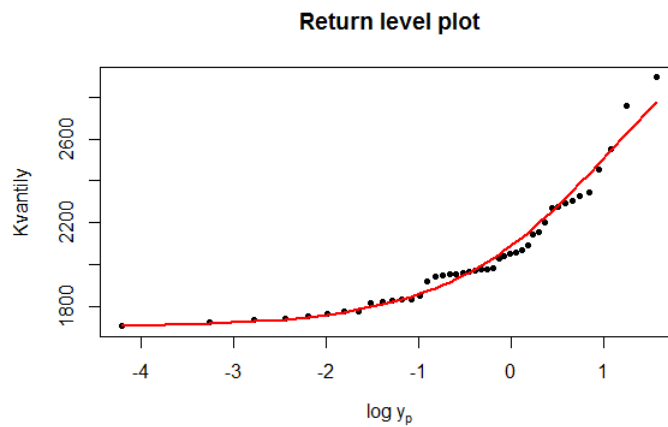


[b]

Obrázek 3.16: Return level plot pro data při prahové hodnotě  $u = 1\,500$  a odhadu parametrů metodou a) maximální věrohodnosti b) pravděpodobnostně vážených momentů.

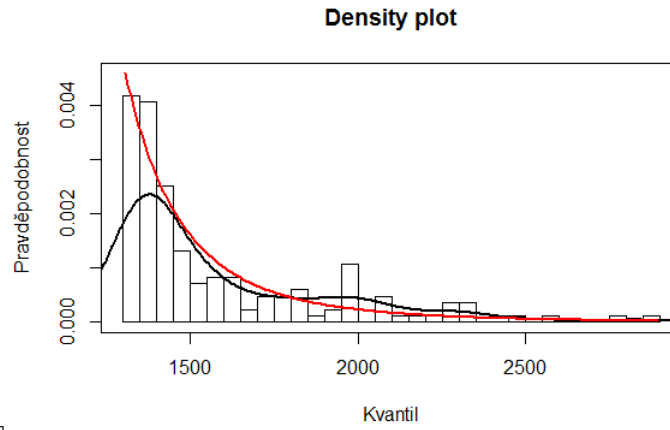


[a]

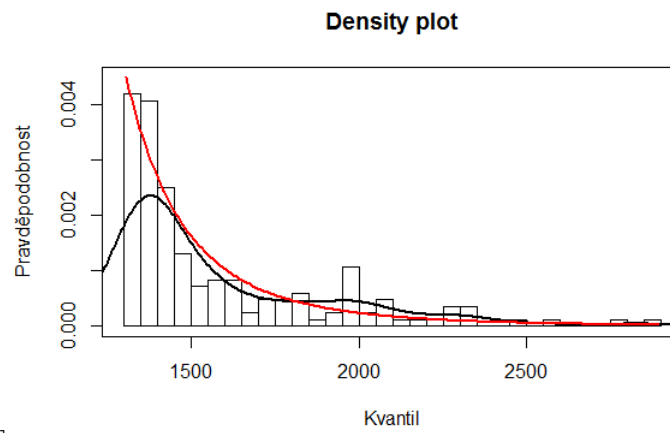


[b]

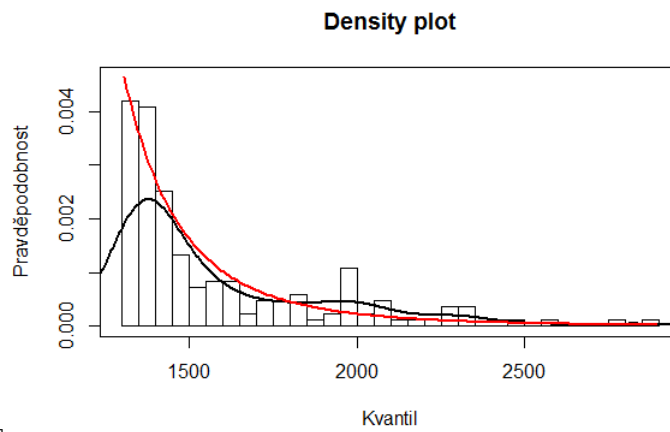
Obrázek 3.17: Return level plot pro data při prahové hodnotě  $u = 1\,700$  a odhadu parametrů metodou a) maximální věrohodnosti b) pravděpodobnostně vážených momentů.



[a]

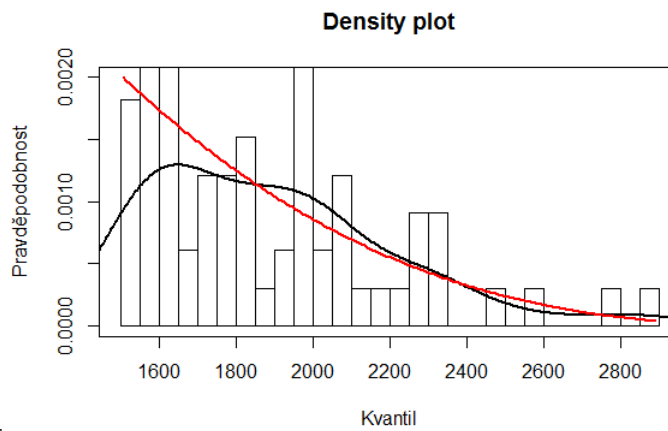


[b]

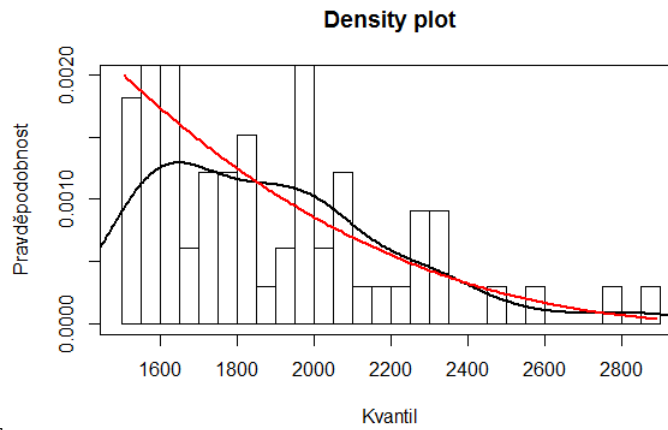


[c]

Obrázek 3.18: Density plot pro data při prahové hodnotě  $u = 1\,300$  a odhadu parametrů metodou a) maximální věrohodnosti b) penalizované maximální věrohodnosti c) pravděpodobnostně vážených momentů.

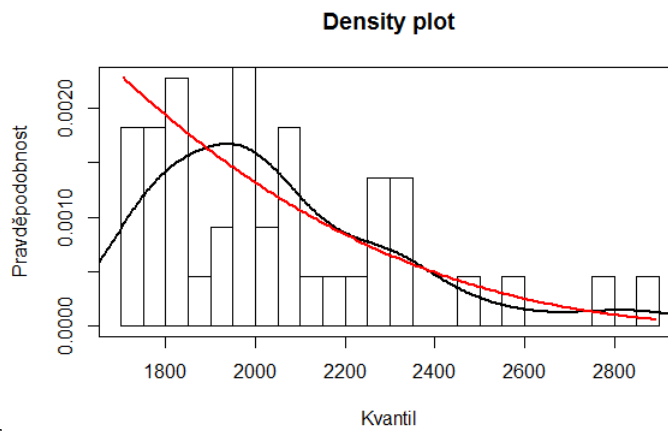


[a]

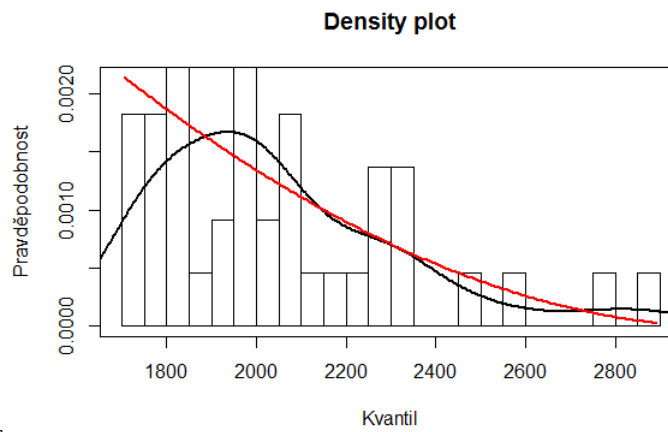


[b]

Obrázek 3.19: Density plot pro data při prahové hodnotě  $u = 1\,500$  a odhadu parametrů metodou a) maximální věrohodnosti b) pravděpodobnostně vážených momentů.



[a]



[b]

Obrázek 3.20: Density plot pro data při prahové hodnotě  $u = 1\,700$  a odhadu parametrů metodou a) maximální věrohodnosti b) pravděpodobnostně vážených momentů.

# Závěr

Cílem této práce bylo seznámit se a popsat možné přístupy modelování extrémních škod v neživotním pojištění. Celá práce se skládá ze tří kapitol. První kapitola se zabývá základními pojmy z pojišťovnictví, které jsou potřebné pro pochopení systému zajištění. Tato kapitola dále pojednává o typech možného zajištění, přičemž se zaměřuje na  $WXL - R$  zajištění, pro které stanovuje i výpočet ceny netto zajistného. Tato cena za zajištění závisí na počtu a výši škod. Druhá kapitola seznamuje čtenáře s teorií extrémních hodnot a s modely, které jsou vhodné pro jejich popis. Hodnoty lze za extrémní považovat z pohledu blokových maxim nebo excedentů přes vysoký práh. Tato práce se více zaměřuje na modely pro excedenty přes vysoký práh, neboť tato metoda je aplikovaná v praktické části práce.

Třetí kapitola na datech popisujících výši škod vzniklých na vozidlech demonstruje, poznatky z předešlých teoretických částí. Po seznámení čtenáře s daty je představen model založený na gamma rozdělení, popisující výši škod. Ten je dále porovnán s modelem založeným na zobecněném Paretově rozdělení, který popisuje pouze škody přes zvolený práh. Výběr optimálního prahu je založen na grafických metodách: Mean residual plot, Threshold choice plot a L-moments plot. K ověření vhodnosti modelů jsou použity statistické testy a grafické metody, jako je probability plot, Q-Q plot, density plot a return level plot.

V závěru práce jsou následně vybrané modely použity při výpočtu ceny netto zajistného v  $WXL - R$  zajištění se sjednanou prioritou  $a$ .



# Literatura

- [1] Anděl, J.: *Základy matematické statistiky*. Vyd. 3., Praha: Matfyzpress, 2011. ISBN 978-80-7378-162-0.
- [2] Beirlant, J., Goegebeur, Y., Teugels, J.: *Statistics of Extremes: Theory and Applications*. 2004. John Wiley Sons, Anglie.
- [3] Cipra, T.: *Pojistná matematika: teorie a praxe*. 2., aktualiz. vyd. Ekopress, Praha, c2006.
- [4] Cipra, T.: *Riziko ve financích a pojišťovnictví: Basel III a Solvency II*. Vydání I. Ekopress, Praha, 2015.
- [5] Cipra, T.: *Zajištění v pojišťovnictví a jeho matematické aspekty*. JČMF, 2004.
- [6] Coles, S. G., Dixon, M. J.: *Likelihood-based inference for extreme value models*. *Extremes*, 1999, 2.1: 5-23.
- [7] Daňhel, J.: *Kapitoly z pojistné teorie*. Nakladatelství Oeconomica, Praha, 2002.
- [8] Ducháčková, E., Daňhel, J.: *Teorie pojistných trhů*. Professional Publishing, Praha, 2010.
- [9] Ducháčková, E.: *Principy pojištění a pojišťovnictví*. Ekopress, Praha, 2009.
- [10] Fukutome, S., Schindler, A., Capobianco, A.: *MeteoSwiss extreme value analyses: User manual and documentation*. Federal Office of Meteorology and Climatology, MeteoSwiss, 2017.
- [11] Fusek, M.: *Rozdělení extrémních hodnot a jejich aplikace*. Vysoké učení technické v Brně, 2013.
- [12] github.com: *Prediction-of-Claim-Amount-in-Vehicle-Insurance*. [online]. [cit. 2021-03-20]. Dostupné z: <https://github.com>.
- [13] Hosking, J. R. M., Wallis, J. R.: *Parameter and quantile estimation for the generalized pareto distribution*. *Technometrics*, 1987, 29:339–349.

- [14] Pešta, M., Petrová B., Procházka J., Smolárová T., Zimmermann P.: *Exercises for non-life insurance*. University of Economics, Prague and Charles University in Prague, 2016.
- [15] Skřivánková, V.: *Štatistická analýza extrémnych hodnôt a metódy ich registrácie v neživotnom poistení*. 3. Mezinárodní konference Řízení a modelování finančních rizik (pp. 270-277), 2006.
- [16] Zuzáková, B.: *Modelování velkých škod*. Univerzita Karlova v Praze: Matematicko-fyzikální fakulta, 2013.