Univerzita Karlova v Praze

Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE

Bc. Tereza Havlíková

Výpočty variability vývojových trojúhelníků v neživotním pojištění

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: RNDr. Martin Branda, Ph.D.

Studijní program: Matematika

Studijní obor: Finanční a pojistná matematika

Praha 2013

Ráda bych na tomto místě poděkovala RNDr. Martinovi Brandovi, Ph.D., vedoucímu mé diplomové práce, za odborné vedení, věnovaný čas a cenné rady. Dále děkuji svým rodičům za jejich podporu během mého studia.

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V Praze dne 1. 8. 2013 Tereza Havlíková

Název práce: Výpočty variability vývojových trojúhelníků v neživotním pojištění

Autor: Bc. Tereza Havlíková

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: RNDr. Martin Branda, Ph.D., KPMS MFF UK

Abstrakt: Cílem této práce je popsat metody výpočtu variability odhadu rezervy na pojistná plnění v neživotním pojištění. Práce se zabývá popisem tří postupů pro výpočet variability odhadu rezervy – Mackovým stochastickým modelem Chain-Ladder, zobecněnými lineárními modely a metodou bootstrap. Práce obsahuje jak teoretickou, tak praktickou část, která je věnována aplikaci zmíněných modelů na reálná a nasimulovaná data.

Klíčová slova: vývojový trojúhelník, rezerva na pojistná plnění, střední kvadratická chyba, Mackův model Chain-Ladder, zobecněný lineární model, bootstrap

Title: Variability estimation of development triangles in non-life insurance

Author: Bc. Tereza Havlíková

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Martin Branda, Ph.D., KPMS MFF UK

Abstract: The aim of this thesis is to describe calculation methods for variability estimation of claims reserve in non-life insurance. The thesis focuses on three main categories of models: Mack's stochastic Chain-Ladder, generalized linear models and bootstrap. Both the theoretical and also the empirical parts are included. Empirical part is devoted to application of all the models described above on both real and simulated data.

Keywords: run-off triangle, claims reserve, mean squared error, Mack's Chain-Ladder, generalized linear model, bootstrap

Obsah

Úvod **2**

1 Metoda Chain-Ladder **3**
1.1 Mackův model pro metodu Chain-Ladder . . . 3
1.2 Výpočet střední kvadratické chyby a standardní chyby . . . 9
1.3 Rekurzivní výpočet standardní chyby a zahrnutí tail faktoru . . . 13

2 Zobecněné lineární modely **19**
2.1 Teorie zobecněných lineárních modelů . . . 19
2.2 Zobecněný lineární model v pojišťovnictví . . . 24

3 Metoda bootstrap **29**
3.1 Základní principy metody bootstrap . . . 29
3.2 Bootstrap v praxi . . . 31

4 Praktická část **37**
4.1 Reálná data . . . 37
4.2 Simulace . . . 46
4.3 Výběr metody . . . 48

Závěr **49**

Seznam použité literatury **50**

Úvod

Každá pojišťovna v České republice je ze Zákona č. 277/2009 Sb., o pojišťovnictví povinna vytvářet technické rezervy k plnění závazků z provozované pojišťovací činnosti, které jsou pravděpodobné nebo jisté, ale nejistá je jejich výše nebo okamžik, ke kterému tyto závazky vzniknou.

V neživotním pojištění má velký význam rezerva na pojistná plnění, někdy také označovaná jako škodní rezerva. Rezerva na pojistná plnění je určena ke krytí závazků z pojistných událostí v období před rozvahovým dnem vzniklých, hlášených, ale v tomto období neuhrazených a nebo vzniklých, ale nehlášených. První typ rezervy se nazývá RBNS rezerva (reported but not settled), druhá se nazývá IBNR rezerva (incurred but not reported). Výše RBNS rezervy je stanovena likvidátorem pojistných událostí. Oproti tomu výše IBNR rezervy se stanoví pomocí matematicko-statistických metod.

Od 1. ledna 2014 začne na území Evropské Unie platit směrnice o přístupu k pojišťovací a zajišťovací činnosti a jejím výkonu, nazývaná Solvency II. Kompletní znění tohoto předpisu je k nalezení pod názvem Směrnice Evropského parlamentu a Rady 2012/23/EU. V rámci této směrnice se hovoří o solventním kapitálovém požadavku, který pojišťovna potřebuje ke krytí všech rizik, kterým je vystavena. Jedním z těchto rizik je riziko pojistného a rezerv v neživotním pojištění. V našem případě potřebujeme odhadnout riziko plynoucí z nedostatečnosti škodních rezerv ke krytí budoucích škod. Kapitálový požadavek potom odpovídá hodnotě v riziku na hladině spolehlivosti $99,5\%$ v časovém horizontu jednoho roku. K odhadu tohoto kvantilu můžeme využít standardní chybu rezervy [16]. Z tohoto důvodu je důležité počítat kromě odhadu rezervy i standardní chybu tohoto odhadu.

Cílem diplomové práce *Výpočet variability vývojových trojúhelníků v neživotním pojištění* je představit hlavní metody výpočtu IBNR rezervy na pojistná plnění a její střední kvadratické chyby a následně je aplikovat a porovnat na reálných a nasimulovaných datech.

Vlastní text je členěn do čtyř kapitol. První tři kapitoly jsou věnovány konkrétním matematickým modelům - Mackově stochastickému modelu metody Chain-Ladder společně se zobecněním této metody pomocí rekurze a zahrnutím tail faktoru do výpočtu rezervy, zobecněným lineárním modelům a metodě bootstrap. Poslední kapitola uvádí aplikaci zmíněných metod na reálná a nasimulovaná data. Dále tato kapitola obsahuje srovnání výsledků získaných pomocí jednotlivých metod.

Poznamenejme, že předpokládáme, že čtenář má základní znalosti z oboru pravděpodobnosti a matematické statistiky.

1. Metoda Chain-Ladder

V neživotním pojištění má velký význam rezerva na pojistná plnění (škodní rezerva), jejíž výše na konci účetního období odpovídá odhadu výše pojistných plnění, která ještě mají být v budoucnu vyplacena za škody vzniklé v daném účetním období nebo v obdobích předchozích. Nejčastější postupy k odhadování rezervy na pojistná plnění jsou metoda Chain-Ladder, separační metoda, Cape Cod metoda, Bornhuetter-Fergusonova metoda a De Vylderova metoda nejmenších čtverců. V této kapitole se zaměříme na metodu Chain-Ladder. Jedná se o početně jednoduchou metodu, u které nepotřebujeme znát rozdělení vzniklých škod. Na druhou stranu jsou získané odhady rezerv pro jednotlivé roky citlivé na změny v pozorovaných datech. Proto je užitečné znát standardní odchylku odhadovaných rezerv jako míru nejistoty obsažené v datech.

1.1 Mackův model pro metodu Chain-Ladder

V této kapitole popíšeme stochastický model pro metodu Chain-Ladder, který v roce 1993 zavedl Thomas Mack. Dále zavedeme pojmy a značení, s nimiž budeme v této kapitole pracovat. Původní text může čtenář nalézt v [10], dále čerpáme z článků [1] a [12].

Nechť $C_{i,j}$ označuje celkový úhrn pojistného plnění za škody, které vznikly v roce i a byly uhrazeny (nahlášeny) do konce vývojového roku j, kde $i = 0, 1, \ldots, I$, $j = 0, 1, \ldots, J$, $I = J$. Nechť $C_{i,\infty}$ je celková výše plnění škodního roku i. Rok J budeme považovat za poslední rok vývoje, tedy $C_{i,\infty} = C_{i,J}$. Po celou dobu budeme předpokládat, že $C_{i,j}$ je nezáporná náhodná veličina a známe její pozorování, pokud $i + j \leq J$.

Data potřebná k výpočtu rezervy na pojistná plnění na základě minulého škodního vývoje se často prezentují ve formě vývojových trojúhelníků. Jedná se o schéma, v němž řádek reprezentuje rok vzniku škody a sloupec označuje zpoždění v úhradě (hlášení) dané škody.

Uvažujme kumulativní trojúhelník $\{C_{i,j}\}$, který obsahuje všechna dosud známá pozorování.

	0	1	$\ldots$	j	$\ldots$	$J-1$	J
0	$C_{0,0}$	$C_{0,1}$		$C_{0,j}$		$C_{0,J-1}$	$C_{0,J}$
1	$C_{1,0}$	$C_{1,1}$		$C_{1,j}$		$C_{1,J-1}$	
$\vdots$							
$J-1$	$C_{J-1,0}$	$C_{J-1,1}$					
J	$C_{J,0}$						

Cílem je odhadnout konečnou výši škod $C_{i,J}$ a rezervu nesplacených pojistných plnění

$$R_i = C_{i,J} - C_{i,J-i} \tag{1.1}$$

pro všechny roky $i = 0, \dots, J$.

Předpoklady Mackovy metody

(MCL1) Existují vývojové faktory f_j, pro které platí

$$E\left[C_{i,j+1} | C_{i,1}, \dots, C_{i,j}\right] = f_j C_{i,j},\ i = 0, \dots, I-1,\ j = 0, \dots, J-1.$$

(MCL2) Jednotlivé roky vzniků škod jsou nezávislé, tj.

$$(C_{i,0}, \dots, C_{i,J}) \text{ a } (C_{k,0}, \dots, C_{k,J})$$

jsou vzájemně nezávislé náhodné vektory pro $i \neq k$.

(MCL3) Pro podmíněný rozptyl platí

$$var\left[C_{i,j+1} | C_{i,1}, \dots, C_{i,j}\right] = \sigma_j^2 C_{i,j},$$

pro $i = 0, \dots, I-1$, $j = 0, \dots, J-1$, kde σ_j^2 je neznámý parametr.

Odhady parametrů

Součástí metody Chain-Ladder je nalezení odhadů vývojových faktorů f_j pomocí vztahu

$$\hat{f}_j = \frac{\sum\limits_{i=0}^{J-1-j} C_{i,j+1}}{\sum\limits_{i=0}^{J-1-j} C_{i,j}} \tag{1.2}$$

pro $j = 0, \dots, J-1$.

Trojúhelník $\{C_{i,j}\}$ doplníme na čtverec násobením odhadnutými faktory $\hat{f}_j$ počínaje od diagonály. Tím získáme odhad konečné výše škodního plnění $C_{i,J}$

$$\hat{C}_{i,J} = C_{i,J-i} \hat{f}_{J-i} \cdot \ldots \cdot \hat{f}_{J-1}, \tag{1.3}$$

ekvivalentně pro škodní rezervu za rok i můžeme psát

$$\hat{R}_i = C_{i,J-i}(\hat{f}_{J-i} \cdot \ldots \cdot \hat{f}_{J-1} - 1). \tag{1.4}$$

Pro další výpočty budeme potřebovat vzorec pro odhad parametru σ_j^2. Tento odhad získáme ze vztahu

$$\hat{\sigma}_j^2 = \frac{1}{J-j-1} \sum_{i=0}^{J-j-1} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^2 \tag{1.5}$$

pro $j = 0, \dots, J-2$.

V následující větě pomocí předpokladů (MCL1), (MCL2) odvodíme předpověď $C_{i,J}$ založenou na všech dosud známých pozorováních.

Věta 1. *Nechť* $\Delta = \{C_{i,j}; i + j \leq J\}$ *je množina všech doposud napozorovaných dat. Pak za předpokladů (MCL1) a (MCL2) platí rovnost*

$$E(C_{i,J}|\Delta) = C_{i,J-i} f_{J-i} \cdot \ldots \cdot f_{J-1}. \tag{1.6}$$

Než přejdeme k důkazu právě vyslovené věty, zmíníme známé vztahy z teorie pravděpodobnosti, které budeme později využívat.

Lemma 1. *Nechť* $X, Y \in L_1(\Omega, \mathcal{A}, P)$, $\mathcal{C} \subset \mathcal{F} \subset \mathcal{A}$, *potom platí*

$$E\left(X|\mathcal{C}\right) = E\left(E\left(X|\mathcal{F}\right)|\mathcal{C}\right).$$

Lemma 2. *Nechť* $X, Y \in L_1(\Omega, \mathcal{A}, P)$, $\mathcal{F} \subset \mathcal{A}$. *Dále nechť* X *je* $F-$*měřitelná a* $E|XY| < \infty$, *potom platí*

$$E\left(XY|\mathcal{F}\right) = XE\left(Y|\mathcal{F}\right).$$

Nyní dokážeme větu 1.

Důkaz. (Mack [10])

Pro jednoduchost zápisu zavedeme označení

$$E_i(X) = E(X|C_{i,1}, \ldots C_{i,J-i}).$$

Na základě předpokladu (MCL2) a opakovanou aplikací předpokladu (MCL1) a pomocí lemmatu 1 a lemmatu 2 dostáváme

$$\begin{aligned} E(C_{i,J}|\Delta) &= E_i(C_{i,J}) \\ &= E_i(E(C_{i,J}|C_{i,1}, \ldots C_{i,J-i})) \\ &= E_i(C_{i,J-1} f_{J-1}) \\ &= E_i(C_{i,J-1}) f_{J-1} \\ &= \ldots \\ &= E_i(C_{i,J-i}) f_{J-i} \cdot \ldots \cdot f_{J-1} \\ &= C_{i,J-i} f_{J-i} \cdot \ldots \cdot f_{J-1} \end{aligned}$$

a tím je rovnost (1.6) dokázána.

□

Předchozí věta nám ukazuje, že odhad $\hat{C}_{i,J}$ má stejný tvar jako $E(C_{i,J}|\Delta)$, který je odhadem $C_{i,J}$ založeném na pozorování Δ.

Věta 2. *Za předpokladů (MCL1) a (MCL2) jsou odhady* $\hat{f}_j$, $j = 0, \ldots, J-1$, *nestranné a nekorelované a platí rovnost*

$$E(\hat{f}_{J-i} \cdot \ldots \cdot \hat{f}_{J-1}) = f_{J-i} \cdot \ldots \cdot f_{J-1}. \tag{1.7}$$

Důkaz. (Mack [10])

Označme

$$\Delta_j = \{C_{i,k}; k \leq j, i + k \leq J\},$$

pro $j = 0, \ldots, J$, jako množinu známých pozorování výší škod, které mají nejvýše j let zpoždění v hlášení.

Pomocí předpokladů (MCL1) a (MCL2) dostáváme

$$E(C_{i,j+1}|\Delta_j) = E(C_{i,j+1}|C_{i,0}\dots C_{i,j}) = f_j C_{i,j}.$$

Nyní vyjdeme z linearity podmíněné střední hodnoty a z obou předpokladů (MCL1) a (MCL2). Tedy můžeme psát

$$\begin{aligned}
E(\hat{f}_j|\Delta_j) &= E\left(\frac{\sum_{k=0}^{J-k-1} C_{k,j+1}}{\sum_{k=0}^{J-k-1} C_{k,j}}\middle|\Delta_j\right) \\
&= E\left(\frac{\sum_{k=0}^{J-k-1} (C_{k,j+1}|\Delta_j)}{\sum_{k=0}^{J-k-1} C_{k,j}}\right) \\
&= \frac{E\left(\sum_{k=0}^{J-k-1} (C_{k,j+1}|\Delta_j)\right)}{\sum_{k=0}^{J-k-1} C_{k,j}} \\
&= \frac{\sum_{k=0}^{J-k-1} E(C_{k,j+1}|\Delta_j)}{\sum_{k=0}^{J-k-1} C_{k,j}} \\
&= \frac{\sum_{k=0}^{J-k-1} E(C_{k,j+1}|C_{k,1},\dots,C_{k,j})}{\sum_{k=0}^{J-k-1} C_{k,j}} \\
&= \frac{\sum_{k=0}^{J-k-1} f_j C_{k,j}}{\sum_{k=0}^{J-k-1} C_{k,j}} \\
&= f_j,
\end{aligned}$$

čímž je dokázána nestrannost odhadů vývojových faktorů, jelikož

$$E(\hat{f}_j) = E\left(E(\hat{f}_j|\Delta_j)\right) = f_j,$$

pro $j = 0, \dots, J-1$.

Odhady jsou také nekorelované, bez újmy na obecnosti předpokládejme $k \geq j$ a pomocí právě dokázané nestrannosti a lemmatu 2 platí

$$\begin{aligned} E(\hat{f}_k \hat{f}_j) &= E(E(\hat{f}_k \hat{f}_j|\Delta_j)) \\ &= E(\hat{f}_k E(\hat{f}_j|\Delta_j)) \\ &= E(\hat{f}_k f_j) \\ &= E(\hat{f}_k) f_j \\ &= E(\hat{f}_k) E(\hat{f}_j) \\ &= f_k f_j. \end{aligned}$$

Rovnost (1.7) vyplývá z nestrannosti a nekorelovanosti odhadů parametrů

$$E(\hat{f}_{J-i} \cdot \ldots \cdot \hat{f}_{J-1}) = E(\hat{f}_{J-i} \cdot \ldots \cdot \hat{f}_{J-2}) E(\hat{f}_{J-1}|\Delta_{J-1}) = \ldots = f_{J-i} \cdot \ldots \cdot f_{J-1}.$$

□

Nekorelovanost $\hat{f}_j$ je překvapující, neboť $\hat{f}_{j-1}$ a $\hat{f}_j$ závisí na stejných datech $C_{1,j} + \ldots + C_{J-i,j}$. Je však pro nás velmi důležitá, neboť díky ní dostáváme nestranné odhady pro $C_{i,J}$ a R_i, jak dokazuje následující věta.

Věta 3. *Nechť $\Delta = \{C_{i,j}; i + j \leq J\}$ je množina všech doposud napozorovaných dat. Nechť platí předpoklady (MCL1) a (MCL2). Potom platí:*

I. $\hat{C}_{i,J}$ je nestranným odhadem $E\left(C_{i,J}|\Delta\right)$, kde

$$\begin{aligned} \hat{C}_{i,J} &= C_{i,J-i} \hat{f}_{J-i} \cdot \ldots \cdot \hat{f}_{J-1} \\ E\left(C_{i,J}|\Delta\right) &= C_{i,J-i} f_{J-i} \cdot \ldots \cdot f_{J-1}. \end{aligned}$$

II. Odhad rezervy $\hat{R}_i = \hat{C}_{i,J}$–$C_{i,J-i}$ je nestranným odhadem skutečné výše rezervy $R_i = C_{i,J}$–$C_{i,J-i}$.

Důkaz. Označme

$$\Delta_j = \{C_{i,k}; k \leq j, i + k \leq J\},$$

pro $j = 0, \ldots, J$, jako množinu známých pozorování výší škod, které mají nejvýše j let zpoždění v hlášení.

Vyjdeme z dokázaných předchozích vztahů a lemmatu 2. Nejprve dokážeme vztah I.:

$$\begin{aligned} E\left(\hat{C}_{i,J}\right) &= E\left(E\left(\hat{C}_{i,J}|\Delta_{J-1}\right)\right) \\ &= E\left(E\left(C_{i,J-i}\hat{f}_{J-i} \cdot \ldots \cdot \hat{f}_{J-1}|\Delta_{J-1}\right)\right) \\ &= E\left(E\left(C_{i,J-i}\hat{f}_{J-i} \cdot \ldots \cdot \hat{f}_{J-2}\right) E\left(\hat{f}_{J-1}|\Delta_{J-1}\right)\right) \\ &= E\left(E\left(C_{i,J-i}\hat{f}_{J-i} \cdot \ldots \cdot \hat{f}_{J-3}\right) E\left(\hat{f}_{J-2}|\Delta_{J-2}\right) f_{J-1}\right) \\ &= \ldots \\ &= E\left(E\left(C_{i,J-i}|\Delta_{J-i}\right)\right) f_{J-i} \cdot \ldots \cdot f_{J-1} \\ &= C_{i,J-i} f_{J-i} \cdot \ldots \cdot f_{J-1}. \end{aligned}$$

Poté vztah II.:

$$\begin{aligned}
E\left(\hat{R}_i\right) &= E\left(E\left(\hat{R}_i|\Delta\right)\right) \\
&= E\left(E\left(\hat{C}_{i,J}-C_{i,J-i}|\Delta\right)\right) \\
&= E\left(E\left(\hat{C}_{i,J}|\Delta\right) - E\left(C_{i,J-i}|\Delta\right)\right) \\
&= C_{i,J} - C_{i,J-i}
\end{aligned}$$

a tím je důkaz hotov.

□

Vlastnosti odhadu σ_j^2 shrneme v následující větě.

Věta 4. *Za předpokladu (MCL3) je $\hat{\sigma}_j^2$ nestranným odhadem σ_j^2, kde*

$$\hat{\sigma}_j^2 = \frac{1}{J-j-1} \sum_{i=0}^{J-j-1} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j\right)^2,$$

pro $j = 0, \ldots, J-2$.

Důkaz. Vyjdeme ze vztahu $E\left(\frac{C_{i,j+1}}{C_{i,j}}|C_{i,0}, \ldots, C_{i,j}\right) = f_j$. Pak dostáváme

$$\begin{aligned}
E\hat{\sigma}_j^2 &= E\left(\frac{1}{J-j-1} \sum_{i=0}^{J-j-1} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j\right)^2\right) \\
&= E\left(\frac{1}{J-j-1} \sum_{i=0}^{J-j-1} E\left(C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j\right)^2 |C_{i,0}, \ldots, C_{i,j}\right)\right) \\
&= E\frac{1}{J-j-1} \sum_{i=0}^{J-j-1} C_{i,j} var\left(\frac{C_{i,j+1}}{C_{i,j}}|C_{i,0}, \ldots, C_{i,j}\right) \\
&= E\frac{1}{J-j-1} \sum_{i=0}^{J-j-1} C_{i,j} \frac{\sigma_j^2}{C_{i,j}} \\
&= \sigma_j^2,
\end{aligned}$$

čímž je důkaz dokončen.

□

Pomocí vztahu (1.5) jsme schopni odhadnout $\sigma_0^2, \ldots, \sigma_{J-2}^2$, zbývá nám tedy najít odhad pro σ_{J-1}^2. Postup výpočtu odhadu σ_{J-1}^2 je navržen v [10]. Pokud $\hat{f}_{J-1} = 1$ a vývoj škod je po $J-1$ letech považován za ukončený, potom položíme $\hat{\sigma}_{J-1}^2 = 0$. V opačném případě získáme odhad σ_{J-1}^2 extrapolací předchozích hodnot $\hat{\sigma}_0^2, \ldots, \hat{\sigma}_{J-2}^2$, které tvoří klesající posloupnost. Odhad získáme například logaritmicko-lineární regresí nebo požadujeme, aby $\frac{\hat{\sigma}_{J-3}^2}{\hat{\sigma}_{J-2}^2} = \frac{\hat{\sigma}_{J-2}^2}{\hat{\sigma}_{J-1}^2}$. Poslední možnost vede k požadavku, aby

$$\hat{\sigma}_{J-1}^2 = min\left(\frac{\hat{\sigma}_{J-2}^4}{\hat{\sigma}_{J-3}^2}, min\left(\hat{\sigma}_{J-3}^2, \hat{\sigma}_{J-2}^2\right)\right). \tag{1.8}$$

1.2 Výpočet střední kvadratické chyby a standardní chyby

V úvodu kapitoly bylo vysvětleno, proč je střední kvadratická odchylka důležitá a k čemu se využívá. V následující části zavedeme definici odchylky odhadů výší škod, ze které budeme vycházet a ukážeme si, jak se počítá.

Definice 1. *Střední kvadratickou chybu (mean squared error) odhadu celkové výše rezervy na pojistná plnění $\hat{C}_{i,J}$ definujeme jako*

$$mse(\hat{C}_{i,J}) = E\left((\hat{C}_{i,J} - C_{i,J})^2|\Delta\right), \tag{1.9}$$

kde $\Delta = \{C_{i,j}; i + j \leq J\}$ je množina všech doposud napozorovaných dat.

Definice 2. *Standardní odchylka (standard error) odhadu $\hat{C}_{i,J}$ je definována jako druhá odmocnina ze střední kvadratické chyby odhadu $\hat{C}_{i,J}$, tj.*

$$se(\hat{C}_{i,J}) = \sqrt{mse(\hat{C}_{i,J})}. \tag{1.10}$$

Můžeme vidět, že

$$\begin{aligned} mse(\hat{R}_i) &= E\left((\hat{R}_i - R_i)^2|\Delta\right) \\ &= E\left((\hat{C}_{i,J} - C_{i,J-i} - C_{i,J} + C_{i,J-i})^2|\Delta\right) \\ &= E\left((\hat{C}_{i,J} - C_{i,J})^2|\Delta\right) \\ &= mse(\hat{C}_{i,J}). \end{aligned}$$

Dále při použití pravidla $E(X - a)^2 = var(X) + (E(X)–a)^2$ a vlastností podmíněné střední hodnoty máme

$$mse(\hat{C}_{i,J}) = var(C_{i,J}|\Delta) + \left(E(C_{i,J}|\Delta) - \hat{C}_{i,J}\right)^2, \tag{1.11}$$

což ukazuje, že střední kvadratická chyba se rovná součtu rozptylu procesu (process variance) a chyby odhadu (estimation error).

Věta 5. *Nechť platí předpoklady (MCL1), (MCL2) a (MCL3). Potom odhad střední kvadratické chyby je dán vztahem*

$$\widehat{mse}(\hat{R}_i) = \hat{C}_{i,J}^2 \sum_{k=J-i}^{J-1} \frac{\hat{\sigma}_k^2}{\hat{f}_k^2} \left(\frac{1}{\hat{C}_{i,k}} + \frac{1}{\sum\limits_{j=1}^{J-k-1} C_{j,k}} \right), \tag{1.12}$$

kde $\hat{C}_{i,j} = C_{i,J-i}\hat{f}_{I-i} \cdot \ldots \cdot \hat{f}_{j-1}$, $j \geq J - i$, jsou odhady budoucích hodnot $C_{i,j}$ a dále uvažujeme $\hat{C}_{i,J-i} = C_{i,J-i}$.

Důkaz. (Mack [10])

Pro zjednodušení zápisu zavedeme následující značení

$$\begin{aligned} E_i(X) &= E(X|C_{I,1},\dots,C_{I,J+1-i}),\\ var_i(X) &= var(X|C_{I,1},\dots,C_{I,J+1-i}). \end{aligned}$$

Dále vyjdeme ze vztahu

$$mse(\hat{R}_i) = var(C_{i,J}|\Delta) + \left(E(C_{i,J}|\Delta) - \hat{C}_{i,J}\right)^2.$$

Opakovanou aplikací předpokladů (MCL1) a (MLC3) metody Chain-Ladder a užitím vztahu pro nepodmíněný rozptyl $varX = E[var(X|Y)] + var(E[X|Y])$ dostáváme pro první člen $mse(\hat{R}_i)$

$$\begin{aligned} var(C_{i,J}|\Delta) &= var_i(C_{i,J})\\ &= E_i\left(var(C_{i,J}|C_{i,1},\dots,C_{i,J-1})\right) + var_i\left(E(C_{i,J}|C_{i,1},\dots,C_{i,J-1})\right)\\ &= E_i(C_{i,J-1})\sigma^2_{J-1} + var_i(C_{i,J-1})f^2_{J-1}\\ &= E_i(C_{i,J-2})f_{J-2}\sigma^2_{J-1} + E_i(C_{i,J-2})\sigma^2_{J-2}f^2_{J-1} + var_i(C_{i,J-2})f^2_{J-2}f^2_{J-1}\\ &= \dots\\ &= C_{i,J-i}\sum_{k=J-i}^{J-1} f_{J-i}\cdot\ldots\cdot f_{k-1}\sigma^2_k f^2_{k+1}\cdot\ldots\cdot f^2_{J-1}, \end{aligned}$$

neboť $var_i(C_{i,J-i}) = 0$.

Druhý člen $mse(\hat{R}_i)$ získáme díky větě 1

$$\left(E(C_{i,J}|\Delta) - \hat{C}_{i,J}\right)^2 = C_{i,J-i}\left(f_{J-i}\cdot\ldots\cdot f_{J-1} - \hat{f}_{J-i}\cdot\ldots\cdot\hat{f}_{J-1}\right)^2. \quad (1.13)$$

V praxi bychom chtěli najít odhady pro tyto dva členy $mse(\hat{R}_i)$. V prvním členu stačí nahradit neznámé parametry f_k a σ^2_k jejich odhady $\hat{f}_k$ a $\hat{\sigma}^2_k$. Tedy pro odhad $var(C_{i,J}|\Delta)$ dostáváme

$$\widehat{var(C_{i,J}|\Delta)} = C_{i,J-i}\sum_{k=J-i}^{J-1}\hat{f}_{J-i}\cdot\ldots\cdot\hat{f}_{k-1}\hat{\sigma}^2_k\hat{f}^2_{k+1}\cdot\ldots\cdot\hat{f}^2_{J-1} = \hat{C}^2_{i,J}\sum_{k=J-i}^{J-1}\frac{\hat{\sigma}^2_k/\hat{f}^2_k}{\hat{C}_{i,k}}, \quad (1.14)$$

kde jsme použili zápis $\hat{C}_{i,k} = C_{i,J-i}\hat{f}_{I-i}\cdot\ldots\cdot\hat{f}_{k-1}$.

Bohužel ve druhém členu $mse(\hat{R}_i)$ nemůžeme nahradit f_k jejich odhady $\hat{f}_k$, jelikož by celý výraz byl nulový. Použijeme zde jiný postup. Můžeme psát

$$\begin{aligned} F &= f_{J-i}\cdot\ldots\cdot f_{J-1} - \hat{f}_{J-i}\cdot\ldots\cdot\hat{f}_{J-1}\\ &= S_{J-i} + \ldots + S_{J-1}, \end{aligned}$$

kde $S_k = \hat{f}_{J-i}\cdot\ldots\cdot\hat{f}_{k-1}(f_k - \hat{f}_k)f_{k+1}\cdot\ldots\cdot f_{J-1}$.

Tedy

$$
\begin{aligned}
F^2 &= \left(S_{J-i} + \ldots + S_{J-1}\right)^2 \\
&= \sum_{k=J-i}^{J-1} S_k^2 + 2\sum_{j<k} S_j S_k.
\end{aligned}
$$

Nyní nahradíme S_k^2 podmíněnou střední hodnotou $E(S_k^2|\Delta_k)$ a součin S_jS_k nahradíme $E(S_jS_k|\Delta_k)$, pro $j<k$. To znamená, že S_k^2 a S_jS_k přiblížíme průměrováním přes co nejmenší počet dat tak, aby co nejvíce hodnot $C_{i,k}$ z pozorovaných dat zůstalo pevnými. Vzhledem k tomu, že platí $E(f_k - \hat{f}_k|\Delta_k) = 0$ dostáváme rovnost $E(S_jS_k|\Delta_k) = 0$, pro $j<k$.

Protože

$$
\begin{aligned}
E\left((f_k - \hat{f}_k)^2|\Delta_k\right) &= var(\hat{f}_k|\Delta_k) \\
&= \frac{\sum_{l=0}^{J-k-1} var(C_{l,k+1}|\Delta_k)}{\left(\sum_{l=0}^{J-k-1} C_{l,k}\right)^2} \\
&= \frac{\sigma_k^2}{\sum_{l=0}^{J-k-1} C_{l,k}},
\end{aligned}
$$

dostáváme

$$
E(S_k^2|\Delta_k) = \frac{\hat{f}_{J-i}^2 \cdot \ldots \cdot \hat{f}_{k-1}^2 \hat{\sigma}_k^2 \hat{f}_{k+1}^2 \cdot \ldots \cdot \hat{f}_{J-1}^2}{\sum_{l=0}^{J-k-1} C_{l,k}}.
$$

Nakonec $F^2 = \left(S_{J-i} + \ldots + S_{J-1}\right)^2$ nahradíme výrazem $\sum_k E(S_k^2|\Delta_k)$. Jelikož jsou všechny členy v této sumě kladné, můžeme všechny neznámé parametry f_k a σ_k^2 nahradit jejich nestrannými odhady $\hat{f}_k$ a $\hat{\sigma}_k^2$.

Výraz $F^2 = \left(f_{J-i} \cdot \ldots \cdot f_{J-1} - \hat{f}_{J-i} \cdot \ldots \cdot \hat{f}_{J-1}\right)^2$ lze odhadnout jako

$$
\begin{aligned}
\hat{F}^2 &= \sum_{k=J-i}^{J-1} \frac{\hat{f}_{J-i}^2 \cdot \ldots \cdot \hat{f}_{k-1}^2 \hat{\sigma}_k^2 \hat{f}_{k+1}^2 \cdot \ldots \cdot \hat{f}_{J-1}^2}{\sum_{l=0}^{J-k-1} C_{l,k}} \\
&= \hat{f}_{J-i}^2 \cdot \ldots \cdot \hat{f}_{k-1}^2 \sum_{k=J-i}^{J-1} \frac{\hat{\sigma}_k^2/\hat{f}_k^2}{\sum_{l=0}^{J-k-1} C_{l,k}}.
\end{aligned}
$$

Dosazením do (1.13) dostáváme odhad

$$
\left(E(C_{i,J}|\Delta) - \hat{C}_{i,J}\right)^2 = C_{i,J-i} \hat{f}_{J-i}^2 \cdot \ldots \cdot \hat{f}_{k-1}^2 \sum_{k=J-i}^{J-1} \frac{\hat{\sigma}_k^2/\hat{f}_k^2}{\sum_{l=0}^{J-k-1} C_{l,k}}. \tag{1.15}
$$

Sečtením (1.14) a (1.15) dostáváme vztah pro odhad $mse(\hat{R}_i)$ a tím je věta 5 dokázána.

$\square$

Předchozí věta udává vztah pro odhad nejistoty při stanovování hodnoty rezervy pro jednotlivé vznikové roky.

Často je důležitá také střední kvadratická chyba celkového součtu odhadů rezerv $\hat{R} = \hat{R}_0 + \ldots + \hat{R}_J$. V tomto případě však nelze pouze sečíst jednotlivé hodnoty $\widehat{mse}(\hat{R}_i)$, $i = 0, \ldots, J$, neboť tyto hodnoty jsou vzájemně korelované. Následující věta slouží pro výpočet střední kvadratické chyby celkové rezervy.

Věta 6. *Za předpokladů uvedených v předchozí větě (Věta 5) získáme odhad střední kvadratické chyby celkové rezervy $\hat{R} = \hat{R}_0 + \ldots + \hat{R}_J$ jako*

$$\widehat{mse}(\hat{R}) = \sum_{i=0}^{J} \left[\widehat{mse}(\hat{R}_i) + \hat{C}_{i,J} \left(\sum_{j=i+1}^{J} \hat{C}_{j,J} \right) \sum_{k=J-i}^{J-1} \frac{2\hat{\sigma}_k^2 / \hat{f}_k^2}{\sum_{n=0}^{J-k-1} C_{n,k}} \right]. \tag{1.16}$$

Důkaz. (Mack [10])

Máme

$$\begin{aligned}
\widehat{mse}\left(\sum_{i=0}^{J} \hat{R}_i\right) &= E\left(\left(\sum_{i=0}^{J} \hat{R}_i - \sum_{i=0}^{J} R_i\right)^2 |\Delta\right) \\
&= E\left(\left(\sum_{i=0}^{J} \hat{C}_{i,J} - \sum_{i=0}^{J} C_{i,J}\right)^2 |\Delta\right) \\
&= var\left(\sum_{i=0}^{J} C_{i,J}|\Delta\right) + \left(E\left(\sum_{i=0}^{J} C_{i,J}|\Delta\right) - \sum_{i=0}^{J} \hat{C}_{i,J}\right)^2.
\end{aligned}$$

Vidíme, že střední kvadratická chyba celkové rezervy je rovna součtu rozptylu procesu a chyby odhadu. Díky nezávislosti jednotlivých let vzniku škody lze rozptyl procesu přepsat jako

$$var\left(\sum_{i=0}^{J} C_{i,J}|\Delta\right) = \sum_{i=0}^{J} var\,(C_{i,J}|\Delta),$$

jehož sčítance jsme vypočetli v důkazu věty 5.

Sčítanec pro chybu odhadu přepíšeme jako

$$\begin{aligned}
\left(E\left(\sum_{i=0}^{J} C_{i,J}|\Delta\right) - \sum_{i=0}^{J} \hat{C}_{i,J}\right)^2 &= \left(\sum_{i=0}^{J} \left(E\,(C_{i,J}|\Delta) - \hat{C}_{i,J}\right)\right)^2 \\
&= \sum_{i,j} \left(E\,(C_{i,J}|\Delta) - \hat{C}_{i,J}\right)\left(E\,(C_{j,J}|\Delta) - \hat{C}_{j,J}\right) \\
&= \sum_{i,j} C_{i,J+1-i} C_{j,J+1-j} F_i F_j,
\end{aligned}$$

kde $F_i = f_{J-i} \cdot \ldots \cdot f_{J-1} - \hat{f}_{J-i} \cdot \ldots \cdot \hat{f}_{J-1}$.

Střední kvadratická chyba rezervy pro rok i je rovna

$$\widehat{mse}(\hat{R}_i) = var(C_{i,J}|\Delta) + (C_{i,J}F_i)^2.$$

Pro střední kvadratickou chybu celkové rezervy musíme uvažovat, že rezervy pro jednotlivé roky vzniku jsou korelované. Tedy

$$\widehat{mse}\left(\sum_{i=0}^{J} \hat{R}_i\right) = \sum_{i=0}^{J} \widehat{mse}(\hat{R}_i) + \sum_{1 \le i \le j \le J} 2C_{i,J-i}C_{j,J-j}F_iF_j.$$

Nakonec k nalezení odhadu F_iF_j, $i < j$, použijeme vztah

$$\hat{F}_i\hat{F}_j = \sum_{k=J-i}^{J-1} \frac{\hat{f}_{J-j} \cdot \ldots \cdot \hat{f}_{J-i}\hat{f}_{J-i}^2 \cdot \ldots \cdot \hat{f}_{k-1}^2 \hat{\sigma}_k^2 \hat{f}_{k+1}^2 \cdot \ldots \cdot \hat{f}_{J-1}^2}{\sum_{n=0}^{J-k-1} C_{n,k}},$$

který je analogický odhadu pro F z důkazu věty 5 . □

1.3 Rekurzivní výpočet standardní chyby a zahrnutí tail faktoru

V předchozí části jsme ukázali výpočet standardní chyby pro data reprezentující výše škod za předpokladu, že vývoj škod pozorujeme do J let od vzniku škody. Nyní se zaměříme na odvození rekurzivního vztahu standardní chyby a zahrnutí tail faktoru do rekurze zobecněním Mackova modelu. Vycházíme z článku [11].

Nechť $C_{i,j}$ označuje celkový úhrn pojistného plnění za škody, které vznikly v roce i a byly uhrazeny (nahlášeny) do konce vývojového roku j, kde $i = 0, 1, \ldots, I$, $j = 0, 1, \ldots, J$, $I = J$. Známe pozorování $C_{i,j}$ pro $j \le J - i$, ostatní hodnoty budeme předpovídat. Algoritmus metody Chain-Ladder obsahuje vztah pro předpovídání

$$\hat{C}_{i,j+1} = \hat{f}_j \hat{C}_{i,j}, \tag{1.17}$$

počínaje od $\hat{C}_{i,J-i} = C_{i,j-i}$.

Součástí je nalezení odhadů vývojových faktorů f_j pomocí vztahu

$$\hat{f}_j = \frac{\sum_{i=0}^{J-1-j} w_{i,j} C_{i,j}^{\alpha} F_{i,j}}{\sum_{i=0}^{J-1-j} w_{i,j} C_{i,j}^{\alpha}} \tag{1.18}$$

pro $j = 0, \ldots, J-1$, $\alpha \in \{0, 1, 2\}$, kde

$$F_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$$

jsou jednotlivé roční faktory a kde

$$w_{i,j} \in [0,1]$$

jsou libovolné váhy, které mohou být použity pro odlišné $F_{i,j}$.

Obvykle $w_{i,j} = 1$ pro všechna i, j. Pak pro $\alpha = 1$ dostáváme historické roční faktory $F_{i,j}$, pro $\alpha = 0$ rovnou získáme průměr pozorovaných jednotlivých vývojových faktorů rozvoje f_j. Pro $\alpha = 2$ ukážeme, že se jedná o odhad pomocí lineární regrese $C_{i,j+1}$ proti $C_{i,j}$.

Na základě vztahu (1.17) uvažujme lineární závislost

$$C_{i,j+1} = f_j C_{i,j}$$

pro $j = 0, \ldots, J-1$, $i = 0, \ldots, J-1-j$ a neznámé f_j.

Součet čtverců je dán vztahem

$$S\left(\hat{f}_j\right) = \sum_{i=0}^{J-1-j} \left(\hat{f}_j C_{i,j} - C_{i,j+1}\right)^2.$$

K nalezení minima součtu čtverců položíme parciální derivaci rovnu nule, tj.

$$\frac{\partial S\left(\hat{f}_j\right)}{\partial \hat{f}_j} = 2 \sum_{i=0}^{J-1-j} \left(\hat{f}_j C_{i,j} - C_{i,j+1}\right) C_{i,j} = 0.$$

Úpravami obdržíme soustavu

$$\hat{f}_j \sum_{i=0}^{J-1-j} C_{i,j}^2 = \sum_{i=0}^{J-1-j} C_{i,j} C_{i,j+1}.$$

Řešením předchozí soustavy dostaneme hledaný odhad parametru f_j

$$\hat{f}_j = \frac{\sum_{i=0}^{J-1-j} C_{i,j} C_{i,j+1}}{\sum_{i=0}^{J-1-j} C_{i,j}^2}.$$

Nakonec ukážeme, že ke stejnému odhadu vývojových faktorů dospějeme dosazením $w_{i,j} = 1$ a $\alpha = 2$ do vztahu (1.18)

$$
\begin{aligned}
\hat{f}_j &= \frac{\sum_{i=0}^{J-1-j} C_{i,j}^2 F_{i,j}}{\sum_{i=0}^{J-1-j} C_{i,j}^2} \\
&= \frac{\sum_{i=0}^{J-1-j} C_{i,j}^2 \frac{C_{i,j+1}}{C_{i,j}}}{\sum_{i=0}^{J-1-j} C_{i,j}^2} \\
&= \frac{\sum_{i=0}^{J-1-j} C_{i,j} C_{i,j+1}}{\sum_{i=0}^{J-1-j} C_{i,j}^2}.
\end{aligned}
$$

Výše popsané pravidlo vede ke stejnému tvaru předpovědi $\hat{C}_{i,J}$ jakou jsme dostali u (1.3), tj.

$$\hat{C}_{i,J} = C_{i,J-i}\hat{f}_{J-i} \cdot \ldots \cdot \hat{f}_{J-1}.$$

Vývoj škod pro rok vzniku i nemusí být ukončen v roce J. Z tohoto důvodu používáme tail faktor $\hat{f}_{ult} \geq 1$ za účelem odhadnout konečnou výši škod $C_{i,ult}$ pomocí vztahu

$$\hat{C}_{i,ult} = \hat{f}_{ult}\hat{C}_{i,J}. \tag{1.19}$$

Tail faktor $\hat{f}_{ult}$ můžeme odhadnout pomocí vztahu

$$\hat{f}_{ult} = \prod_{k=J}^{\infty} \hat{f}_k, \tag{1.20}$$

kde $\hat{f}_k$, $k \geq J$ odhadneme lineární extrapolací $\ln(\hat{f}_k - 1)$ přímkou $a \cdot k + b$, $a \leq 0$.

V Mackově modelu byl odvozen vztah pro standardní chybu předpovědi $\hat{C}_{i,j}$ pro $\alpha = 1$ a $w_{i,j} = 1$ pro všechny i, j. V další části tento vztah zobecníme pro případ $\alpha = 0$ nebo $\alpha = 2$ a $w_{i,j} \leq 1$.

Zobecněné předpoklady Mackovy metody

(MCLR1) Existují vývojové faktory, pro které platí

$$E\left[F_{i,j} | C_{i,1}, \ldots, C_{i,j}\right] = f_j,$$

kde $i = 0, \ldots, I-1$, $j = 0, \ldots, J-1$.

(MCLR2) Jednotlivé roky vzniků škod jsou nezávislé, tj.

$$(C_{i,0}, \ldots, C_{i,J}) \text{ a } (C_{k,0}, \ldots, C_{k,J})$$

jsou vzájemně nezávislé náhodné vektory pro $i \neq k$.

(MCLR3) Pro podmíněný rozptyl platí

$$var\left[F_{i,j}|C_{i,1},\dots,C_{i,j}\right] = \frac{\sigma_j^2}{w_{i,j}C_{i,j}^{\alpha}},$$

kde $i = 0,\dots,I-1$, $j = 0,\dots,J-1$, σ_j^2 je neznámý parametr a pro dané $\alpha \in \{0,1,2\}$, $w_{i,j} \in [0,1]$.

Model (MCLR1) – (MCLR3) nazýváme základní model Chain-Ladder. V tomto modelu také platí následující pravidla

- $E\left[C_{i,j+1}|C_{i,1},\dots,C_{i,j}\right] = f_j C_{i,j}$,
- $E\left[C_{i,J}|C_{i,1},\dots,C_{i,j}\right] = C_{i,J+1-i} f_{J-i}\dots f_{J-1}$,
- $\hat{f}_j$ je nestranný odhad f_j s minimálním rozptylem (pro daná α a $w_{i,j}$),
- $\hat{f}_{J-i}\cdot\ldots\cdot\hat{f}_{J-1}$ je nestranný odhad $f_{J-i}\cdot\ldots\cdot f_{J-1}$.

Dále platí, že

$$\hat{\sigma}_j^2 = \frac{1}{J-j-1}\sum_{i=0}^{J-j-1} w_{i,j}C_{i,j}^{\alpha}\left(F_{i,j}-\hat{f}_j\right)^2, \tag{1.21}$$

pro $j = 0,\dots,J-2$, je nestranný odhad σ_j^2, který můžeme doplnit vztahem (1.8).

Thomas Mack [11] v roce 1999 odvodil vztah pro standardní chybu $\hat{C}_{i,J}$, která je standardní chybou odhadu $\hat{R}_i = \hat{C}_{i,J} - C_{i,J-i}$ skutečné výše rezervy $R_i = C_{i,J} - C_{i,J-i}$. Tento vztah odpovídá vzorci (1.12) ve větě 4, který můžeme přepsat do tvaru

$$se(\hat{C}_{i,J})^2 = \hat{C}_{i,J}^2 \sum_{k=J-i}^{J-1} \frac{\left(se(F_{i,k})\right)^2 - \left(se(\hat{f}_k)\right)^2}{\hat{f}_k^2}, \tag{1.22}$$

kde $\left(se(F_{i,k})\right)^2$ je odhad $var(F_{i,k}|C_{i,1},\dots,C_{i,k})$ a $\left(se(\hat{f}_k)\right)^2$ je odhad

$$var(\hat{f}_k) = \frac{\sigma_k^2}{\sum_{j=0}^{J-k-1} w_{j,k}C_{j,k}^{\alpha}}.$$

Analogickou aplikací důkazu věty 4 pro $\alpha = 0$ a $w_{i,j} = 1$ dojdeme k závěru, že vztah (1.22) platí také pro $\alpha = 0$ nebo $\alpha = 2$ a jakoukoliv volbu $w_{i,j} \in [0,1]$. Kromě výše zmíněného můžeme aplikací důkazu získat rekurzivní vzorec pro výpočet $se(\hat{C}_{i,j+1})^2$, který je snadněji naprogramovatelný. Rekurzivní vzorec je dán vztahem

$$se(\hat{C}_{i,j+1})^2 = \hat{C}_{i,j}^2\left(\left(se(F_{i,j})\right)^2 - \left(se(\hat{f}_j)\right)^2\right) + se(\hat{C}_{i,j})^2\hat{f}_k^2 \tag{1.23}$$

s počáteční hodnotou $se(\hat{C}_{i,J-i}) = 0$ [11].

Tato rekurze, která vede ke vztahu (1.22), je hodně intuitivní: $(se(F_{i,j}))^2$ odhaduje (kvadratickou) náhodnou chybu $var(F_{i,j}) = E(F_{i,j} - f_j)^2$, tj. střední kvadratickou odchylku jednotlivých $F_{i,j}$ od jejich průměrů f_j, a $\left(se(\hat{f}_j)\right)^2$ odhaduje (kvadratickou) chybu odhadu $var(\hat{f}_j) = E(\hat{f}_j - f_j)^2$, tj. střední kvadratickou odchylku odhadovaných průměrů $\hat{f}_j$ a $F_{i,j}$, $i = 0, \ldots, J$ od f_j. Z našeho výkladu je zřejmé, že platí nerovnost $var(\hat{f}_j) \leq var(F_{i,j})$, pokud odhad $\hat{f}_j$ je nestranný a škodní rok i patří do let, přes které průměrujeme $\hat{f}_j$ [11].

Zahrnutí tail faktoru

Rekurzi můžeme rozšířit o zahrnutí tail faktoru $\hat{f}_{ult}$:

$$se(\hat{C}_{i,ult})^2 = \hat{C}_{i,J}^2 \left((se(F_{i,ult}))^2 - \left(se(\hat{f}_{ult})\right)^2 \right) + se(\hat{C}_{i,J})^2 \hat{f}_{ult}^2. \tag{1.24}$$

Poznamenejme, že pojistní matematici, kteří vyvíjejí odhad pro f_{ult} mohou také vypracovat odhad $se(\hat{f}_{ult})$ pro chybu odhadu $\sqrt{var(\hat{f}_{ult})}$ a odhad $se(F_{i,ult})$ pro odpovídající náhodnou chybu $\sqrt{var(F_{i,ult})}$. Jinými slovy, jak hodně je v průměru vychýlené $\hat{f}_{ult}$ od f_{ult} a $F_{i,ult}$ od f_{ult}.

Obvykle budeme moci najít index $j \leq J$, pro který

$$\hat{f}_{j-1} > \hat{f}_{ult} > \hat{f}_j.$$

Potom můžeme zkontrolovat, zda je rozumné předpokládat platnost nerovnic

$$se(\hat{f}_{j-1}) > se(\hat{f}_{ult}) > se(\hat{f}_j)$$

$$se(\hat{F}_{i,j-1}) > se(\hat{F}_{i,ult}) > se(\hat{F}_{i,j})$$

nebo zda je lepší nastavit $se(\hat{f}_{ult})$ a/nebo $se(F_{i,j})$ mimo tyto nerovnosti.

Nakonec dostáváme rekurzivní vztah pro standardní chybu rezervy přes všechny škodní roky

$$\begin{aligned} \left(se\left(\hat{R}\right)\right)^2 = \quad & \left(se\left(\sum_{i=J+1-j}^{J} \hat{C}_{i,j}\right)\right)^2 \cdot \hat{f}_j^2 \\ & + \sum_{i=J-j}^{J} \hat{C}_{i,j} \cdot (se(F_{i,j}))^2 \\ & + \left(\sum_{i=J-j}^{J} \hat{C}_{i,j}\right)^2 \cdot \left(se(\hat{f}_j)\right)^2, \end{aligned} \tag{1.25}$$

počínaje od $j = 1$. Tento vztah můžeme též získat z důkazu věty 4.

Poznamenejme, že některé statistiky (např. $\hat{f}_{ult}$, $s.e(\hat{f}_{ult})$) není možné odhadnout exaktně matematicky, ale je nutné přistoupit k heuristickým postupům. Obvykle záleží na úsudku aktuára, zabývajícího se daným problémem.

2. Zobecněné lineární modely

V posledních letech byla velká pozornost věnována vztahu mezi různými stochastickými modely a metodou Chain-Ladder. Stochastické modely byly zkonstruovány s cílem vytvořit přesně stejné odhady rezerv jako tradiční deterministická metoda Chain-Ladder. Na první pohled se to může zdát jako nesmyslné cvičení: proč používat komplexní stochastickou metodu k nalezení odhadu rezervy, když by stačila jednoduchá deterministická metoda? Odpověd je, že stejně jako odhad rezervy existují i jiné důležité aspekty modelu. Těmi jsou například základní distribuční předpoklady odhadovaného modelu nebo odhady variability odhadů parametrů. Je také užitečné vědět, kde se data liší od odhadovaného modelu, a mít představu, jak mohou být jiné modely odhadovány a porovnávány [5].

Doposud byly navrhnuty dva modely, které poskytují odhady rezerv a které jsou shodné s odhady, které poskytuje deterministická metoda Chain-Ladder a umožňují vypočítat odhady variability rezerv. Jedná se o Mackův model [10] a Renshawův a Verrallův přístup pomocí zobecněných lineárních modelů [5].

Mackův model jsme popsali v minulé kapitole. Nyní se budeme zabývat zobecněnými lineárními modely. Jak naznačuje sám název, zobecněný lineární model je zobecněním klasického lineárního modelu, který uplatňujeme při analýze dat. Mnohá data ale vykazují odchylky od standardních předpokladů klasického lineárního modelu a v těchto situacích odhady parametrů získané klasickou metodou nejmenších čtverců ztrácejí své optimální vlastnosti. V posledních letech nastal rozvoj statistických metod, které umožňují upravovat příslušné odhady a tím získat odhady, které si i při porušení základních předpokladů lineárního modelu ponechávají dobré statistické vlastnosti. Teorie lineárního modelu byla zobecněna pro případ, že rozdělení náhodných chyb v modelu není normální a střední hodnota vysvětlované proměnné není identickou funkcí lineárního prediktoru. Tyto modely se nazývají zobecněné lineární modely.

2.1 Teorie zobecněných lineárních modelů

Než přejdeme k samotnému modelu, zavedeme důležité pojmy a vztahy, se kterými budeme dále pracovat. Vycházíme z [2] a [4].

Označení dat a předpoklady

Stejně jako v lineárních modelech, tak i v zobecněných lineárních modelech vystupují pozorování závisle proměnné, což je sloupcový vektor náhodných veličin, tedy:

$$Y^T = (Y_1, \dots, Y_m).$$

Tyto proměnné nazýváme vysvětlovanými proměnnými.

Dále se v modelu vyskytují nezávisle proměnné, které budeme nazývat vysvětlující veličiny (prediktory):

$$x_i^T = (X_{i,1}, \ldots, X_{i,n})$$

pro $i = 1, \ldots, m$.

Můžeme je zapsat pomocí maticového zápisu:

$$\begin{pmatrix} X_{1,1} & \ldots, & X_{1,n} \\ \vdots & & \vdots \\ X_{m,1} & \ldots, & X_{m,n} \end{pmatrix},$$

kde předpokládáme, že matice má plnou sloupcovou hodnost. Tato podmínka vyžaduje, aby mezi vysvětlujícími proměnnými nebyla funkční lineární závislost, tedy v matici X nesmí existovat lineárně závislé sloupce. Počet vysvětlujících proměnných nesmí být pochopitelně větší než počet pozorování a v praxi by měl být počet pozorování výrazně větší než počet vysvětlujících proměnných.

Vektor neznámých parametrů je:

$$\beta^T = (\beta_1, \ldots, \beta_n).$$

Na naše data klademe předpoklad, aby pozorování Y_i byla nezávislá a prediktory x_i^T byly měřené konstanty.

Rodina exponenciálních rozdělení

Definice 3. *Rozdělení s hustotou tvaru*

$$f(y; \theta_i, \phi) = \exp\left\{\frac{y\theta_i - b(\theta_i)}{a(\phi)} + c(y, \phi)\right\} \tag{2.1}$$

nazýváme rozdělení exponenciálního typu s disperzním parametrem ϕ, kde $\theta_i \in R$ nazýváme kanonický parametr, $a(\phi) \in (0, \infty)$ a $a : R_+ \to R_+$, b, c jsou známé funkce. Po funkci b požadujeme, aby byla ryze monotónní, dvakrát spojitě diferencovatelná a s kladnou druhou derivací.

V literatuře se běžně objevují různé parametrizace, pro známé funkce a, b, c (jejich tvar je dán konkrétním rozdělením z exponenciální rodiny) a neznámé parametry θ, φ, kromě vztahu (2.1) např.:

$$f(y; \theta_i, \varphi) = \exp\left\{\frac{y\theta_i - b(\theta_i)}{\varphi} + c(y, \varphi)\right\}. \tag{2.2}$$

Věta 7. *Pro náhodnou veličinu Y patřící do rodiny exponenciálních rozdělení s hustotou*

$$f(y;\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\}$$

platí: Je-li b dvakrát spojitě diferencovatelná, potom

$$E[Y] = b'(\theta), \tag{2.3}$$

$$var(Y) = a(\phi)b''(\theta). \tag{2.4}$$

Než přejdeme k samotnému důkazu, uvedeme pomocné lemma.

Lemma 3. *Nechť Y je reálná náhodná veličina s hustotou $f(y,\theta)$. Označme*

$$l(y,\theta) = \log f(y,\theta).$$

Pak platí

$$E\left(\frac{\partial l}{\partial \theta}\right) = 0 \tag{2.5}$$

a

$$E\left(\frac{\partial^2 l}{\partial^2 \theta}\right) + \left(E\left(\frac{\partial l}{\partial \theta}\right)\right)^2 = 0. \tag{2.6}$$

Nyní dokážeme větu 7.

Důkaz. Vyjdeme z logaritmu hustoty rozdělení exponenciálního typu

$$\log f(y;\theta,\phi) = \frac{y\theta - b(\theta)}{a(\phi)} + \log(c(y,\phi)).$$

Pro první parciální derivaci platí

$$\frac{\partial \log f(y;\theta,\phi)}{\partial \theta} = \frac{y - b'(\theta)}{a(\phi)}.$$

Z rovnice (2.5) plyne

$$\frac{EY - b'(\theta)}{a(\phi)} = 0,$$

tedy

$$EY = b'(\theta).$$

Nyní dokážeme vztah pro rozptyl. Pro druhou parciální derivaci platí

$$\frac{\partial^2 \log f(y;\theta,\phi)}{\partial \theta^2} = -\frac{b''(\theta)}{a(\phi)}.$$

Podobně ze vztahu (2.6) máme

$$\begin{aligned} 0 &= \frac{E[(Y - b'(\theta))^2]}{(a(\phi))^2} - \frac{b''(\theta)}{a(\phi)} \\ &= \frac{var(Y)}{(a(\phi))^2} - \frac{b''(\theta)}{a(\phi)}. \end{aligned}$$

Úpravou dostaneme

$$var(Y) = a(\phi)b''(\theta),$$

a tím je důkaz dokončen.

□

Stavební elementy zobecněných lineárních modelů

První komponenta modelu je reprezentována nezávislými náhodnými veličinami $Y_1, \ldots, Y_m$, o kterých předpokládáme, že jejich rozdělení je exponenciálního typu. To znamená, že mají hustotu tvaru

$$f(Y_i; \theta_i, \varphi) = \exp\left\{ \frac{Y_i\theta_i - b(\theta_i)}{\varphi} + c(Y_i, \varphi) \right\}, \tag{2.7}$$

pro známé funkce b, c a neznámé parametry θ_i, φ, kde θ_i nazýváme kanonický parametr a φ disperzní parametr. Po funkci b požadujeme, aby byla ryze monotónní, dvakrát spojitě diferencovatelná a s kladnou druhou derivací.

Druhou složku, kterou nazýváme systematická složka, tvoří sloupcový vektor

$$\eta^T = \{\eta_1, \ldots, \eta_n\},$$

jenž je lineární funkcí vysvětlujících proměnných reprezentovanými maticí X,

$$\eta_i = \sum_{j=1}^{n} x_{i,j}\beta_j = X\beta, \tag{2.8}$$

kde β_j jsou neznámé parametry a $x_{i,j}$ jsou známé hodnoty prediktorů. Tuto funkci nazýváme lineární prediktor.

Dále předpokládáme, že systematická složka η je propojena s lineárním prediktorem takzvanou linkovou (spojovací) funkcí, která je diferencovatelná a ryze monotónní a platí tedy vztah

$$g(\mu_i) = g(EY_i) = \eta_i, \tag{2.9}$$

neboli

$$\mu_i = g^{-1}(\eta_i). \tag{2.10}$$

Poznamenejme, že v zobecněných lineárních modelech se nevyskytuje chybový člen jako u lineární regrese. Důvodem je, že levá strana předchozí rovnice je funkcí střední hodnoty veličiny Y_i, nikoliv samotné veličiny Y_i.

Dále poznamenejme, že vztah pro rozptyl můžeme přepsat pomocí rozptylové funkce $V(\mu) = b''[(b')^{-1}(\mu)]$ jako

$$var(Y) = a(\phi)V(\mu) = \varphi V(\mu).$$

Odhad parametrů

Máme zvolený model. Nyní nám zbývá odhadnout vektor parametrů β. Ten je nejčastěji odhadován metodou maximální věrohodnosti. Nyní naznačíme postup odhadování parametrů β.

Nechť Y_i jsou nezávislé náhodné veličiny s hustotou exponenciálního typu $f(Y_i; \theta_i, \varphi)$, jejichž hustota závisí na prediktorech skrze vztah

$$\theta_i = (b')^{-1}(g^{-1}({x'}_i\beta)).$$

Věrohodnostní funkce má tvar

$$L(Y; \beta, \varphi) = \prod_{i=1}^{m} f(Y_i; \theta_i, \varphi)$$

a logaritmická věrohodnostní funkce má tvar

$$\begin{aligned} l(Y; \beta, \varphi) &= \sum_{i=1}^{m} \log(f(Y_i; \theta_i, \varphi)) \\ &= \sum_{i=1}^{m} \frac{Y_i\theta_i - b(\theta_i)}{\varphi} + c(Y_i, \varphi). \end{aligned} \tag{2.11}$$

Odhad β_j hledáme maximalizací $l(Y; \beta, \varphi)$. Derivaci logaritmické věrohodnostní funkce (2.11) podle β_j získáme pomocí řetízkového pravidla

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{m} \frac{\partial \log f}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \sum_{i=1}^{m} \frac{(Y_i - \mu_i)X_{i,j}}{g'(\mu_i)\varphi V(\mu_i)}, \tag{2.12}$$

kde

$$\begin{aligned} \frac{\partial \log f}{\partial \theta_i} &= \frac{Y_i - b'(\theta_i)}{\varphi} = \frac{Y_i - \mu_i}{\varphi}, \\ \frac{\partial \theta_i}{\partial \mu_i} &= \frac{1}{b''(\theta_i)} = \frac{1}{V(\mu_i)}, \\ \frac{\partial \mu_i}{\partial \eta_i} &= \frac{1}{g'(\mu_i)}, \\ \frac{\partial \eta_i}{\partial \beta_j} &= X_{i,j}. \end{aligned}$$

Maximum hledáme tak, že všechny derivace (2.12) podle β_j položíme rovny nule, tj.

$$\frac{\partial l}{\partial \beta_j} = 0$$

pro $j = 0, \ldots, n$.

Poznamenejme, že pro získání odhadu β_j se často využívá následujícího tvaru parciální derivace

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{m} \frac{\partial \log f}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j} = \sum_{i=1}^{m} \frac{(Y_i - \mu_i)}{\varphi V(\mu_i)} \left(\frac{\partial \mu_i}{\partial \beta_j} \right).$$

Parametry β_j můžeme kromě metody maximální věrohodnosti odhadnout i pomocí jiných metod. Dalšími možnostmi jsou Newton-Raphsonův algoritmus nebo metoda iterativních vážených čtverců.

Odhad disperzního parametru

Odhad parametru ϕ lze provést například pomocí zobecněné Pearsonovy chí-kvadrát statistiky. Vzorec pro odhad disperzního parametru je dán vztahem

$$\hat{\phi} = \frac{1}{m-n} \sum_{i=1}^{m} \frac{(Y_i \; - \; \hat{\mu}_i)}{V(\hat{\mu}_i)},$$

kde m je počet pozorování a n je počet parametrů.

Poznamenejme, že pro odhad disperzního parametru můžeme uvažovat jiné vztahy - viz níže (3.9) a (3.10).

2.2 Zobecněný lineární model v pojišťovnictví

V této části popíšeme jednotlivé zobecněné lineární modely, se kterými se můžeme setkat v pojišťovnictví. Jedná se o logaritmicko-normální model, zobecněný Poissonův model a Gamma model. Nakonec se dostaneme k analytickému výpočtu odhadu variability rezerv. Původní text může čtenář nalézt v [5] a [18].

V minulé kapitole jsme při definování Mackova modelu vycházeli z celkového úhrnu pojistného plnění $C_{i,j}$ za škody, které vznikly v roce i a byly uhrazeny do konce vývojového roku j. Tyto veličiny jsou závislé. Pro výpočet rezerv u zobecněných lineárních modelů potřebujeme nezávislé náhodné veličiny. Proto uvažujme nekumulativní hodnoty $D_{i,j}$, které představují škody z roku i uhrazené právě v roce $i + j$. Pro tyto škody platí následující vztahy:

$$D_{i,j} = C_{i,j} - C_{i,j-1}, \tag{2.15}$$

$$C_{i,j} = \sum_{k=0}^{j} D_{i,k}. \tag{2.16}$$

Logaritmicko-normální model

Nechť $D_{i,j} > 0$, $i = 0, 1, \ldots, J$, $j = 0, 1, \ldots, J$ označuje škody z roku i, které byly uhrazeny právě v roce $i + j$. Označme $Y_{i,j} = \log(D_{i,j})$ a uvažujme model $Y_{i,j} = m_{i,j} + \varepsilon_{i,j}$.

Předpokládáme, že normální odezvy $Y_{i,j}$ se aditivně rozloží mezi deterministicky nenáhodnou složku se střední hodnotou $m_{i,j} = \eta_{i,j}$ a homoskedastickou normálně rozdělenou náhodnou chybu s nulovou střední hodnotou $\varepsilon_{i,j}$. Použití logaritmické transformace v modelu nám dává omezení, že jednotlivé škodní částky musí být kladné.

Formálně:

$$Y_{i,j} \sim LN(m_{i,j}, \sigma^2) \tag{2.17}$$

$$\varepsilon_{i,j} \sim LN(0, \sigma^2) \tag{2.18}$$

$$m_{i,j} = \eta_{i,j} = \gamma + \alpha_i + \beta_j, \ \alpha_1 = \beta_1 = 0 \tag{2.19}$$

Rovnice (2.17) – (2.19) definují model zavedený E. Kremerem (1982). Škodní rok a vývojový rok jsou považovány za faktory, kde parametr α_i odpovídá jednotlivým škodním rokům i a parametr β_j odpovídá jednotlivým vývojovým rokům j. Parametry v lineárním prediktoru $\eta_{i,j}$ jsou odhadovány metodou maximální věrohodnosti, která v případě logaritmicko-normálního rozdělení chyb $\varepsilon_{i,j}$ odpovídá minimalizaci reziduálního součtu čtverců. Neznámý rozptyl σ^2 je odhadován reziduálním součtem čtverců děleným počtem stupňů volnosti (rozdíl počtu pozorovaných dat a počtu odhadovaných parametrů).

Zobecněný Poissonův model

Pro Poissonův model obecně platí, že střední hodnota je shodná s rozptylem. Porušením této rovnosti dosáváme zobecněný lineární model. A. E. Renshaw a R. J. Verrall (1994) navrhli modelování nekumulativních škod $D_{i,j}$ přímo jako odezvy se stejným lineárním prediktorem jako E. Kremer, ale použili logaritmickou linkovou funkci k popisu průměru a zobecněné Poissonovo rozdělení chyb.

Formálně:

$$E[D_{i,j}] = m_{i,j} \tag{2.20}$$

$$var[D_{i,j}] = \phi E[D_{i,j}] = \phi m_{i,j} \tag{2.21}$$

$$\log(m_{i,j}) = \eta_{i,j} = \gamma + \alpha_i + \beta_j, \ \alpha_1 = \beta_1 = 0 \tag{2.22}$$

Rovnice (2.20) – (2.22) definují zobecněný lineární model, ve kterém jsou odezvy modelovány logaritmickou linkovou funkcí a rozptyl je přímo úměrný průměru. Parametr ϕ je neznámý disperzní parametr.

Renshaw a Verrall nebyli prvními, kteří si všimli souvislosti mezi metodou Chain-Ladder a Poissonovým rozdělením, ale byli prvními, kteří zavedli model používající standardní postupy ve statistickém modelování a poskytli propojení s analýzou kontingenčních tabulek.

Vzhledem ke způsobu, jakým je struktura modelu parametrizována, je nutné zavést omezení, že součet jednotlivých škod v každém řádku a každém sloupci vývojového trojúhelníku musí být kladný. Kromě tohoto omezení musíme kvůli logaritmické linkové funkci uvažovat pouze kladné hodnoty předpovídaných hodnot. To ovšem obvykle vede k nepoužitelnosti modelu u vzniklých škod, které často obsahují přecenění rezerv v raných fázích vývoje vedoucích k sérii záporných škod v pozdější fázi vývoje.

Gamma model

T. Mack (1991) navrhl další stochastický model. V tomto modelu je struktura parametrů navržena pro střední hodnotu jednotlivých kumulativních škod, které jsou modelovány jako proměnné s Gamma rozdělením a k získání odhadů parametrů používá metodu maximální věrohodnosti. Renshaw a Verrall poznamenali, že stejný model získáme použitím zobecněného lineárního modelu, ve kterém budou jednotlivé škodní částky modelovány jako nezávislé proměnné s Gamma rozdělením, s logaritmickou linkovou funkcí a stejným lineárním prediktorem, který použil Kremer.

Formálně:

$$E[D_{i,j}] = m_{i,j} \tag{2.23}$$

$$var[D_{i,j}] = \phi E[D_{i,j}]^2 = \phi m_{i,j}^2 \tag{2.24}$$

$$\log(m_{i,j}) = \eta_{i,j} = \gamma + \alpha_i + \beta_j,\ \alpha_1 = \beta_1 = 0 \tag{2.25}$$

Jediný rozdíl mezi tímto modelem a modelem navrženým Renshawem a Verrallem je, že rozptyl celkového úhrnu pojistného plnění je přímo úměrný druhé mocnině střední hodnoty.

Analytické odhady chyb předpovědí

Jednou z výhod stochastických modelů je dostupnost odhadů variability rezerv. Běžným problémem v předpovídání je standardní chyba předpovědi, známá také jako chyba předpovědi, nebo střední kvadratická chyba předpovědi. Uvažujme škodní rok i a výplaty škod ve vývojovém roce j. Střední kvadratická chyba předpovědi je dána vztahem

$$E\left[(D_{i,j} - \hat{D}_{i,j})^2\right] = var[D_{i,j}] + var[\hat{D}_{i,j}]. \tag{2.26}$$

Rovnice (2.26) platí pro logaritmicko-normální model, zobecněný Poissonův model a Gamma model. Poznamenejme, že střední kvadratická chyba předpovědí může být považována za součet dvou členů, rozptylu procesu a chyby odhadu (viz vztah (1.10)).

Obecný vztah pro proces rozptylu můžeme odvodit pro zobecněný Poissonův a Gamma model. Z rovnic (2.21) a (2.24) vidíme

$$var[D_{i,j}] = \phi m_{i,j}^{\rho},$$

kde $\rho = 1$ pro zobecněný Poissonův model a $\rho = 2$ pro Gamma model.

Pro odhad máme

$$\hat{D}_{i,j} = \hat{m}_{i,j} = \exp(\hat{\eta}_{i,j}).$$

Užitím Taylorova rozvoje

$$var[\hat{D}_{i,j}] = \left[\frac{\partial m_{i,j}}{\partial \eta_{i,j}}\right]^2 var[\eta_{i,j}]$$

dostáváme vztah pro střední kvadratickou chybu předpovědi

$$E\left[(D_{i,j} - \hat{D}_{i,j})^2\right] = \phi m_{i,j}^{\rho} + m_{i,j}^2 var[\eta_{i,j}]. \tag{2.27}$$

Také můžeme spočítat standardní chybu předpovědi odhadů škod pro daný rok a odhad celkové rezervy. Označme trojúhelník předpovídaných škodních nároků jako $\hat{\Delta} = \left\{\hat{D}_{i,j}; i + j \geq J\right\}$. Odhady rezerv v daném roce i získáme součtem předpovídaných hodnot v daném řádku trojúhelníku předpovídaných škodních nároků $\hat{\Delta}$, tj.

$$\hat{R}_i = \sum_{j=J-i}^{J-1} \hat{D}_{i,j}.$$

Střední kvadratická chyba předpovědi rezervy pro daný rok je přibližně dána vztahem

$$\begin{aligned} E\left[(\hat{R}_i)^2\right] \cong & \sum_{j\in\hat{\Delta}_i} \phi m_{i,j}^{\rho} + \sum_{j\in\hat{\Delta}_i} m_{i,j}^2 var[\eta_{i,j}] \\ & +2 \sum_{j_1,j_2\in\hat{\Delta}_i, j_1\geq j_2} m_{i,j_1} m_{i,j_2} cov\left[\eta_{i,j_1}\eta_{i,j_2}\right]. \end{aligned} \tag{2.28}$$

Odhad celkové rezervy spočítáme pomocí vzorce

$$\hat{R} = \sum_{i=0}^{J-1} \sum_{j=J-i}^{J-1} D_{i,j}$$

a střední kvadratickou chybu předpovědi celkové rezervy vypočítáme ze vztahu

$$\begin{aligned} E\left[(\hat{R})^2\right] \cong & \sum_{i,j\in\hat{\Delta}} \phi m_{i,j}^{\rho} + \sum_{i,j\in\hat{\Delta}} m_{i,j}^2 var[\eta_{i,j}] \\ & +2 \sum_{i_1,i_2,j_1,j_2\in\hat{\Delta},\{i_1,j_1\}\neq\{i_2,j_2\}} m_{i_1,j_1} m_{i_2,j_2} cov\left[\eta_{i_1,j_1}\eta_{i_2,j_2}\right]. \end{aligned} \tag{2.29}$$

Rovnice (2.28) a (2.29) vyžadují opatrnost při sčítání jednotlivých členů. Výrazy obsahující kovarianci nejsou snadno dostupné z balíčků statistických softwarů. Za předpokladu, že máme k dispozici statistický software, můžeme vypočítat kovarianční matici. Rozptyly lineárních prediktorů jsou prvky na diagonále této matice, mimo diagonálu leží kovariance lineárních prediktorů. Připomeňme vztahy pro výpočet těchto rozptylů a kovariancí.

Rozptyl lineárního prediktoru vypočítáme pomocí vztahu

$$\begin{aligned} var\left(\eta_{i,j}\right) &= var\left(\gamma + \alpha_i + \beta_j\right) \\ &= var\left(\alpha_i\right) + var\left(\beta_j\right) + 2cov\left(\alpha_i\beta_j\right) \end{aligned}$$

pro $i = 0, \dots, J-1$, $j = 0, \dots, J-1$.

Kovarianci lineárních prediktorů spočteme na základě vztahu

$$cov\left(\eta_{i_1,j_1}\eta_{i_2,j_2}\right) = cov\left(\alpha_{i_1}\alpha_{i_2}\right) + cov\left(\alpha_{i_1}\beta_{j_2}\right) + cov\left(\alpha_{i_2}\beta_{j_1}\right) + cov\left(\beta_{j_1}\beta_{j_2}\right),$$

pro $i_1, i_2 = 0, \dots, J-1$, $i_1 \neq i_2$, $j_1, j_2 = 0, \dots, J-1$, $j_1 \neq j_2$.

3. Metoda bootstrap

3.1 Základní principy metody bootstrap

V této části popíšeme základní myšlenku metody bootstrap, definujeme elementární pojmy a jednoduchým způsobem nastíníme algoritmus metody. Vycházíme z článku [17].

Představme si následující situaci: Uvažujme nezávislé stejně rozdělené (*iid*) náhodné veličiny $X_1, \ldots, X_n$, které mají rozdělení s distribuční funkcí F. Tato distribuční funkce není blíže specifikována. Nechť vektor $x = (x_1, \ldots, x_n)$ je realizací náhodných veličin $X_1, \ldots, X_n$. Nechť $\theta = \theta(F)$ je nějaká charakteristika rozdělení - může to být neznámý parametr, který má být odhadnut na základě realizace náhodného výběru x.

Nechť $T_n = T_n(X_1, \ldots, X_n)$ je statistika pro odhad parametru θ, nechť $R_n = R_n(X_1, \ldots, X_n; F)$ je její vhodně standardizovaná verze, např. $R_n = \sqrt{n}(T_n - \theta)$, nebo nějaká její funkce. Nechť

$$H_n(x) = P\left[R_n(X_1, \ldots, X_n, F) \leq x\right]$$

značí distribuční funkci statistiky R_n.

Naším cílem je získat rozdělení H_n. Bohužel explicitní odvození rozdělení H_n i výpočet číselných charakteristik mohou být v jednotlivých případech značně obtížné, či dokonce analyticky neproveditelné. Obecně mohou nastat dvě možnosti:

1. Distribuční funkci F známe, ale H_n je tak složitá funkce proměnných $X_1, \ldots, X_n$, že ji nejsme schopni analyticky najít.
2. Rozdělení náhodných veličin $X_1, \ldots, X_n$ s danou distribuční funkcí F neznáme a nemáme o něm žádné informace.

V prvním případě, při známé distribuční funkci F, můžeme pravděpodobnostní rozdělení H_n odhadnout bez složitých analytických výpočtů a postupovat metodou Monte Carlo: z rozdělení s danou distribuční funkcí vygenerujeme dlouhou sérii nezávislých náhodných výběrů velikosti n. Pro každé opakování spočteme hodnotu příslušné charakteristiky a její skutečné rozdělení aproximujeme empirickým rozdělením získaným z řady takto uměle získaných hodnot.

V druhé situaci, kdy je skutečná distribuční funkce F neznámá, je možné aproximovat H_n asymptotickým rozdělením, které lze odvodit na základě limitních vět teorie pravděpodobnosti. Základní idea je popsána na následujících řádcích.

Metoda bootstrap nabízí řešení, které kombinuje tzv. substituční princip a metodu Monte Carlo. Vysvětleme nejdříve substituční princip.

Nechť $F_n(x)$ je nějaký odhad distribuční funkce. Nejčastěji se uvažuje empirická distribuční funkce založená na náhodném výběru $X_1, \ldots, X_n$, tj.

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}[X_i \leq x],$$

kde $\mathbf{I}[A]$ značí indikátor množiny A. Při daných hodnotách $X_1, \ldots, X_n$ je F_n známá funkce.

Nechť $X_1^*, \ldots, X_n^*$ je nezávislý náhodný výběr z F_n, tj. při daných pozorováních $X_1, \ldots, X_n$ jsou $X_1^*, \ldots, X_n^*$ nezávislé, stejně rozdělené náhodné veličiny, z nichž každá nabývá hodnot $X_1, \ldots, X_n$ s pravděpodobností $\frac{1}{n}$. Soubor $X_1^*, \ldots, X_n^*$ nazýváme bootstrapový výběr. Jinak řečeno, bootstrapový výběr je náhodný výběr s vracením z množiny $\{x_1, \ldots, x_n\}$. V dalších úvahách původní výběr nahradíme bootstrapovým výběrem a neznámou distribuční funkci F známou distribuční funkcí F_n. Dostaneme parametr $\theta^* = \theta(F_n)$ a statistiky $T_n^* = T_n(X_1^*, \ldots, X_n^*)$ a $R_n^* = R_n(X_1^*, \ldots, X_n^*, F_n)$.

Nechť P^* značí pravděpodobnostní míru indukovanou bootstrapem, tj.

$$P^*\left(R_n(X_1^*, \ldots, X_n^*, F_n) \leq x\right) = P\left(R_n(X_1^*, \ldots, X_n^*, F_n) \leq x | X_1, \ldots, X_n\right).$$

Nyní můžeme definovat charakteristiky jako střední hodnotu

$$E^*T_n^* = \int_{R^n} T_n(x_1, \ldots, x_n) d\left(F_n(x_1), \ldots, F_n(x_n)\right),$$

rozptyl

$$var^*T_n^* = \int_{R^n} \left[T_n(x_1, \ldots, x_n) - E^*T_n^*\right]^2 d\left(F_n(x_1), \ldots, F_n(x_n)\right)$$

a distribuční funkci

$$\begin{aligned} H_n^*(x) &= P^*\left(R_n(X_1^*, \ldots, X_n^*, F_n) \leq x\right) \\ &= P\left(R_n(X_1^*, \ldots, X_n^*, F_n) \leq x | X_1, \ldots, X_n\right); \end{aligned}$$

jsou to tzv. teoretické charakteristiky a teoretická distribuční funkce získané metodou bootstrap.

Pro praktické použití jsou teoretické bootstrapové charakteristiky vhodné jen v případě, že jsou explicitními funkcemi pozorování $X_1, \ldots, X_n$. Přesné stanovení bootstrapového rozdělení by vyžadovalo provedení všech n^n možných výběrů s vracením z populace pozorovaných hodnot $X_1, \ldots, X_n$. To je však uskutečnitelné jen pro výběry o malém rozsahu. Nejčastěji se proto na bootstrapový výběr $X_1^*, \ldots, X_n^*$ a známou distribuční funkci F_n aplikuje metoda Monte Carlo, kdy se mnohokrát ($B-$krát) generuje nezávislý náhodný výběr z rozdělení F_n, při každém opakování se spočtou hodnoty T_n^*, R_n^* a z nich se stanoví aritmetický průměr. Dostaneme tak bootstrapové odhady původního rozdělení a původních charakteristik.

Např. bootstrapový odhad rozptylu T_n dostaneme tak, že opakujeme nezávislý náhodný výběr z rozdělení F_n celkem $B-$krát a spočteme vždy hodnotu statistiky T_n^*. Dostáváme tak hodnoty $T_{n,1}^*, \ldots, T_{n,B}^*$, ze kterých spočteme

$$\widehat{var}^* T_n^* = \frac{1}{B}\sum_{b=1}^{B}\left(T_{n,b}^* - \frac{1}{B}\sum_{k=1}^{B} T_{n,k}^*\right)^2.$$

Podobně odhadneme distribuční funkci statistiky R_n jako

$$\widehat{H_n^*}(x) = \frac{1}{B}\sum_{b=1}^{B}\mathbf{I}\left\{R_n\left(X_{1,b}^*, \ldots, X_{n,b}^*, F_n\right) \leq x\right\},$$

kde $X_{1,b}^*, \ldots, X_{n,b}^*$, $b = 1, \ldots, B$ jsou nezávislé výběry z F_n.

V následující části ukážeme použití metody bootstrap v praxi. Původní text může čtenář nalézt v [5] a [7]. Dále čerpáme z přednášek [14] a [15].

3.2 Bootstrap v praxi

V minulých kapitolách jsme si ukázali metody, podle kterých můžeme analyticky spočítat standardní chybu. Pokud je těžké nebo dokonce nemožné odhadnout standardní chybu analyticky použijeme metodu bootstrap. U škodních rezerv se zajímáme o chybu předpovědi součtu náhodných veličin a bootstrap je vhodnou metodou pro její získání. V regresních modelech je běžné bootstrapovat rezidua místo bootstrapování samotných dat. U modelů lineární regrese s normálně rozdělenými chybami jsou rezidua dána rozdílem pozorované a odhadnuté hodnoty. Pro zobecněné lineární modely požadujeme rozšířenou definici reziduí. Možnými rezidui jsou Pearsonova rezidua, rezidua založená na devianci a Anscombe rezidua.

Při definování těchto reziduí budeme vycházet z již dříve používaného značení z předchozí kapitoly. Nechť $D_{i,j}$ jsou pozorované nezávislé stejně rozdělené náhodné veličiny a nechť $\hat{D}_{i,j}$ jsou odhadnuté hodnoty těchto veličin. Dále nechť b je funkce z hustoty u zobecněných lineárních modelů - ryze monotónní, dvakrát spojitě diferencovatelná s kladnou druhou derivací. Nechť dále platí $\hat{\theta}_i^* = (b')^{-1}(D_{i,j})$. Pak můžeme zavést následující definice.

Definice 4. *Normované Pearsonovo reziduum je definováno vztahem*

$$_{(P)}r_{i,j} = \frac{D_{i,j} - \hat{D}_{i,j}}{\sqrt{var\hat{D}_{i,j}}}. \tag{3.1}$$

Definice 5. *Anscombe reziduum je definováno vztahem*

$$_{(A)}r_{i,j} = \frac{A(D_{i,j}) - A(\hat{D}_{i,j})}{A'(\hat{D}_{i,j})\sqrt{var\hat{D}_{i,j}}}, \tag{3.2}$$

kde

$$A(\mu) = \int \frac{1}{var(\mu)^{1/3}} d\mu$$

a $var(\mu)$ je varianční funkce vysvětlované náhodné veličiny.

Definice 6. *Reziduum založené na devianci je definováno vztahem*

$$_{(D)}r_{i,j} = sign(D_{i,j} - \hat{D}_{i,j})\sqrt{d_i}, \tag{3.3}$$

kde

$$d_i = 2\left(D_{i,j}\left(\hat{\theta}_i^* - \hat{\theta}_i\right) - b\left(\hat{\theta}_i^*\right) + b\left(\hat{\theta}_i\right)\right).$$

Přesná forma definice reziduí je dána rozdělením chyb. Pro zobecněný Poissonův model definovaný rovnicemi (2.20) – (2.22) používáme rezidua [15]:

$$_{(P)}r_{i,j} = \frac{D_{i,j} - \hat{D}_{i,j}}{\sqrt{\hat{D}_{i,j}}} \tag{3.4}$$

$$_{(A)}r_{i,j} = \frac{3}{2}\frac{D_{i,j}^{2/3} - \hat{D}_{i,j}^{2/3}}{\hat{D}_{i,j}^{1/6}} \tag{3.5}$$

$$_{(D)}r_{i,j} = sign(D_{i,j} - \hat{D}_{i,j})\sqrt{2\left(D_{i,j}\log\left(\frac{D_{i,j}}{\hat{D}_{i,j}}\right) - D_{i,j} + \hat{D}_{i,j}\right)} \tag{3.6}$$

Pro Gamma model definovaný rovnicemi (2.23) – (2.25) používáme rezidua [15]:

$$_{(P)}r_{i,j} = \frac{D_{i,j} - \hat{D}_{i,j}}{\hat{D}_{i,j}} \tag{3.7}$$

$$_{(A)}r_{i,j} = 3\frac{D_{i,j}^{2/3} - \hat{D}_{i,j}^{2/3}}{\hat{D}_{i,j}^{1/3}} \tag{3.8}$$

Bootstrap proces zahrnuje výběr reziduí s vracením. Takto dostaneme bootstrapová rezidua $_{(\cdot)}r_{i,j}^*$. Bootstrapový výběr dat poté získáme osamostatněním jednotlivých hodnot $D_{i,j}$ v jednom ze vzorců (3.1) – (3.3), kde rezidua $_{(\cdot)}r_{i,j}$ nahradíme bootstrapovými rezidui $_{(\cdot)}r_{i,j}^*$. Máme-li k dispozici určitá rezidua $_{(\cdot)}r_{i,j}^*$ a hodnoty $\hat{D}_{i,j}$, pak vidíme, že rovnici (3.3) neumíme analyticky vyřešit pro jednotlivé škody $D_{i,j}$. Tedy v tomto případě není vhodné bootstrapovat rezidua, ale naopak je snadnější vyřešit rovnici (3.1) pro $D_{i,j}$. Je také možné získat hodnoty $D_{i,j}$ ze vztahu (3.2) pro Anscombe rezidua , ale dále je nebudeme uvažovat, neboť jsou méně běžná a je vhodné použít bootstrapování u reziduí, které obsahují odhad disperzního parametru.

Disperzní parametr je odhadován buď jako statistika deviance dělená počtem stupňů volnosti nebo jako Pearsonova chí-kvadrát statistika dělená počtem stupňů volnosti. Statistika deviance a Pearsonova chí-kvadrát statistika jsou získány jako součet druhých mocnin odpovídajících reziduí. Stupně volnosti jsou definovány jako rozdíl počtu dat v modelu (z původního výběru) a počtu parametrů v odhadovaném modelu. Formálně, disperzní parametr založený na devianci je dán vztahem

$$\phi_D = \frac{1}{J-n} \sum_{i=0}^{J} {}_{(D)}r_{i,j}^2 \tag{3.9}$$

a disperzní Pearsonův parametr je definován

$$\phi_P = \frac{1}{J-n} \sum_{i=0}^{J} {}_{(P)}r_{i,j}^2 \tag{3.10}$$

kde J je počet pozorovaných dat a n je počet odhadovaných parametrů.

Algoritmus bootstrap pro Mackův model

K dispozici máme celkový úhrn pojistného plnění $C_{i,j}$ za škody, které vznikly v roce i a byly uhrazeny (nahlášeny) do konce vývojového roku j, kde $i = 0, 1, \ldots, I$, $j = 0, 1, \ldots, J$, $I = J$. Uvažujme kumulativní trojúhelník $\{C_{i,j}\} = \Delta$, který obsahuje všechna dosud známá pozorování $C_{i,j}$. Nechť dále platí předpoklady Mackovy metody (MCL1), (MCL2) a (MCL3). Vycházíme z článku [7].

1. Z kumulativních dat $C_{i,j}$ vypočítáme odhad vývojových faktorů f_j pomocí vztahu (1.2), kde $j = 0, \ldots, J$.

2. Trojúhelník $\{C_{i,j}\}$, $i + j \geq J + 1$ doplníme na čtverec násobením odhadnutými faktory $\hat{f}_j$ počínaje od diagonály. Pravý dolní trojúhelník odhadujeme pomocí vztahu

$$\hat{C}_{i,j} = C_{i,j-i}\hat{f}_{j-i} \cdot \ldots \cdot \hat{f}_{j-1}, \tag{3.11}$$

   kde $i + j \geq J + 1$.

3. Pomocí rekurze ve škodním trojúhelníku $\{C_{i,j}\}$ zpětně odhadujeme původní škody (kromě diagonály). Tím získáme kumulativní trojúhelník odhadnutých hodnot $\hat{C}_{i,j}$.

$$\hat{C}_{i,J-i} = C_{i,J-i}, \tag{3.12}$$

$$\hat{C}_{i,j} = \frac{C_{i,J-i}}{\hat{f}_{J-i} \cdot \ldots \cdot \hat{f}_j}, \tag{3.13}$$

pro $i + j \leq J$.

4. Rozdílem sousedních hodnot v trojúhelníku $\left\{\hat{C}_{i,j}\right\}$ dostaneme nekumulativní trojúhelník odhadnutých hodnot $\left\{\hat{D}_{i,j}\right\}$. Formálně:

$$D_{i,0} = C_{i,0}, \tag{3.14}$$

$$D_{i,j} = C_{i,j} - C_{i,j-1}, \tag{3.15}$$

pro $i + j \leq J$.

5. Pomocí vztahu (3.4) spočteme normovaná Pearsonova rezidua ${}_{(P)}r_{i,j}$ pro všechna $j = 0, \ldots, J$.

6. Bootstrapový proces zahrnuje výběr reziduí s vracením. Tj. $B-$krát s vracením vybereme normovaná Pearsonova rezidua, která jsme vypočítali v minulém kroku. Takto dostaneme B nových trojúhelníků naplněných bootstrapovými rezidui

$$\{{}_{(P)}r_{i,j}\} \rightarrow \{{}^{(b)}_{(P)}r^*_{i,j}\},$$

pro $i + j \geq J + 1$, $b = 1, \ldots, B$.

7. Ve vzorci (3.4) nahradíme normovaná Pearsonova rezidua ${}_{(P)}r_{i,j}$ bootstrapovými rezidui ${}^{(b)}_{(P)}r^*_{i,j}$. Bootstrapové jednotlivé škodní částky ${}^{(b)}D^*_{i,j}$ získáme osamostatněním členu $D_{i,j}$ ve vzorci (3.4), tj.

$$^{(b)}D^*_{i,j} =^{(b)}_{(P)} r^*_{i,j}\sqrt{\hat{D}_{i,j}} + \hat{D}_{i,j}. \tag{3.16}$$

Výsledkem je B nekumulativních trojúhelníků bootstrapových hodnot $\{{}^{(b)}D^*_{i,j}\}$, $b = 1, \ldots, B$.

8. Součtem hodnot v bootstrapovém trojúhelníku $\{{}^{(b)}D^*_{i,j}\}$ dostaneme kumulativní trojúhelník bootstrapových hodnot $\{{}^{(b)}C^*_{i,j}\}$, kde

$$\begin{aligned} {}^{(b)}C^*_{i,0} &=^{(b)} D^*_{i,0} \\ {}^{(b)}C^*_{i,j} &= \sum_{k=0}^{J-1-i} {}^{(b)}D^*_{i,k} \end{aligned}$$

pro $i = 0, \ldots, J$. Tento postup opakujeme $B-$krát, $b = 1, \ldots, B$.

9. Standardní Mackovu metodu Chain-Ladder aplikujeme na každý kumulativní trojúhelník bootstrapových hodnot $\{{}^{(b)}C^*_{i,j}\}$ a pomocí vztahu (1.2) spočteme bootstrapové vývojové faktory ${}^{(b)}f^*_j$, $b = 1, \ldots, B$.

10. Bootstrapové škodní rezervy ${}^{(b)}R_i^*$ pro jednotlivé roky i vypočítáme pomocí vztahu (1.4), ve kterém hodnoty $C_{i,J-i}$ nahradíme bootstrapovými hodnotami ${}^{(b)}C_{i,J-i}^*$ a vývojové faktory $\hat{f}_j$ nahradíme bootstrapovými vývojovými faktory ${}^{(b)}f_j^*$, $b = 1, \ldots, B$.

11. Dále nahrazením ${}^{(b)}C_{i,J-i}^* \to C_{i,J-i}$ a ${}^{(b)}f_j^* \to \hat{f}_j$ ve vztahu (1.5) pro $\hat{\sigma}_j^2$ získáme bootstrapový odhad $({}^{(b)}\hat{\sigma}_j^2)^*$, $b = 1, \ldots, B$.

12. Jednotlivé rezervy získáme jako aritmetický průměr jednotlivých bootstrapových rezerv, tj.

$$R_i^* = \frac{1}{B} \sum_{b=1}^{B} {}^{(b)}R_i^*.$$

13. Nakonec pomocí vztahu (1.16) spočteme odhad střední kvadratické chyby celkové rezervy $mse(R^*)$, kde $R^* = R_0^* + \ldots + R_J^*$.

Poznamenejme, že disperzní Pearsonův parametr používáme v analytickém odhadu rozptylu a Pearsonova rezidua v bootstrap procesu. Bootstrapový odhad rozptylu je analogický analytickému odhadu rozptylu bez uvažování počtu parametrů (jako kdyby byl disperzní parametr počítán vydělením J místo $J - n$). Chceme-li povolit srovnání odhadů rozptylu mezi postupy zobecněného lineárního modelu a metody bootstrap, je nezbytné provést úpravu bootstrapového odhadu rozptylu a vzít v úvahu počet parametrů použitých ve fitovaném modelu. Vhodná úprava je vynásobit bootstrapový odhad rozptylu výrazem $J/(J - n)$.

Pro získání bootstrapové standardní chyby rezervy je nutné přidat odhad rozptylu procesu, který je dán součinem disperzního parametru a odhadu rezerv. Bootstrapová standardní odchylka rezervy je pak dána vztahem

$$_{(B)}se(R)^* = \sqrt{\phi_P R^* + \frac{J}{J-n} mse(R^*)}, \tag{3.17}$$

kde R^* je celková rezerva, $R^* = R_0^* + \ldots + R_J^*$ a $mse(R^*)$ je bootstrapová střední směrodatná chyba odhadu rezervy.

Pro úplnost shrneme aplikaci metody bootstrap na zobecněné lineární modely. Vyjdeme ze zobecněného Poissonova modelu, viz rovnice (2.20)−(2.22). Čerpáme z přednášky [14].

Algoritmus bootstrap pro zobecněný Poissonův model

1. Pomocí statistického softwaru získáme odhady $\hat{\gamma}$, $\hat{\alpha}_i$, $\hat{\beta}_j$ a $\hat{\phi}$ v zobecněném Poissonově modelu.

2. Nafitujeme očekávané hodnoty škod v nekumulativním trojúhelníku $\{D_{i,j}\}$ pomocí vztahu

$$\hat{D}_{i,j} = \hat{m}_{i,j} = \exp\left\{\hat{\gamma} + \hat{\alpha}_i + \hat{\beta}_j\right\}.$$

3. Vypočítáme Pearsonova rezidua pomocí vztahu

$$_{(P)}r_{i,j} = \frac{D_{i,j} - \hat{D}_{i,j}}{\sqrt{\hat{\phi}\hat{D}_{i,j}}}. \tag{3.18}$$

4. Vybereme rezidua - budeme vybírat rezidua $\left\{_{(P)}r_{i,j}\right\}$ B-krát s vracením. Získáme B nových vývojových trojúhelníků naplněných bootstrapovými rezidui $\left\{^{(b)}_{(P)}r^*_{i,j}\right\}$, kde $b = 1, \ldots, B$.

5. Inverzní postup, zpětně zkonstruujeme B bootstrapových trojúhelníků pomocí vztahu

$$^{(b)}D^*_{i,j} =^{(b)}_{(P)} r^*_{i,j}\sqrt{\hat{\phi}\hat{D}_{i,j}} + \hat{D}_{i,j},$$

kde $i + j \leq J$, $b = 1, \ldots, B$.

6. Aplikujeme zobecněnou Poissonovu metodu na každý bootstrapový trojúhelník a získáme bootstrapové odhady $^{(b)}\hat{\gamma}$, $^{(b)}\hat{\alpha}_i$, $^{(b)}\hat{\beta}_j$ a $^{(b)}\hat{\phi}$, $b = 1, \ldots, B$.

7. Vypočítáme bootstrapové rezervy

$$^{(b)}\hat{R}^*_i = \sum_{j=J+2-i}^{J} \hat{D}^*_{i,j} = \exp\left\{^{(b)}\hat{\gamma} +^{(b)} \hat{\alpha}_i\right\} \sum_{j=J+2-i}^{J} \exp\left\{^{(b)}\hat{\beta}_j\right\}. \tag{3.19}$$

pro $b = 1, \ldots, B$.

8. Nakonec vypočítáme střední kvadratickou chybu předpovědi celkové rezervy podle vztahu (2.29). Proces opakujeme B-krát. Výsledná bootstrapová střední kvadratická chyba předpovědi je dána jako aritmetický průměr jednotlivých chyb pro $b = 1, \ldots, B$.

Poznamenejme nakonec, že ve výpočtech metody bootstrap nemáme k dispozici zavedení tail faktoru. Není totiž zřejmé, jak bychom měli vzít v úvahu nejistotu v předpovídaných hodnotách mimo rozsah sledovaných údajů.

4. Praktická část

V předchozích kapitolách jsme se zabývali popisem stochastických metod pro odhad celkové rezervy na pojistná plnění a výpočet její střední kvadratické chyby. V této kapitole se budeme věnovat praktické aplikaci zmíněných modelů na reálná a nasimulovaná data.

Cílem praktické části je porovnat čtyři představené modely (Mackův model Chain-Ladder, zobecněný Poissonův model, Gamma model a metodu bootstrap aplikovanou na Chain-Ladder).

K modelování budeme využívat matematický software Mathematica 8.0.

4.1 Reálná data

Přestože v České republice působí mnoho pojišťoven zabývajících se neživotním pojištěním, je velmi problematické reálná data získat. Pojišťovny totiž nemají povinnost tato data zveřejňovat. Pro aplikaci modelů použijeme data z tabulky 4.1, která jsme získali z volně přístupné studie PartnerRe Ltd. 2011 Loss Development Triangles [13]. Tato data vyjadřují kumulované hrubé hodnoty nahlášených pojistných plnění za pojištění motorových vozidel v Severní Americe mezi lety 2002 až 2011. Hrubými hodnotami označujeme škody, které nebyly upraveny o podíl zajišťoven a také neuvažujeme jiné vlivy. Všechna data byla převedena na americké dolary s použitím směnných kurzů k 31. prosinci 2011. Škody uvádíme v tisících. Předpokládáme, že vývoj škod je po deseti letech ukončen.

i/j	0	1	2	3	4	5	6	7	8	9
0	18 825	78 917	99 690	106 762	109 280	111 157	112 093	112 788	113 280	113 488
1	21 019	84 679	105 686	113 190	117 600	119 251	120 013	120 695	121 176	
2	25 211	80 403	104 706	115 509	120 995	122 132	125 362	125 854		
3	19 413	53 887	67 005	71 947	75 166	75 837	76 574			
4	18 897	54 612	62 441	69 738	74 647	77 505				
5	13 613	36 725	48 750	54 415	55 751					
6	18 712	46 519	57 998	66 714						
7	25 328	70 545	89 597							
8	16 011	51 211								
9	14 433									

Tabulka 4.1: Kumulativní trojúhelník $\{C_{i,j}; i + j < 10\}$

Mackův model Chain-Ladder

Uvažujme kumulativní trojúhelník hlášených škod $\{C_{i,j}; i+j < 10\}$ uvedených v tabulce 4.1. Pomocí vztahu (1.2) odhadneme vývojové faktory $\hat{f}_j$. Následně pro $i + j \geq 10$ odhadneme budoucí škody $\left\{\hat{C}_{i,j}; 10 \leq i+j \leq 18\right\}$, tj. kumulativní trojúhelník doplníme na čtverec násobením odhadnutými faktory $\hat{f}_j$.

Tabulka 4.2 obsahuje hodnoty odhadů vývojových faktorů $\hat{f}_j$ a směrodatných odchylek $\hat{\sigma}_j^2$. Pro výpočet $\hat{\sigma}_8^2$ jsme použili extrapolaci uvedenou v (1.8):

$$\hat{\sigma}_8^2 = min\left(\frac{\hat{\sigma}_7^4}{\hat{\sigma}_6^2}, min\left(\hat{\sigma}_6^2, \hat{\sigma}_7^2\right)\right).$$

j	0	1	2	3	4	5	6	7	8
$\hat{f}_j$	3.149	1.256	1.095	1.041	1.016	1.013	1.005	1.004	1.002
$\hat{\sigma}_j^2$	6893.5	155.47	64.133	22.736	11.652	10.140	0.1719	0.0082	0.0004

Tabulka 4.2: Hodnoty odhadů $\hat{f}_j$ a $\hat{\sigma}_j^2$

V tabulce 4.3 jsou zachyceny odhady predikovaných škod pomocí Mackovy metody Chain-Ladder.

i/j	0	1	2	3	4	5	6	7	8	9
0										
1										121 398
2									126 378	126 611
3								76 974	77 295	77 437
4							78 530	78 941	79 270	79 415
5						56 669	57 418	57 719	57 959	58 065
6					69 460	70 603	71 537	71 911	72 211	72 343
7				98 126	102 164	103 846	105 220	105 770	10 6210	106 406
8			64 319	70 441	73 340	74 548	75 534	75 928	76 245	76 385
9		45 452	57 086	6 2520	65 093	66 165	67 040	67 390	67 671	67 795

Tabulka 4.3: Odhad škod $\hat{C}_{i,j}$ pomocí Mackovy metody Chain-Ladder

Na základě vztahu (1.4) vypočítáme odhad jednotlivých rezerv. Součtem těchto rezerv dostaneme odhad celkové rezervy na pojistná plnění. Pomocí tvrzení ve větě 5 spočteme střední kvadratickou chybu jednotlivých rezerv. Střední kvadratickou chybu celkové rezervy na pojistná plnění získáme ze vztahu (1.16). Tabulka 4.4 obsahuje hodnoty jednotlivých rezerv a celkové rezervy společně s jejich střední kvadratickou chybou a standardní odchylkou. Vidíme, že odhad celkové rezervy na pojistná plnění je 107 041 tisíc dolarů se standardní odchylkou 17 769 tisíc dolarů.

i	$\hat{R}_i$	$mse(\hat{R}_i)$	$se(\hat{R}_i)$
0	0	0	0
1	222	100	10
2	757	1 717	41
3	863	17 075	131
4	1 910	966 732	983
5	2 314	1 436 064	1 198
6	5 629	3 688 825	1 921
7	16 809	13 601 158	3 688
8	25 174	21 708 266	4 659
9	53 362	258 444 395	16 076
Celkem	107 041	315 723 853	17 769

Tabulka 4.4: Odhad rezerv na pojistná plnění Mackovou metodou Chain-Ladder

Zobecněný Poissonův model

V této části vypočítáme odhad rezervy na pojistná plnění a její střední kvadratickou chybu (resp. standardní odchylku) zobecněného Poissonova modelu daného rovnicemi (2.20) – (2.22). Pro pozdější porovnání výsledků s ostatními metodami budeme při výpočtu vycházet z kumulativního vývojového trojúhelníku uvedeného v tabulce 4.1. Pro zobecněný lineární model potřebujeme nekumulativní trojúhelník jednotlivých škod. Nekumulativní trojúhelník škod $D_{i,j}$ je uveden v tabulce 4.5.

i/j	0	1	2	3	4	5	6	7	8	9
0	18 825	60 092	20 773	7 072	2 518	1 877	936	695	492	208
1	21 019	63 660	21 007	7 504	4 410	1 651	762	682	481	
2	25 211	55 192	24 303	10 803	5 486	1 137	3 230	492		
3	19 413	34 474	13 118	4 942	3 219	671	737			
4	18 897	35 715	7 829	7 297	4 909	2 858				
5	13 613	23 112	12 025	5 665	1 336					
6	18 712	27 807	11 479	8 716						
7	25 328	45 217	19 052							
8	16 011	35 200								
9	14 433									

Tabulka 4.5: Nekumulativní trojúhelník $\{C_{i,j}; i + j < 10\}$

Odhadované hodnoty parametrů zobecněného Poissonova modelu jsou zobrazeny v tabulce 4.6.

Parametr	Odhad parametrů	Standardní chyba
γ	10.0925	0.0909
α_2	0.0673	0.1025
α_3	0.1094	0.1017
α_4	-0.3822	0.1162
α_5	-0.3570	0.1158
α_6	-0.6701	0.1286
α_7	-0.4502	0.1215
α_8	-0.0644	0.1118
α_9	-0.3959	0.1338
α_{10}	-0.5152	0.2257
β_2	0.7650	0.0714
β_3	-0.2156	0.0913
β_4	-0.9768	0.1248
β_5	-1.7243	0.1790
β_6	-2.6002	0.2814
β_7	-2.8030	0.3361
β_8	-3.7178	0.5780
β_9	-3.9395	0.7991
β_{10}	-4.7549	1.7234

Tabulka 4.6: Odhad parametrů zobecněného Poissonova modelu

Ze znalosti parametrů γ, α_i a β_j jsme schopni spočítat fitované hodnoty a očekávané hodnoty jednotlivých škod $\hat{D}_{i,j}$ pro $i = 0, \ldots, 9$, $j = 0, \ldots, 9$. Očekávané nekumulativní hodnoty škod vypočítané na základě zobecněného Poissonova modelu jsou zaneseny v tabulce 4.7.

i/j	0	1	2	3	4	5	6	7	8	9
0	24 161	51 926	19 475	9 096	4 307	1 794	1 465	587	470	208
1	25 845	55 545	20 832	9 730	4 608	1 919	1 567	628	503	222
2	26 954	57 930	21 726	10 148	4 806	2 001	1 634	655	524	232
3	16 486	35 431	13 288	6 207	2 939	1 224	999	400	321	142
4	16 907	36 336	13 628	6 365	3 014	1 255	1 025	411	329	146
5	12 362	26 567	9 964	4 654	2 204	918	749	300	241	106
6	15 401	33 100	12 414	5 798	2 746	1 144	934	374	300	133
7	22 653	48 685	18 259	8 529	4 039	1 682	1 373	550	441	195
8	16 262	34 949	13 108	6 122	2 899	1 207	986	395	316	140
9	14 433	31 019	11 634	5 434	2 573	1 072	875	351	281	124

Tabulka 4.7: Odhad nekumulativních škod $\hat{D}_{i,j}$

Součtem hodnot za diagonálou dostaneme odhady rezerv pro jednotlivé roky. Součtem těchto rezerv dostaneme odhad celkové rezervy získané na základě zobecněného Poissonova modelu. Odhad disperzního parametru ϕ pomocí Pearsonovy statistiky je 616,1. Pomocí vztahů (2.28) a (2.29) vypočítáme střední kvadratickou chybu rezerv. Odmocněním dostaneme standardní odchylku rezerv na pojistná plnění. Vypočítané hodnoty jsou zobrazeny v tabulce 4.8.

i	$\hat{R}_i$	$mse(\hat{R}_i)$	$se(\hat{R}_i)$
0	0	0	0
1	222	285 191	534
2	757	812 984	902
3	863	727 087	853
4	1 910	1 557 571	1 248
5	2 314	1 779 545	1 334
6	5 629	4 715 540	2 172
7	16 809	17 800 448	4 219
8	25 174	32 829 945	5 730
9	53 362	211 516 474	14 544
Celkem	107 041	162 345 818	12 741

Tabulka 4.8: Odhad rezerv na pojistná plnění pomocí zobecněného Poissonova modelu

Srovnáním hodnot v tabulce 4.4 a 4.8 vidíme, že zobecněný Poissonův model dává stejný odhad rezerv jako Mackova metoda Chain-Ladder. Výsledek není překvapující, naopak je uklidňující. Ve studii [18] je dokázán vztah, že zmíněné metody dávají stejné výsledky pro odhad rezerv. Důvod, proč jsme hledali jiný model, který by měl stejný výstup jako Mackův model Chain-Ladder, byla možnost hodnocení modelů pomocí střední kvadratické chyby (resp. standardní chyby) rezervy.

Gamma model

Stejným postupem jako v minulé části vypočítáme odhad rezervy na pojistná plnění a její střední kvadratickou chybu (resp. standardní odchylku) Gamma modelu daného rovnicemi (2.23) – (2.25). Vyjdeme z nekumulativního trojúhelníku uvedeného v tabulce 4.5. Odhadneme hodnoty parametrů tohoto modelu a spočítáme očekávané hodnoty jednotlivých škod. Dále vypočteme odhady rezerv pro jednotlivé roky a odhad celkové rezervy získané na základě Gamma modelu. Odhad disperzního parametru ϕ pomocí Pearsonovy statistiky je 0,107. Pomocí vztahů (2.28) a (2.29), kde volíme $\rho = 2$ vypočítáme střední kvadratickou chybu rezerv. Odmocněním dostaneme standardní odchylku rezerv na pojistná plnění. Vypočítané hodnoty jsou zobrazeny v tabulce 4.9.

i	$\hat{R}_i$	$mse(\hat{R}_i)$	$se(\hat{R}_i)$
0	0	0	0
1	218	13 567	116
2	901	12 9816	360
3	865	89 029	298
4	2 561	712 213	844
5	2 385	570 696	755
6	6 332	4 261 670	2 064
7	17 503	36 376 741	6 031
8	25 194	89 995 566	9 487
9	52 344	591 221 710	24 315
Celkem	108 302	420 225 117	20 499

Tabulka 4.9: Odhad rezerv na pojistná plnění pomocí Gamma modelu

Bootstrap pro Mackův model

I v případě bootstrapování vyjdeme z dat v tabulce 4.1. Při výpočtu rezervy budeme postupovat podle algoritmu popsaného ve třetí kapilote. Body 1 a 2 máme spočítané v části zabývající se Mackovou metodou Chain-Ladder. Dále pomocí rekurze (vzorec (3.12)-(3.13)) zpětně odhadneme původní škody. Kumulativní trojúhelník fitovaných hodnot nalezneme v tabulce 4.10.

i/j	0	1	2	3	4	5	6	7	8	9
0	24 161	76 086	95 561	104 657	108 964	110 758	112 223	112 810	113 280	113 488
1	25 845	81 390	102 222	111 952	116 560	118 479	120 045	120 673	121 176	
2	26 954	84 884	106 610	116 758	121 564	123 565	125 199	125 854		
3	16 486	51 916	65 205	71 411	74 350	75575	76 574			
4	16 907	53 243	66 870	73 235	76 250	77505				
5	12 362	38 929	48 893	53 547	55 751					
6	15 401	48 501	60 916	66 714						
7	22 653	71 338	89 597							
8	16 262	51 211								
9	14 433									

Tabulka 4.10: Kumulativní trojúhelník fitovaných hodnot $\left\{\hat{C}_{i,j}; i+j < 10\right\}$

Tabulka 4.11 obsahuje nekumulativní trojúhelník fitovaných hodnot, které získáme rozdílem sousedních hodnot v předchozí tabulce 4.10.

i/j	0	1	2	3	4	5	6	7	8	9
0	24 161	51 926	19 475	9 096	4 307	1 794	1 465	587	470	208
1	25 845	55 545	20 832	9 730	4 608	1 919	1 567	628	503	
2	26 954	57 930	21 726	10 148	4 806	2 001	1 634	655		
3	16 486	35 431	13 288	6 207	2 939	1 224	999			
4	16 907	36 336	13 628	6 365	3 014	1 255				
5	12 362	26 567	9 964	4 654	2 204					
6	15 401	33 100	12 414	5798						
7	22 653	48 685	18 259							
8	16 262	34 949								
9	14 433									

Tabulka 4.11: Nekumulativní trojúhelník fitovaných hodnot $\left\{\hat{C}_{i,j}; i+j < 10\right\}$

Dále podle vztahu (3.4) spočteme normovaná Pearsonova rezidua ${}_{(P)}r_{i,j}$. Vypočítané hodnoty reziduí obsahuje tabulka 4.12.

i/j	0	1	2	3	4	5	6	7	8	9
0	-34.32	35.83	9.30	-21.22	-27.26	1.95	-13.81	4.46	1.00	0.00
1	-30.01	34.43	1.21	-22.56	-2.91	-6.11	-20.33	2.16	-0.97	
2	-10.61	-11.37	17.48	6.50	9.81	-19.32	39.48	-6.35		
3	22.79	-5.08	-1.47	-16.05	5.16	-15.80	-8.30			
4	15.30	-3.256	-49.67	11.67	34.51	45.23				
5	11.25	-21.19	20.64	14.81	-18.48					
6	26.67	-29.09	-8.39	38.31						
7	17.77	-15.71	5.86							
8	-1.96	1.34								
9	0.00									

Tabulka 4.12: Normovaná Pearsonova rezidua ${}_{(P)}r_{i,j}$

Nyní nastává hlavní krok v bootstrapování - z tabulky reziduí 4.12 budeme tisíckrát s vracením vybírat jednotlivá rezidua. Tím dostaneme 1 000 trojúhelníků naplněných bootstrapovými rezidui $\{{}^{(b)}_{(P)}r^*_{i,j}\}$. Jeden takový výběr je ukázán v tabulce 4.13.

i/j	0	1	2	3	4	5	6	7	8	9
0	35.83	-15.80	-3.25	2.16	-11.37	0.00	-15.80	-20.33	17.48	-6.11
1	1.21	1.00	-1.47	-13.81	22.79	11.25	-18.48	-13.81	-2.91	
2	-21.22	6.50	35.83	-15.71	38.31	26.67	5.16	14.81		
3	-49.67	35.83	34.51	17.77	-8.39	35.83	-1.96			
4	-29.09	20.64	5.86	-1.96	-34.32	-18.48				
5	0.00	45.23	11.25	-18.48	22.79					
6	-29.09	34.51	-2.91	-29.09						
7	-22.56	-21.22	9.30							
8	-21.19	2.16								
9	-6.35									

Tabulka 4.13: Bootstrapová normovaná Pearsonova rezidua ${}^{(1)}_{(P)}r^*_{i,j}$

Tabulka 4.14 obsahuje nekumulativní trojúhelník bootstrapových hodnot vypočtených podle vztahu (3.16).

i/j	0	1	2	3	4	5	6	7	8	9
0	29 731	48 323	19 020	9 303	3 561	1 794	860	94	849	120
1	26 040	55 783	20 619	8 368	6 155	2 412	835	282	438	
2	23 470	59 495	27 009	8 565	7 462	3 195	1 843	1 034		
3	10 108	42 177	17 267	7 607	2 484	2 478	937			
4	13 124	40 271	14 313	6 208	1 130	600				
5	12 362	33 940	11 088	3 393	3 274					
6	11 791	39 379	12 090	3 583						
7	19 256	44 002	19 516							
8	13 558	35 355								
9	13 670									

Tabulka 4.14: Nekumulativní trojúhelník bootstrapových hodnot $\{{}^{(b)}D^*_{i,j}\}$

Součtem hodnot v nekumulativním trojúhelníku bootstrapových hodnot $\{{}^{(b)}D^*_{i,j}\}$ dostaneme kumulativní trojúhelník těchto hodnot $\{{}^{(b)}C^*_{i,j}\}$. Hodnoty jsou zaneseny v tabulce 4.15.

i/j	0	1	2	3	4	5	6	7	8	9
0	29 731	78 054	97 074	106 377	109 938	111 732	112 592	112 686	113 535	113 655
1	26 040	81 822	102 441	110 809	116 964	119 376	120 211	120 493	120 931	
2	23 470	82 964	109 973	118 538	125 999	129 194	131 037	132 071		
3	10 108	52 284	69 551	77 158	79 642	82 120	83 057			
4	13 124	53 395	67 708	73 916	75 046	75 646				
5	12 362	46 302	57 389	60 782	64 056					
6	11 791	51 170	63 259	66 842						
7	19 256	63 258	82 774							
8	13 558	48 913								
9	13 670									

Tabulka 4.15: Nekumulativní trojúhelník bootstrapových hodnot $\{{}^{(b)}D^*_{i,j}\}$

Pomocí Mackovy metody Chain-Ladder spočteme bootstrapové vývojové faktory ${}^{(b)}f^*_j$. Dále spočteme bootstrapový odhad $({}^{(b)}\hat{\sigma}^2_j)^*$. Tabulka 4.16 obsahuje vypočtené bootstrapové odhady.

j	0	1	2	3	4	5	6	7	8
$\hat{f}^*_j$	3.5008	1.2767	1.0828	1.0439	1.0206	1.0101	1.0038	1.0055	1.0010
$({}^{(b)}\hat{\sigma}^2_j)^*$	8592.09	99.831	24.870	29.349	6.3959	1.3929	1.7163	0.8874	0.4588

Tabulka 4.16: Hodnoty bootstrapových odhadů ${}^{(b)}f^*_j$ a $({}^{(b)}\hat{\sigma}^2_j)^*$

Na základě vztahu (1.4) Mackovy metody Chain-Ladder spočteme jednotlivé bootstrapové rezervy. Aritmetickým průměrem získáme rezervy na pojistná plnění pro jednotlivé roky. Součtem těchto hodnot dostaneme celkovou rezervu. Dále analogicky Mackově modelu vypočítáme střední kvadratickou chybu jednotlivých rezerv a střední kvadratickou chybu celkové rezervy na pojistná plnění. Tabulka 4.17 zahrnuje hodnoty jednotlivých rezerv společně s jejich střední kvadratickou chybou a standardní odchylkou. Mimo jiné tabulka obsahuje i celkovou hodnotu rezervy a její střední kvadratickou chybu a standardní chybu, kterou spočítáme na základě vztahu (3.17). Z tabulky vidíme, že odhad celkové rezervy na pojistná plnění založený na metodě bootstrap je 106 534 tisíc dolarů se standardní odchylkou 17 857 tisíc dolarů.

i	$\hat{R}_i$	$mse(\hat{R}_i)$	$se(\hat{R}_i)$
0	0	0	0
1	220	148 102	385
2	564	503 867	710
3	808	456 571	676
4	1 807	990 509	995
5	1 941	1 127 550	1 062
6	4 734	2 938 923	1 714
7	17 570	9 752 817	3 123
8	29 393	17 097 240	4 135
9	49 495	115 122 223	10 730
Celkem	106 534	318 860 080	17 857

Tabulka 4.17: Odhad rezerv na pojistná plnění pomocí metody bootstrap

Porovnání jednotlivých metod

Nyní porovnáme výsledky získané podle jednotlivých metod. Odhady rezerv na pojistná plnění získané Mackovým modelem Chain-Ladder, zobecněným Poissonovým modelem, Gamma modelem a pomocí metody bootstrap jsou zaneseny v tabulce 4.18. Ekvivalentní standardní odchylky vyjádřené jako procento z odhadu rezervy obsahuje tabulka 4.19.

Připomeňme, že zobecněný Poissonův model dává úplně stejné odhady rezerv jako Mackův stochastický model Chain-Ladder. Dále můžeme vidět, že odhady rezerv vypočtené pomocí Gamma modelu jsou velice blízké Mackovým odhadům. Metoda bootstrap aplikovaná na Mackův model dává trochu nižší hodnoty rezerv.

Rok	Mackův model	Poisson ZLM	Gamma ZLM	Bootstrap
0	0	0	0	0
1	222	222	218	220
2	757	757	901	564
3	863	863	865	808
4	1 910	1 910	2 561	1 807
5	2 314	2 314	2 385	1 941
6	5 629	5 629	6 332	4 734
7	16 809	16 809	17 503	17 570
8	25 174	25 174	25 194	29 393
9	53 362	53 362	52 344	49 495
Celkem	107 041	107 041	108 302	106 534

Tabulka 4.18: Rezervy na pojistná plnění

Podívejme se nyní na standardní odchylky předpovídaných rezerv. Ačkoliv standardní odchylky pro celkové rezervy získané pomocí zobecněného Poissonova modelu jsou skoro identické s odchylkami ostatních modelů, vidíme velké rozdíly díváme-li se na jednotlivé škodní roky. Největší rozdíl je pro škodní rok $i = 1$, kdy Poissonův model dává standardní odchylku 240 %. Je třeba poznamenat, že jmenovatel (odhad rezervy) je velice nízký a velká odchylka v prvním roce není neočekávaná. Bootstrapová standardní chyba (založená na 1 000 simulacích) je velice blízká analytické chybě Poissonova modelu pro jednotlivé rezervy. Naopak u celkové rezervy se výsledek blíží Mackově modelu Chain-Ladder.

Rok	Mackův model	Poisson ZLM	Gamma ZLM	Bootstrap
0	0 %	0 %	0 %	0 %
1	4 %	240 %	53 %	175 %
2	5 %	119 %	40 %	126 %
3	15 %	99 %	35 %	84 %
4	51 %	65 %	33 %	55 %
5	52 %	58 %	32 %	55 %
6	34 %	39 %	33 %	36 %
7	22 %	25 %	34 %	18 %
8	19 %	23 %	38 %	14 %
9	30 %	27 %	46 %	22 %
Celkem	17 %	12 %	19 %	17 %

Tabulka 4.19: Standardní chyba jako procento z rezervy

4.2 Simulace

V druhé podkapitole praktické části aplikujeme teoretické poznatky na nasimulovaná data. Tabulka 4.20 obsahuje nekumulativní trojúhelník nasimulovaných škod. Při simulaci se omezíme na hodnoty v rozmezí 10 000 až 65 000.

i/j	0	1	2	3	4	5	6	7	8	9
0	36 956	21 189	34 088	61 689	34 873	17 398	64 203	46 928	62 570	45 089
1	49 264	59 621	61 535	22 314	16 614	10 827	29 415	39 359	61 354	
2	62 716	38 474	63 941	27 320	28 394	56 555	63 502	33 552		
3	28 828	19 844	41 144	63 302	48 252	32 494	44 868			
4	26 079	51 177	42 238	32 603	53 471	36 686				
5	56 306	15 712	51 995	51 598	12880					
6	33 048	49 617	18 771	27 369						
7	32 215	44 631	18 559							
8	37 740	10 982								
9	11 424									

Tabulka 4.20: Nekumulativní trojúhelník nasimulovaných hodnot

Při výpočtech odhadů rezerv a příslušných standardních odchylek postupujeme stejně jako v případě s reálnými daty. Nebudeme zde již popisovat jednotlivé dílčí kroky výpočtu, ale uvedeme pouze výsledné odhady rezerv a jejich standardní odchylky.

V tabulce 4.21 vidíme odhady rezerv na pojistná plnění. Odhady rezerv pro Mackův model a zobecněný Poissonův model vyjdou shodně. Gamma model dává odhad rezervy velice blízký hodnotě Mackova modelu. Výsledný odhad rezervy na pojistná plnění získaný metodou bootstrap je vyšší než odhady pro ostatní modely.

Rok	Mackův model	Poisson ZLM	Gamma ZLM	Bootstrap
0	0	0	0	0
1	41 577	41 577	41 193	25 226
2	130 067	130 067	134 681	128 682
3	149 096	149 096	157 746	156 336
4	209 686	209 686	224 354	162 217
5	215 922	215 922	206 840	309 044
6	200 246	200 246	202 792	216 615
7	229 190	229 190	232 648	252 268
8	205 076	205 076	192 034	228 661
9	99 088	99 088	95 876	128 484
Celkem	1 479 947	1 479 947	1 488 164	1 607 532

Tabulka 4.21: Rezervy na pojistná plnění pro nasimulovaná data

Nakonec se podívejme na tabulku 4.22, která obsahuje standardní odchylky rezerv vyjádřené jako procento z příslušné rezervy. Zobecněný Poissonův model, Gamma model i bootstrap dávají skoro stejné odchylky celkové rezervy. V případě nasimulovaných dat dává Mackův model nejmenší standardní odchylku. V případě odchylek pro jednotlivé rezervy vidíme, že pro Mackův model vychází nejmenší odchylky kromě posledního roku $i = 9$, kdy standardní odchylka vyšla 85 %. U Poissonova modelu je největší rozdíl také ve škodním roce $i = 9$, a to 93 %. Pro Gamma model a metodu bootstrap vyšel největší rozdíl standardní odchylky v prvním škodním roce, číselně 75 % pro Gamma model (resp. 62 % pro bootstrap).

Rok	Mackův model	Poisson ZLM	Gamma ZLM	Bootstrap
0	0 %	0 %	0 %	0 %
1	3 %	68 %	75 %	62 %
2	4 %	46 %	56 %	29 %
3	13 %	44 %	50 %	26 %
4	17 %	41 %	46 %	32 %
5	21 %	41 %	45 %	18 %
6	28 %	44 %	45 %	26 %
7	37 %	45 %	47 %	27 %
8	42 %	54 %	51 %	34 %
9	85 %	93 %	62 %	59 %
Celkem	13 %	17 %	18 %	19 %

Tabulka 4.22: Standardní chyba jako procento z rezervy pro nasimulovaná data

4.3 Výběr metody

Zaměřme se nyní na analýzu získaných výsledků. V případě reálných dat metoda bootstrap určila nejmenší hodnotu rezervy (číselně 106 534), ale standardní chyba vyšla na 17 % z hodnoty rezervy, zatímco podle zobecněného Poissonova modelu vyšla standardní chyba na 12 % z rezervy o výšce 107 041. Zdůrazněme, že standardní chyba pro první vývojový rok podle zobecněného Poissonova modelu vyšla 285 191, což je 240 % z rezervy pro daný rok. Oproti tomu v simulaci vyšla nejmenší hodnota rezervy v případě výpočtu pomocí Mackovy metody a zobecněného Poissonova modelu. Jak již bylo zmíněno tyto modely dávají stejný odhad rezervy na pojistná plnění. Nejmenší hodnotu standardní chyby vyjádřené jako procento z vypočtené rezervy jsme dostali také pro tyto metody. U Mackova modelu vyšla standardní chyba na 13 % z rezervy, u zobecněného Poissonova modelu na 17 % z této rezervy. Musíme konstatovat, že výběr vhodné metody záleží vždy na aktuárském úsudku a na datech, na která chceme metodu použít.

Závěr

Cílem diplomové práce bylo popsat základní principy stochastických metod pro stanovení odhadu rezervy na pojistná plnění škod vzniklých, ale dosud nenahlášených společně se střední kvadratickou odchylkou odhadu rezervy a následně demonstrovat aplikaci metod na reálných a nasimulovaných datech.

První tři kapitoly jsme věnovali teoretické části. Představili jsme si zde Mackův stochastický model metody Chain-Ladder, zobecněné lineární modely a metodu bootstrap. Uvedli jsme předpoklady jednotlivých metod. Odvodili jsme vztahy jak pro výpočet rezervy na pojistná plnění, tak pro stanovení střední kvadratické chyby (resp. standardní chyby).

V praktické části, které odpovídá čtvrtá kapitola, jsme se zaměřili na aplikaci jednotlivých metod na konkrétní data. Nejprve jsme vyšli z kumulativního trojúhelníku, který byl publikovaný v [13]. Poté jsme aplikovali modely na nasimulovaná data. Nakonec jsme všechny metody porovnali.

Na závěr poznamenejme, že bychom se dále mohli věnovat jiným modelům pro odhad rezervy na pojistná plnění, jako např. Bornhuetter-Fergusonově metodě, či Merz-Wüthrichově metodě vhodné pro účely Solvency II.

Seznam použité literatury

[1] Braun, Ch.: *The prediction error of the Chain Ladder method applied to correlated run-off triangles*, ASTIN Bulletin, Vol. 34, No. 2, 399-423, 2004.

[2] Dobson, A.J.: *An Introduction to Generalized Linear Models*, Chapman Hall/CRC, 2002.

[3] Branda, M.: *Zobecněné lineární modely v pojišťovnictví*, Seminář z aktuárských věd, 2012.

[4] De Jong, P., Heller, G.: *Generalized linear models for insurance data*, Cambridge, 2008.

[5] England, P. , Verrall, R.: *Analytic and bootstrap estimates of prediction errors in claims reserving*, British Act. Journal, 2002.

[6] England, P. , Verrall, R.: *Stochastic claims reserving in general insurance*, Insurance: Mathematics and Economics 25, 281–293, 1999.

[7] England, P.: *Addendum to "Analytic and bootstrap estimates of prediction errors in claims reserving"*, Insurance: Mathematics and Economics 31, 461–466, 2002.

[8] Kulich, M.: *Zobecněné lineární modely*, Zápisky z přednášky (NSTP196), 2012.

[9] Mack, T., Venter, G.: *A Comparison of stochastic model that reproduce Chain Ladder reserves estimates*, ASTIN Colloquium International Actuarial Association, 263-274, 1999.

[10] Mack, T.: *Distribution-free calculation of the standard error of Chain Ladder reserve estimates*, ASTIN Bulletin, No. 2, 1993.

[11] Mack, T.: *The standard error of Chain Ladder reserve estimates: Recursive calculation and inclusion of a tail factor*, ASTIN Bulletin, No. 2, 361-366, 1999.

[12] Mandl, P., Mazurová, L.: *Matematické základy neživotního pojištění*, MATFYZPRESS, 1999.

[13] PartnerRe Ltd.: *2011 Loss Development Triangles*, 2011.

[14] Pešta, M.: *Bootstrap methods in reserving*, Seminář z aktuárských věd, 2011.

[15] Pešta, M.: *Bootstrapping the triangles*, Seminář z aktuárských věd, 2012.

[16] Petrová, I.: *Komparace Value at Risk a Expected Shortfall v rámci Solvency II*, 7. mezinárodní konference Finanční řízení podniků a finančních institucí, 2009.

[17] Prášková, Z.: *Metoda bootstrap*, ROBUST´2004, (J. Antoch and G. Dohnal, eds), JČMF, Praha, 299-314, 2004.

[18] Renshaw, A. , Verrall, R.: *A stochastic model underlying the Chain-ladder technique*, British Act. Journal, 1998.

[19] Šimurda, M.: *Zobecněný lineární model (GLM)*, Seminář z aktuárských věd, 2008.