Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE

Bc. Jan Pechanec

Statistické vlastnosti odhadu technických rezerv neživotního pojištění

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: RNDr. Petr Jedlička, Ph.D.

Studijní program: matematika, finanční a pojistná matematika

2008

Na tomto místě bych chtěl poděkovat vedoucímu diplomové práce panu RNDr. Petru Jedličkovi, Ph.D. za cenné rady a připomínky, které mi při psaní práce poskytl.

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

V Praze dne 10. prosince 2008 Jan Pechanec

Obsah

Úvod **6**

1 Úvod do problematiky **7**
1.1 Technické rezervy 7
1.2 Vývojové trojúhelníky 7

2 Metoda Chain ladder **10**
2.1 Deterministická verze Chain ladder 10
2.2 Stochastický pohled na Chain ladder 11
2.2.1 Základní vlastnosti 11
2.2.2 Variabilita odhadů 12
2.2.3 Odhady škodních úhrnů a rezerv 14

3 Modely založené na vývojových faktorech **17**
3.1 Motivace 17
3.2 Geometrické pojetí trendů 17
3.3 Deterministické modely 21
3.4 Stochastické modely 22

4 Modely Probabilistic trend family (PTF) **25**
4.1 Popis modelu 25
4.2 Regresní model 27
4.3 Hledání modelu 30
4.4 Odhady budoucích hodnot 31
4.4.1 Stabilita modelu 31
4.4.2 Predikce hodnot 31

5 Modely Chain ladder a PTF v praxi **33**
5.1 Aplikace Chain ladder 33

5.1.1 Výsledky CL 33
5.1.2 Statistické testování předpokladů 35
5.2 Aplikace PTF 41
5.2.1 Tvorba modelu 41
5.2.2 Statistické vlastnosti modelu 44
5.2.3 Odhad rezervy PTF modelem 47
5.3 Srovnání výsledků 48

Závěr **50**

Literatura **51**

A Matematický aparát **52**
A.1 Normální a logaritmicko-normální rozdělení 52
A.2 Regresní model 53

B Přílohy **56**
B.1 Tabulky 56
B.2 Obrázky 60

Název práce: Statistické vlastnosti odhadu technických rezerv neživotního pojištění
Autor: Bc. Jan Pechanec
Katedra: Katedra pravděpodobnosti a matematické statistiky
Vedoucí diplomové práce: RNDr. Petr Jedlička, Ph.D.
E-mail vedoucího: petr.jedlicka@ckp.cz

Abstrakt: V předložené práci studujeme dvě rozdílné statistické metody pro odhadování IBNR rezervy, která je součástí rezervy na pojistná plnění. První metodou je stochastická verze Chain ladder a druhou jsou PTF modely. Tyto metody teoreticky popisujeme a ukazujeme jejich odlišné vlastnosti. Metoda Chain ladder nepředpokládá žádné rozdělení škodních úhrnů, na druhou stranu PTF modely předpokládají normální rozdělení logaritmovaných přírůstkových dat. V praktické části práce pak aplikujeme obě metody na ilustrativní data, jejichž výsledky následně porovnáváme. Zkoumáme především pravděpodobnostní rozdělení odhadu rezerv a jeho statistické vlastnosti. Důležitou součástí je také statistické testování předpokladů obou metod.
Klíčová slova: IBNR rezervy, Chain ladder, PTF modely

Title: Statistical properties of the estimate of non-life insurance technical reserves
Author: Bc. Jan Pechanec
Department: Department of Probability and Mathematical Statistics
Supervisor: RNDr. Petr Jedlička, Ph.D.
Supervisor's e-mail address: petr.jedlicka@ckp.cz

Abstract: In the presented work we study two different statistical methods for estimating IBNR reserve that is a part of the reserve for outstanding claims reserves. The first method is stochastic version of the Chain ladder method and the second one is a PTF model. We describe theories of the methods and show their different properties. The Chain ladder method does not assume any distribution of claims amount, on the other hand PTF models assume normal distribution of logarithms of incremental data. In practical part of this work we apply both methods on illustrative data and then we compare the results. We inspect especially probabilistic distribution of estimate of reserves and their statistical characteristics. An important part of this work is statistical testing of assumptions of both methods.
Keywords: IBNR reserves, Chain ladder, PTF models

Úvod

Při výpočtu technických rezerv na pojistná plnění u odvětví neživotního pojištění se objevuje celá řada praktických problémů, kterým je nutné v případě aplikace standardních i alternativních matematických a statistických metod čelit a zhodnotit, zda pro analyzovaná data je vůbec uvažovaná metoda vhodná.

V této práci se zaměříme na srovnání celého procesu výpočtu IBNR rezervy pro dvě odlišné metody, přičemž nejprve připomeneme nutné teoretické základy, používané statistické testy pro podkladová data, principy tvorby modelu a ověření výsledků pro analyzované vývojové schéma z dat popisující výplaty z tzv. povinného ručení.

Významnou součástí práce bude zhodnocení statistických vlastností získaných odhadů, detailní ověření splnění předpokladů obou metod a celkové srovnání výsledků.

Nutné připomenutí první analyzované metody (stochastická verze Chain ladder) bude uvedené v druhé kapitole spolu s uvedením aparátu, který se použije pro statistické testování předpokladů a výpočty rozdělení výsledné rezervy.

Ve třetí kapitole uvedeme "spojovací článek" mezi metodou Chain ladder připomenutou a alternativním přístupem k modelování logaritmů přírůstkových dat vycházejících ze škodních úhrnů, který se dále využije k představení tzv. PTF modelů ve čtvrté kapitole práce.

Představení statistických testů pro oba přístupy a detailní porovnání obou metod z hlediska statistických vlastností je pak ilustrováno na základě analyzovaných dat v páté kapitole, kdy pro oba modely předvedeme již používané statistické testy a navrhujeme některé rozšiřující přístupy především pro PTF model. Z hodnocení výsledků těchto testů pak činíme závěry o interpretaci získaných výsledků obou použitých metod.

Kapitola 1

Úvod do problematiky

1.1 Technické rezervy

Pojistitelé jsou ze zákona povinni pro provozování své činnosti vytvářet tzv. technické rezervy ke krytí škod nastalých z pojistných událostí. Základní rezervou, kterou musí pojistitel vytvářet, je rezerva na pojistná plnění. Tato rezerva je určena ke krytí závazků z pojistných událostí, které

1. byly nahlášeny, ale nebyly dosud zlikvidovány (RBNS rezervy),
2. již vznikly, ale nebyly dosud nahlášeny (IBNR rezervy).

RBNS rezervy stanovují zpravidla likvidátoři na základě odhadnuté výše škody z jednotlivých pojistných událostí. Není-li to tímto způsobem možné, použije se matematicko-statistická metoda. Rezerva je dále upravována podle dalšího šetření pojistné události v průběhu její likvidace. Naproti tomu se IBNR rezervy stanovují především na základě matematicko-statistických metod. Pokud není možné z objektivních důvodů použít tyto metody, využije se metoda kvalifikovaného odhadu.

1.2 Vývojové trojúhelníky

Při výpočtu rezervy na pojistná plnění potřebujeme mít kromě fundamentálních znalostí pojistného odvětví také informace o škodním vývoji pojistného rizika, tzn. informace o dosud vyplaceném pojistném plnění a vytvořených RBNS rezervách, což představuje nám dosud známý závazek. Nejčastěji se pozorované hodnoty uvádějí ve formě tzv. vývojových trojúhelníků, které

mohou mít formu kumulativní nebo nekumulativní (přírůstkovou). Cílem je tento trojúhelník doplnit na čtverec. Data se sledují za stejně dlouhá období, přičemž obdobím se nejčastěji rozumí 1 rok.

Jak je uvedeno v [9], pokud uvažujeme kumulativní data, označme $C_{i,k}$ celkovou výši škod, které vznikly v období i, $1 \leq i \leq n$, a byly uhrazeny nebo nahlášeny do konce vývojového období k, $0 \leq k \leq n-1$. Hodnoty $C_{i,k}$, kde $i+k \leq n$, patří do známé části čtverce, tedy vývojového trojúhelníku. Ostatní hodnoty, kde $i+k > n$, patří do neznámé části čtverce a potřebujeme je odhadnout. Především nás zajímají hodnoty $C_{i,n}$, které udávají celkový úhrn škod vzniklých v období i a uhrazených nebo nahlášených se zpožděním n období, $i = 2, \ldots, n$. Tyto hodnoty mohou dobře aproximovat celkový škodní úhrn v daném období, pokud předpokládáme, že škodní vývoj je po n obdobích ukončen.

	0	1	...	k	...	$n-2$	$n-1$
1	$C_{1,0}$	$C_{1,1}$	...	$C_{1,k}$	...	$C_{1,n-2}$	$C_{1,n-1}$
2	$C_{2,0}$	$C_{2,1}$	...	$C_{2,k}$	...	$C_{2,n-2}$	
...	...	...	...	...	...		
i	$C_{i,0}$	$C_{i,1}$	...	$C_{i,k}$			
...	...	...	...				
$n-1$	$C_{n-1,0}$	$C_{n-1,1}$					
n	$C_{n,0}$						

Tabulka 1.1: Kumulativní vývojový trojúhelník

Máme-li přírůstková data, označme $S_{i,k}$ výši škod vzniklých v období i a uhrazených nebo nahlášených ve vývojovém období k. Použijeme-li označení z kumulativního trojúhelníku, můžeme přírůstková data vyjádřit rozdílem

$$S_{i,k} = C_{i,k} - C_{i,k-1}, \qquad 1 \leq i \leq n, \qquad 0 \leq k \leq n-1,$$

s konvencí $C_{i,-1} = 0$. I zde známe hodnoty $S_{i,k}$ pouze pro $i+k \leq n$ a taktéž potřebujeme odhadnout přírůstky pro $i+k > n$.

Zaveďme dále veličinu R_i a R následujícím způsobem:

$$\begin{aligned} R_i &= C_{i,n} - C_{i,n-i} = S_{i,n+1-i} + \ldots + S_{i,n}, \qquad 2 \leq i \leq n, \\ R &= R_2 + \ldots + R_n. \end{aligned}$$

$C_{i,k}$ (resp. $S_{i,k}$) však nemusí označovat pojistné plnění a RBNS rezervu dohromady, může se také jednat například pouze o pojistné plnění. Proto

jestliže $C_{i,k}$ je označení známého závazku, pak R_i značí IBNR rezervu pro období vzniku i, pokud $C_{i,k}$ značí pojistné plnění, pak R_i značí celou rezervu na pojistná plnění, tj. rezervy RBNS a IBNR dohromady. R označuje celkovou rezervu pro všechna uvažovaná období vzniku.

Nakonec pro zjednodušení zápisu označme D množinu dosud pozorovaných dat

$$D = \{C_{i,k}, i + k \leq n\}.$$

Kapitola 2

Metoda Chain ladder

2.1 Deterministická verze Chain ladder

Metoda Chain ladder je pravděpodobně nejznámější metodou pro stanovení odhadu rezervy na pojistná plnění. Klasická deterministická verze metody Chain ladder odhaduje neznámé hodnoty škodních úhrnů $C_{i,k}$, $i + k > n$, ze známých hodnot D podle vzorce

$$\hat{C}_{i,k} = C_{i,n-i}\hat{f}_{n+1-i} \dots \hat{f}_{k-1}, \qquad i + k > n, \qquad i \geq 2. \tag{2.1.1}$$

kde

$$\hat{f}_k = \frac{\sum_{i=1}^{n-k} C_{i,k+1}}{\sum_{i=1}^{n-k} C_{i,k}}, \qquad 0 \leq k \leq n-2, \tag{2.1.2}$$

jsou odhady tzv. *vývojových faktorů* f_k.

Je zřejmé, že metoda Chain ladder využívá k odhadům neznámých hodnot $C_{i,k}$ kromě odhadů vývojových faktorů $\hat{f}_k$ také poslední známou hodnotu $C_{i,n-i}$ v období vzniku i. Samozřejmě by se mohla místo $C_{i,n-i}$ využívat také jakákoliv dřívější hodnota v daném období vzniku $C_{i,k}, 0 \leq k < n - i$, a s ní i více odhadů vývojových faktorů To by však bylo v rozporu právě s předpokladem metody Chain ladder, že poslední známá hodnota v daném období vzniku je považována za nejlepší dostupný údaj pro odhad konečného škodního úhrnu pro toto období vzniku i.

Hlavní nevýhodou deterministické metody Chain ladder je nemožnost jakkoliv změřit a ohodnotit variabilitu jejích odhadů. V následující části k vyřešení tohoto problému využijeme stochastického přístupu, jak navrhuje Mack v [6].

2.2 Stochastický pohled na Chain ladder

2.2.1 Základní vlastnosti

Uvažujme situaci, že jsme na konci k-tého vývojového období v libovolném období vzniku a známe tedy hodnoty $C_{i,0}, \ldots, C_{i,k}$. Proto můžeme hodnoty $C_{i,k+1}, \ldots, C_{i,n}$ považovat za náhodné veličiny. Metoda Chain ladder ve své deterministické podobě předpokládá existenci vývojových faktorů f_k, které jsou závislé pouze na vývojovém období, nikoliv období vzniku. Pokud se na problém podíváme ze stochastického pohledu, můžeme uvést v souladu s článkem [6] základní předpoklad ve tvaru

$$\mathbb{E}[C_{i,k+1}|C_{i,1}, \ldots, C_{i,k}] = C_{i,k} f_k, \qquad 1 \leq i \leq n, \qquad 0 \leq k \leq n-1. \quad (2.2.1)$$

Vývojové faktory f_k vyjadřují očekávaný přírůstek z $C_{i,k}$ do $C_{i,k+1}$. *Individuální vývojové faktory* $f_{i,k} \stackrel{def}{=} C_{i,k+1}/C_{i,k}$ však mohou být různé od f_k, a proto je můžeme považovat za náhodné odchylky od f_k, což lze vyjádřit ekvivalentním zapsáním předpokladu (2.2.1) ve tvaru:

$$\mathbb{E}[C_{i,k+1}/C_{i,k}|C_{i,1}, \ldots, C_{i,k}] = f_k, \qquad 1 \leq i \leq n, \qquad 0 \leq k \leq n-1. \quad (2.2.2)$$

V následující větě ukážeme, že předpoklad (2.2.1) spolu s předpokladem o nezávislosti škodních období, který bývá často při aplikaci Chain ladder přehlížen, přestože nemusí být vždy splněn (například kvůli silnému efektu inflace v některém kalendářním období), je nutnou součástí metody Chain ladder. Budeme proto dále předpokládat, že náhodné vektory škodních úhrnů

$$\{C_{i,1}, \ldots, C_{i,n}\}, \{C_{j,1}, \ldots, C_{j,n}\}, \qquad i \neq j, \quad (2.2.3)$$

jsou nezávislé.

Věta 2.2.1. *Za předpokladu (2.2.1) a (2.2.3) platí následující:*

1. $\mathbb{E}[C_{i,k+1}|D] = C_{i,n-i} f_{n-i} \ldots f_k, \qquad n-i < k \leq n-2,$
2. $\hat{f}_k$ *jsou nejlepší nestranné a navzájem nekorelované odhady* f_k, $0 \leq k \leq n-2$.

Důkaz. Uveden v [8]. □

Jak uvádí Mack v [6], nekorelovanost $\hat{f}_k$ může být překvapující, protože $\hat{f}_{k-1}$ a $\hat{f}_k$ závisí částečně na stejných datech. Z tohoto důvodu by se neměla metoda Chain ladder aplikovat v takových odvětvích, kde po nízké hodnotě vývojových faktorů následují vysoké hodnoty a naopak.

Z Věty 2.2.1 díky nestrannosti a nekorelovanosti plyne, že

$$\mathbb{E}[\hat{f}_{n-i} \dots \hat{f}_k] = f_{n-i} \dots f_k, \qquad n-i \leq k \leq n-2.$$

Odtud ihned dostáváme

$$\begin{aligned} \mathbb{E}[\hat{C}_{i,k+1}] &= \mathbb{E}[C_{i,n-i}\hat{f}_{n-i} \dots \hat{f}_k] = \\ &= C_{i,n-i}\mathbb{E}[\hat{f}_{n-i} \dots \hat{f}_k] = \\ &= C_{i,n-i}f_{n-i} \dots f_k = \\ &= \mathbb{E}[C_{i,k}|D]. \end{aligned}$$

Výsledkem je odhad škodního úhrnu

$$\hat{C}_{i,k+1} = C_{i,n-i}\hat{f}_{n-i} \dots \hat{f}_k,$$

který je nestranným odhadem

$$\mathbb{E}[C_{i,k+1}|D] = C_{i,n-i}f_{n-i} \dots f_k.$$

2.2.2 Variabilita odhadů

V dalším připomeneme základní statistické vlastnosti odhadů celkového škodního úhrnu $\hat{C}_{i,n}$. Víme, že odhad

$$\hat{C}_{i,n} = C_{i,n-i}\hat{f}_{n-i} \dots \hat{f}_{n-2}$$

je pouze bodovým odhadem, který nám nic neříká o tom, s jakou pravděpodobností a o kolik se budoucí realizace $C_{i,n}$ může od tohoto odhadu lišit. Proto by nás mělo zajímat, jak velká může být v průměru odchylka od tohoto odhadu, k čemuž se využívá, jak uvádí Mack v [6], *střední kvadratická chyba* $mse(\hat{C}_{i,n})$ odhadu $\hat{C}_{i,n}$, která je definována

$$mse(\hat{C}_{i,n}) = \mathbb{E}\left[(\hat{C}_{i,n} - C_{i,n})^2|D\right].$$

a obdobně definujeme také

$$mse(\hat{R}_i) = \mathbb{E}\left[(\hat{R}_i - R_i)^2|D\right].$$

Zřejmě platí

$$mse(\hat{R}_i) = \mathbb{E}\left[(\hat{R}_i - R_i)^2|D\right] = \mathbb{E}\left[(\hat{C}_{i,n} - C_{i,n})^2|D\right] = mse(\hat{C}_{i,n})$$

a také

$$mse(\hat{R}_i) = \mathrm{Var}[C_{i,n}|D] + \left[\mathbb{E}[\hat{C}_{i,n}] - \hat{C}_{i,n}\right]^2 = mse(\hat{C}_{i,n}),$$

což ukazuje rozklad $mse(\hat{C}_{i,n})$ na dvě složky - náhodnou chybu a chybu modelu.

Než přejdeme k výpočtu odhadu $mse(\hat{C}_{i,n})$, je nutné se zmínit o podmíněném rozptylu individuálních vývojových faktorů. Jak ukazuje Mack v [7], za podmínky nezávislosti období vzniku (2.2.3) je rozumný předpoklad, že podmíněný rozptyl $C_{i,k+1}/C_{i,k}$ nepřímo úměrný $C_{i,k}$. Tedy

$$\mathrm{Var}\left[\frac{C_{i,k+1}}{C_{i,k}}|C_{i,1},\ldots,C_{i,k}\right] = \frac{\sigma_k^2}{C_{i,k}}, \tag{2.2.4}$$

kde $\sigma_k^2 \geq 0$ je neznámý parametr. Protože $C_{i,k}$ je díky podmínění známá hodnota, můžeme vyjádřit podmíněný rozptyl $C_{i,k+1}$ ve tvaru

$$\mathrm{Var}[C_{i,k+1}|C_{i,1},...,C_{i,k}] = C_{i,k}\sigma_k^2, \qquad 1 \leq i \leq n, \qquad 0 \leq k \leq n-2. \tag{2.2.5}$$

Tento vztah (2.2.5), stejně jako předpoklad o nezávislosti období vzniku (2.2.3) a (2.2.1), je nutné testovat, zda-li je splněn, k čemuž se dostaneme v další části textu.

Mack v článku [7] dále uvádí výpočet nestranných odhadů parametrů σ_k^2 podle vzorce

$$\hat{\sigma}_k^2 = \frac{1}{n-k-1}\sum_{i=1}^{n-k} C_{i,k}\left(\frac{C_{i,k+1}}{C_{i,k}} - \hat{f}_k\right)^2, \qquad 1 \leq k \leq n-3. \tag{2.2.6}$$

Pro $k = n-2$ lze odhad σ_{n-2}^2 vypočítat například způsobem

$$\hat{\sigma}_{n-2}^2 = \min\left\{\hat{\sigma}_{n-3}^4/\hat{\sigma}_{n-4}^2, \min(\hat{\sigma}_{n-4}^2, \hat{\sigma}_{n-3}^2)\right\}.$$

Nyní již máme všechny potřebné podklady k formulaci odhadu $mse(\hat{C}_{i,n})$.

Věta 2.2.2. *Nechť jsou období vzniku škodních událostí nezávislá dle (2.2.3) a předpokládejme (2.2.1) pro $i = 2,\ldots,n$. Pak platí*

$$\widehat{mse(\hat{R}_i)} = \hat{C}_{i,n}^2 \sum_{k=n-i}^{n-2} \frac{\hat{\sigma}_k^2}{\hat{f}_k^2}\left(\frac{1}{\hat{C}_{i,k}} + \frac{1}{\sum_{j=1}^{n-k} C_{j,k}}\right). \tag{2.2.7}$$

Důkaz. Uveden v [6]. □

Pro zjednodušení zápisu vzorce (2.2.7) je v předchozí větě použito označení $\hat{C}_{i,n-i} = C_{i,n-i}$.

Důsledek 2.2.3. *Za předpokladu (2.2.1) a (2.2.3) platí pro odhad celkové rezervy R*

$$\widehat{mse(\hat{R})} = \sum_{i=2}^{n} \left\{ \widehat{mse(\hat{R}_i)} + \hat{C}_{i,n} \left(\sum_{j=i+1}^{n} \hat{C}_{j,n} \right) \sum_{k=n-i}^{n-2} \frac{2\hat{\sigma}_k^2 / \hat{f}_k^2}{\sum_{j=1}^{n-k-1} C_{j,k}} \right\}. \quad (2.2.8)$$

2.2.3 Odhady škodních úhrnů a rezerv

Významnou veličinou, která se vypočítá ze střední kvadratické chyby $mse(\hat{C}_{i,n})$, je standardní odchylka (*standard error*) $s.e.(\hat{C}_{i,n})$ počítaná podle vzorce

$$s.e.(\hat{C}_{i,n}) = \sqrt{mse(\hat{C}_{i,n})} = \sqrt{mse(\hat{R}_i)} = s.e.(\hat{R}_i).$$

Odhad $\widehat{s.e.(\hat{R}_i)}$ je zřejmě stanoven způsobem

$$\widehat{s.e.(\hat{R}_i)} = \sqrt{\widehat{mse(\hat{R}_i)}}.$$

Přirozeně by nás zcela jistě mělo zajímat rozpětí hodnot, ve kterém se může realizace $C_{i,n}$, resp. R_i, pohybovat s jistou pravděpodobností. V následujícím se proto budeme zabývat, jakým způsobem lze odvodit odhad pravděpodobnostního rozdělení$C_{i,n}$, resp. R_i.

Víme, že $C_{i,n} = C_{i,n-i} + R_i$. Protože $C_{i,n-i}$ je známá hodnota, znamená to, že pravděpodobnostní rozdělení $C_{i,n}$ je určeno pouze R_i. Proto nám stačí formulovat odhad rozdělení R_i a odtud posunutím získat odhad rozdělení $C_{i,n}$. Jak uvádí Mack v [7], pokud lze celkový škodní úhrn interpretovat jako součet velkého počtu dílčích nezávislých škodních úhrnů jednotlivých pojistných událostí, což bývá v praxi splněno, můžeme díky centrální limitní větě předpokládat, že rozdělení R_i je normální, přičemž jeho střední hodnotu odhadujeme hodnotou $\hat{R}_i$ a rozptyl hodnotou $\left(\widehat{s.e.(\hat{R}_i)}\right)^2$. Odtud můžeme odvodit intervalový odhad potřebného objemu rezervy R_i o spolehlivosti $1-\alpha, \alpha \in (0,1)$, ve tvaru

$$\left(\hat{R}_i - u_{1-\alpha/2} \cdot \widehat{s.e.(\hat{R}_i)}, \hat{R}_i + u_{1-\alpha/2} \cdot \widehat{s.e.(\hat{R}_i)} \right), \quad (2.2.9)$$

což implikuje intervalový odhad celkového škodního úhrnu $C_{i,n}$ o spolehlivosti $1-\alpha$

$$\left(\hat{R}_i + C_{i,n-i} - u_{1-\alpha/2} \cdot \widehat{s.e.(\hat{R}_i)}, \hat{R}_i + C_{i,n-i} + u_{1-\alpha/2} \cdot \widehat{s.e.(\hat{R}_i)}\right). \quad (2.2.10)$$

Pokud by však bylo rozdělení zešikmené, není použití normálního rozdělení příliš vhodné pro porušení symetrie. Podle Macka v [7] je tento problém v praxi výrazný v případě, kdy $s.e.(R_i)$ je větší než 50% R_i. To pak může vést k zápornému dolnímu intervalovému odhadu, což je v rozporu s požadavkem na nezápornost rezerv. V takovém případě doporučuje Mack využít logaritmicko-normální rozdělení. Budeme-li uvažovat, že R_i má logaritmicko-normální rozdělení s parametry μ_i a σ_i^2, dostáváme rovnosti:

$$\hat{R}_i = \mathrm{e}^{\mu_i + \sigma_i^2/2}, \quad (2.2.11)$$

$$\left(s.e.(\hat{R}_i)\right)^2 = \mathrm{e}^{2\mu_i + \sigma_i^2}\left(\mathrm{e}^{\sigma_i^2} - 1\right). \quad (2.2.12)$$

Úpravou těchto rovnic získáme vyjádření odhadů parametrů μ_i a σ_i^2:

$$\hat{\sigma}_i^2 = \ln\left[1 + \frac{\left(\widehat{s.e.(\hat{R}_i)}\right)^2}{\hat{R}_i^2}\right],$$

$$\hat{\mu}_i = \ln(\hat{R}_i) - \hat{\sigma}_i^2/2.$$

Intervalový odhad rezervy R_i s logaritmicko-normálním rozdělení o spolehlivosti $1-\alpha$ je podle (A.1.2) ve tvaru

$$\left(\mathrm{e}^{\hat{\mu}_i - u_{1-\alpha/2}\hat{\sigma}_i^2/2}, \mathrm{e}^{\hat{\mu}_i + u_{1-\alpha/2}\hat{\sigma}_i^2/2}\right). \quad (2.2.13)$$

Konečně si ukážeme, jak odhadnout celkovou rezervu R bez rozlišení období vzniku škodních událostí. Bodový odhad rezervy R je součtem bodových odhadů rezerv pro všechna období vzniku, tj.

$$\hat{R} = \hat{R}_2 + \hat{R}_3 + \ldots + \hat{R}_n.$$

Odhad směrodatné odchylky $s.e.(\hat{R})$ se vypočte ze vztahu (2.2.8) dle vzorce

$$\widehat{s.e.(\hat{R})} = \sqrt{\widehat{mse(\hat{R})}}.$$

Odtud formulujeme vzorec pro výpočet kvantilů celkové rezervy R ve tvaru

$$\mathrm{e}^{\hat{\mu}+u_{1-\alpha/2}\hat{\sigma}^2/2} \tag{2.2.14}$$

a taktéž určíme intervalové odhady R o spolehlivosti $1-\alpha$ ve tvaru

$$\left(\mathrm{e}^{\hat{\mu}-u_{1-\alpha/2}\hat{\sigma}^2/2}, \mathrm{e}^{\hat{\mu}+u_{1-\alpha/2}\hat{\sigma}^2/2}\right), \tag{2.2.15}$$

kde odhady $\hat{\mu}$ a $\hat{\sigma}^2$ se vypočítají analogicky jako odhady $\hat{\mu}_i$ a $\hat{\sigma}_i^2$.

Po zavedení stochastické verze metody Chain ladder jsme postupně uvedli tři předpoklady (2.2.1), (2.2.3) a (2.2.5), které by měly být splněny, abychom mohli model použít. Pro testování splnění těchto podmínek dat existuje řada statistických testů, z nichž některé představíme a zároveň využijeme v páté kapitole.

Kapitola 3

Modely založené na vývojových faktorech

3.1 Motivace

V této kapitole si ukážeme odlišný pohled na vývojové trojúhelníky prezentovaný Zehnwirthem v [10]. Budeme vycházet z předpokladu, že v nekumulativním trojúhelníku existují tři různé trendy - ve vývojových obdobích, obdobích vzniku a kalendářním období, a zobecníme tak metodu Chain ladder z předchozí kapitoly. Nezbytnou složkou zahrnutou v datech je náhodná fluktuace, která udává odchylku od zmíněných trendů. Schématicky můžeme tento vztah znázornit následovně:

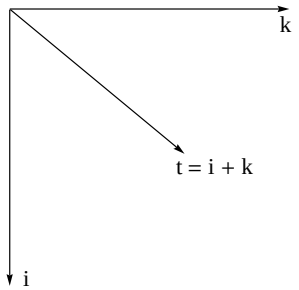
$$data = trendy + náhodná\ fluktuace.$$

Úkolem je odhadnout z pozorovaných hodnot tyto trendy a přiřadit je k odpovídajícím modelům na základě regresní analýzy. V následujícím textu si ukážeme některé přístupy, které budou základem pro stochastické modely vyložené v následující kapitole.

3.2 Geometrické pojetí trendů

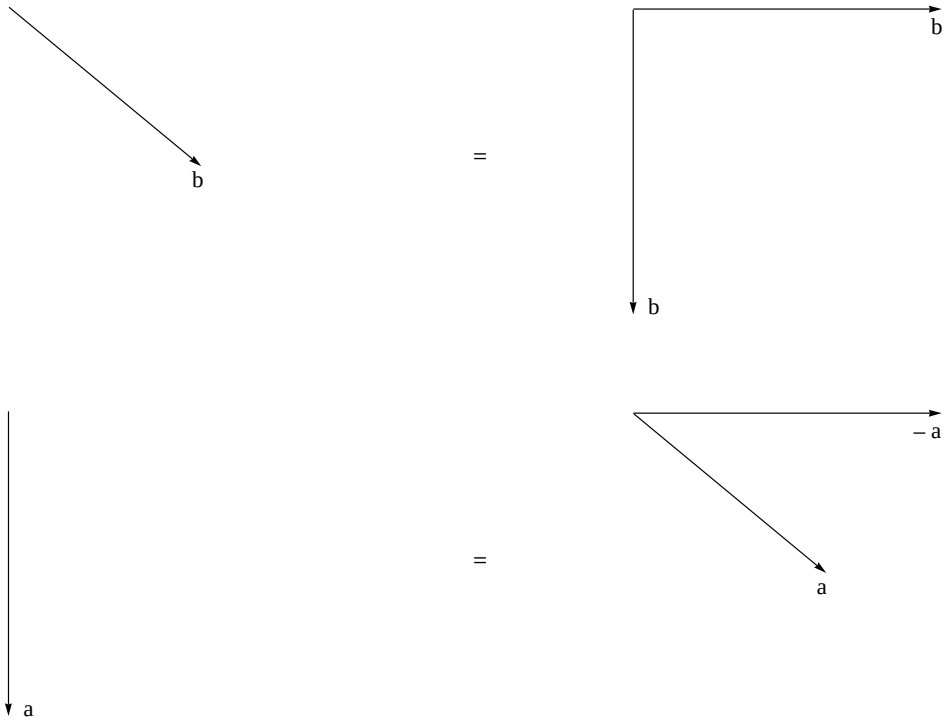
Mějme k-té vývojové období a období vzniku i. Označme dále $t = i + k$, $i + k \leq n$, kalendářní období, ve kterém byla data do trojúhelníku zaznamenána. Odtud je ihned zřejmé, že přestože máme 3 trendy, tak pouze 2 z nich jsou nezávislé.

Vztahy mezi všemi trendy lze jednoduše znázornit také graficky pomocí směrů (směr vzniku, směr vývoje, směr kalendářního období), které jsou v souladu s vývojovým trojúhelníkem, viz Obrázek 3.1.

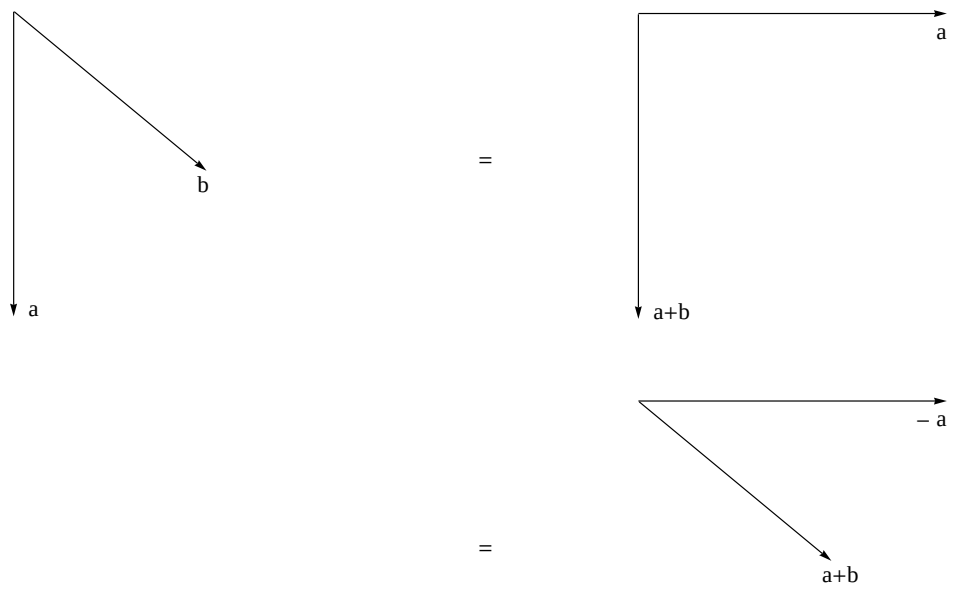


Obrázek 3.1: Trendy v datech

Základní ekvivalentní vztahy mezi trendy vyplývající z definice znázorníme graficky na následujících obrázcích.



Na dalším obrázku znázorníme možnou záměnu trendů ukazující na nejednoznačnost identifikace trendů.



Pro úplnost uvedeme krátký příklad prezentující zmíněné vztahy na nekumulatnvím vývojovém trojúhelníku.

Příklad 3.2.1. Uvažujme přírůstková data uvedená ve vývojovém trojúhelníku 1 (Tabulka 3.1).

	1	2	3	4	5	6
1	50	75	60	50	45	40
2	50	75	60	50	45	
3	50	75	60	50		
4	50	75	60			
5	50	75				
6	50					

Tabulka 3.1: Vývojový trojúhelník 1: homogenní vývoj

Vidíme, že všechna období vzniku mají totožný vývoj, tzn. mají homogenní individuální vývojové faktory. Na tomto trojúhelníku přidejme 10%-ní trend v obdobích vzniku, čímž dostaneme trojúhleník 2.

	1	2	3	4	5	6
1	50	75	60	50	45	40
2	55	83	66	55	50	
3	61	91	73	62		
4	67	100	80			
5	73	110				
6	81					

Tabulka 3.2: Vývojový trojúhelník 2: 10%-ní trend v obdobích vzniku

Dále dodejme ještě 10%-ní trend ve zpoždění a získáme trojúhelník 3, který má v obou kolmých směrech stejný 10%-trend.

	1	2	3	4	5	6
1	50	83	73	67	61	65
2	55	91	80	73	73	
3	62	100	88	82		
4	67	110	97			
5	73	121				
6	81					

Tabulka 3.3: Vývojový trojúhelník 3: 10%-ní trend v obdobích vzniku a 10%-ní trend ve vývojových obdobích

Nyní do trojúhelníku 1 přidejme 10%-ní trend v kalednářním období. Získáme trojúhelník, který je shodný s trojúhelníkem 3, protože tento trend je podle výše uvedených pravidel ekvivalentní se dvěmi trendy, které jsme postupně přidali k původnímu trojúhelníku 1.

Obraťme nyní situaci a představme si, že máme trojúhelník 3 a první analýzou bychom zjistili, že trojúhelník obsahuje právě 10%-ní kalendářní trend. Jinou nezávislou analýzou bychom ale zjistitli, že trojúhelník obsahuje místo tohoto kalendářního trendu 10%-ní trendy v obdobích vzniku a vývojových obdobích, které jsme postupně přidali k původnímu trojúhelníku. Nastává tedy otázka, který výsledek je správný? Odpovedět lze často pouze se znalostí dalších informací, které data v trojúhelníku již neobsahují.

3.3 Deterministické modely

Uvažujme nejprve pouze jedno období vzniku. Označme S_k přírůstek v k-tém vývojovém období a $Y_k = \ln(S_k)$ jeho logaritmovanou hodnotu. Dále definujme hodnotu absolutního členu (*intercept*) $\alpha = \ln(S_0)$ a trend v přírůstcích $\gamma_k = Y_k - Y_{k-1}$.

Zřejmě platí

$$\begin{aligned} Y_k &= Y_0 + Y_1 - Y_0 + \ldots + Y_k - Y_{k-1}, \\ Y_k &= \alpha + \sum_{j=1}^{k} \gamma_j. \end{aligned} \tag{3.3.1}$$

Trend γ_k lze vyjádřit také jako logartimus podílu přírůstků:

$$\begin{aligned} \gamma_k &= Y_k - Y_1 \\ &= \ln(S_k) - \ln(S_{k-1}) \\ &= \ln(S_k/S_{k-1}). \end{aligned}$$

Podíl S_k/S_{k-1} můžeme chápat jako individuální vývojový faktor přírůstkových dat a γ_k je pak logaritmus těchto "vývojových faktorů".

Opusťme předpoklad pouze jednoho období vzniku a zaveďme deterministické modely, které vycházejí z předpokladu existence trendů ve třech diskutovaných směrech, tj. ve vývojových obdobích, obdobích vzniku a kalendářních obdobích. Obecná rovnice pro tyto modely je

$$Y_{i,k} = \alpha_i + \sum_{j=0}^{k-1} \gamma_j + \sum_{t=1}^{i+k} \iota_t, \tag{3.3.2}$$

kde α_i značí počáteční úroveň období vzniku i, γ_j je trend ve vývojovém období j a ι_t označuje trend v kalednářním období t.

Dodejme, že speciálním případem těchto modelů je model

$$Y_{i,k} = \alpha_i + \sum_{j=1}^{k} \gamma_j, \tag{3.3.3}$$

který se nazývá *statistický Chain ladder*, neboť neobsahuje parametry kalendářních trendů.

3.4 Stochastické modely

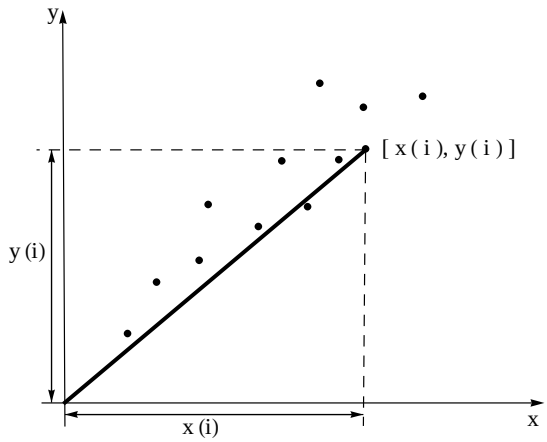
Než přejdeme ke stochastickým modelům odvozených z rovnice (3.3.2), popíšeme ve zkratce v tomto odstavci systém stochastických modelů *Extended link ratio family* (ELRF) prezentovaný Zehnwirthem v článku [10].

Modely ELRF jsou považovány za jakýsi mezičlánek mezi deterministickými metodami vývojových faktorů a stochastickými metodami využívající (zlogaritmovaná) přírůstková data.

ELRF vychází z kumulativních dat $C_{i,k}$. Pro každé období vzniku $i = 1, \ldots, n$ zaveďme označení

$$\begin{aligned} x(i) &= \sum_{s=1}^{i-1} C_{s,i-1}, \\ y(i) &= \sum_{s=1}^{i-1} C_{s,i} \end{aligned}$$

a uvažujme tzv. vývojový poměr (*link ratio*) $y(i)/x(i)$. Každý poměr určuje sklon přímky procházející počátkem grafu $[0, 0]$ a bodem $[x(i), y(i)]$ může být považován za trend, viz Obrázek 3.2. Modely ELRF pak vycházejí z jednoduché myšlenky o existenci průměrného vývojového poměru (trendu).



Obrázek 3.2: Trendy v ELRF modelech

Modely ELRF vychází z regresního modelu

$$y(i) = \beta x(i) + \varepsilon(i), \tag{3.4.1}$$

kde

$$\varepsilon(i) \sim \mathrm{N}(0, \sigma^2 x(i)^\delta). \tag{3.4.2}$$

Parametr β je průměrným vývojovým trendem, který hledáme. Rozptyl $y(i)$ závisí na $x(i)$ přes funkci $x(i)^\delta$. Například volbou $\delta = 1$ lze získat odhad vývojového poměru ve tvaru váženého průměru vývojových poměrů

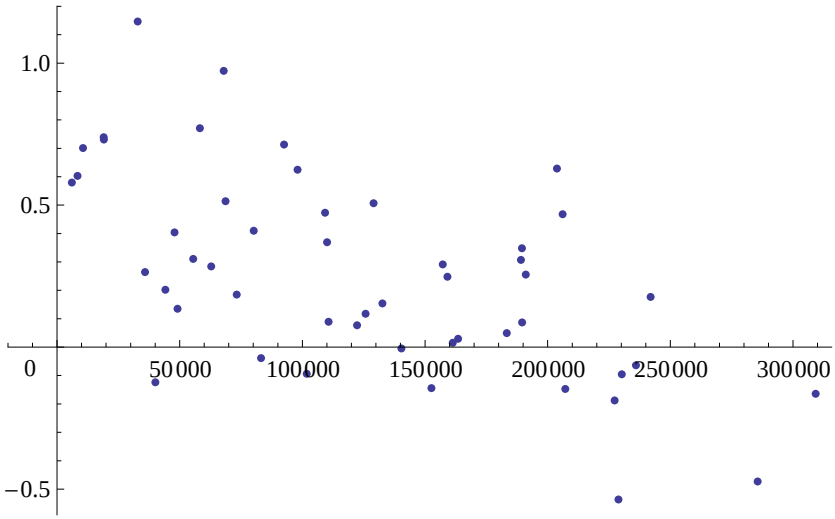
$$\hat{\beta} = \frac{\sum x(i).y(i)/x(i)}{\sum x(i)} = \frac{\sum y(i)}{\sum x(i)}. \tag{3.4.3}$$

Zřejmě platí, že $\hat{\beta}$ je $x(i)$-váženým průměrem vývojových faktorů užívaných v metodě Chain ladder.

Základním předpokladem ELRF je

$$\mathbb{E}[y(i)|x(i)] = \beta x(i). \tag{3.4.4}$$

V praxi však tento předpoklad ale nebývá příliš často splněn, jak můžeme vidět na Obrázku 3.3, kde jsou zobrazena standardizovaná residua v závislosti na $x(i)$ indikující klesající trend. Obrázek značí situaci, ve které jsou nízké pozorované hodnoty v předpovědích podhodnoceny a vysoké naopak nadhodnoceny.



Obrázek 3.3: Standardizovaná rezidua v závislosti na x(i)

Tento nedostatek vedl k vyvoření obecného modelu ELRF

$$y(i) - x(i) = \alpha + \gamma i + (\beta - 1)x(i) + \varepsilon(i), \tag{3.4.5}$$

kde

$$\mathrm{Var}[\varepsilon(i)] = \sigma^2 x(i)^\delta.$$

Bližším pohledem na rovnici (3.4.5) zjistíme, že ELRF model nebere v úvahu možnost změn v trendech v období vývoje pro různé období vzniku, což je vzhledem k praktickému využití velký nedostatek.

Kapitola 4

Modely Probabilistic trend family (PTF)

4.1 Popis modelu

V této kapitole navážeme na deterministický model (3.3.2) a rozšíříme jej o náhodnou složku, čímž získáme novou skupinu modelů, kterou se zde budeme detailněji zabývat.

Mějme logaritmovaná přírůstková data a uvažujme nejprve pouze jedno libovolné období vzniku. Pak náhodná veličina Y_k je přírůstek v k-tém vývojovém období ve tvaru

$$Y_k = \ln S_k = \alpha + \sum_{j=1}^{k} \gamma_j + \varepsilon, \tag{4.1.1}$$

kde $\varepsilon \sim N(0, \sigma^2)$.

Rovnice (4.1.1) popisuje model, ve kterém $Y_0 \sim N(\alpha, \sigma^2)$. γ_j je střední hodnota trendu mezi zpožděním $j-1$ a j, protože z rovnice (4.1.1) lze vyjádřit

$$Y_j - Y_{j-1} \quad = \quad \gamma_j + \varepsilon_j - \varepsilon_{j-1}, \tag{4.1.2}$$

kde ε_j je odchylka pro vývojové období j. Odtud pak

$$\mathbb{E}\left[Y_j - Y_{j-1}\right] = \mathbb{E}\left[\log \frac{S_j}{S_{j-1}}\right] = \gamma_j. \tag{4.1.3}$$

Z dosud uvedeného vyplývá, že přírůstky S_j mají logaritmicko-normální rozdělení následujícími vlastnostmi:

$$
\begin{aligned}
\text{median}(S_j) &= \mathrm{e}^{\alpha+\sum_{j=1}^{k}\gamma_j}, \\
\mathbb{E}[S_j] &= \text{median}(S_j).\mathrm{e}^{0.5\sigma^2}, \\
\sqrt{\text{Var}[S_j]} &= \mathbb{E}[S_j].\sqrt{\mathrm{e}^{\sigma^2}-1}.
\end{aligned}
$$

Zřejmě $Y_j - Y_{j-1} \sim N(\gamma_j, 2\sigma^2)$ a proto podíl přírůstků mezi obdobími $j-1$ a j má střední hodnotu

$$
\mathbb{E}\left[\frac{S_j}{S_{j-1}}\right] = \mathrm{e}^{\gamma_j+\sigma^2}.
$$

Opusťme dále předpoklad pouze jednoho období vzniku a uvažujme model

$$
Y_{i,k} = \alpha_i + \sum_{j=1}^{k}\gamma_j + \sum_{t=2}^{i+k}\iota_t + \varepsilon_{i,k}, \tag{4.1.4}
$$

kde $\varepsilon_{i,k}$ má normální rozdělení s nulovou střední hodnotou. Rovnice (4.1.4) je zápis obecného modelu skupiny nazývané PTF (*Probabilistic trend family*). Rovnice popisuje rozdělení každého období $[i,k]$ zvlášť. Jak uvádí Barnett v [2], pokud by rozptyl náhodných veličin $\varepsilon_{i,k}$ nebyl konstantní, je pak nutné tuto skutečnost vzít v úvahu při sestavování modelu.

Jak upozorňuje Boulter v [3], u přírůstkových dat, především pokud máme data o přírůstcích pojistného plnění i RBNS rezerv, může nastat situace, kdy dojde k zápornému přírůstku - poklesu RBNS rezervy. V takovém případě nelze tyto hodnoty zlogaritmovat. Existuje několik možností (např. v [8]), jak se v praxi s touto situací vypořádat. Hodnoty lze, např. jestliže technické skoky v rezervování nemají z pohledu vývoje význam, interpolovat takovým způsobem, aby kumulativní hodnoty byly po skončení období se zápornými hodnotami shodné pro oba soubory dat. Pokud máme velký počet dat, je další možností úplné vynechání záporných dat. Volba přístupu vždy závisí na posouzení konkrétních dat.

Očekávané trendy v jednotlivých směrech jsou určeny následovně:

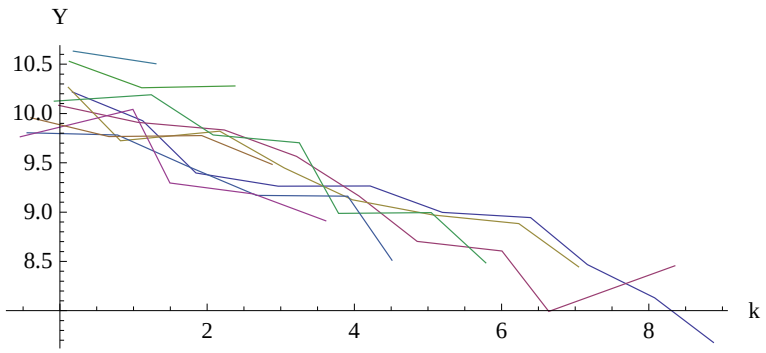
$$
\begin{aligned}
\mathbb{E}\left[Y_{i,k} - Y_{i,k-1}\right] &= \gamma_k + \iota_{i+k}, \\
\mathbb{E}\left[Y_{i,k} - Y_{i-1,k}\right] &= \alpha_i - \alpha_{i-1} + \iota_{i+k}, \\
\mathbb{E}\left[Y_{i,k} - Y_{i-1,k-1}\right] &= \alpha_i - \alpha_{i-1} + \gamma_k + \iota_{i+k}.
\end{aligned} \tag{4.1.5}
$$

Pro názornost se podívejme na jednoduchý příklad.

Příklad 4.1.1. Uvažujme model

$$Y_{i,k} = 10 - 0,3k + 0,05(i + k) + \varepsilon.$$

Model popisuje situaci, ve které přírůstky konstantně klesají se zvyšujícím se vývojovým obdobím a konstantně rostou v obdobích vzniku i v kalendářních obdobích. Na Obrázku 4.1 jsou znázorněny přírůstky v závislosti na období vývoje.



Obrázek 4.1: Přírůstky v závislosti na období vývoje

Kdybychom chtěli z grafu či přímo z dat získat trendy, nelze je stanovit jednoduše pouhým pohledem. Nepomůže k tomu ani technika vývojových faktorů použitá na kumulativní hodnoty. K určení trendů je možné využít pouze regresní přístup, který lze aplikovat na jakýkoliv typ přírůstkového trojúhelníku, tzn. nikoliv pouze na přírůstcích celkových závazků, ale i na počtech škodních událostí či samotné výši pojistného plnění. ◇

4.2 Regresní model

Definujme vektor neznámých parametrů β

$$\beta = (\alpha_1, \ldots, \alpha_n, \gamma_1, \ldots, \gamma_n, \iota_1, \ldots, \iota_{n+1}),$$

a zaveďme náhodný vektor

$$\mathbf{Y} = (Y_{1,0}, Y_{1,1}, \ldots, Y_{1,n-1}, Y_{2,0}, \ldots, Y_{n-1,1}, Y_{n,0}).$$

Ke stanovení odhadu vektoru β potřebujeme mít k dispozici matici $\mathbf{X}$ v regresním modelu

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon.$$

Je zřejmé, že vývojový trojúhelník, kde logaritmy přírůstků na pozici $[i, k]$ popisuje náhodná veličina $Y_{i,k}$, má její střední hodnoty ve tvaru znázorněném v Tabulce 4.1.

	1	$\dots$	k	$\dots$	n
1	α_1	$\dots$	$\alpha_1 + \sum_{j=1}^{k} \gamma_j + \sum_{t=1}^{1+k} \iota_t$	$\dots$	$\alpha_1 + \sum_{j=1}^{n} \gamma_j + \sum_{t=1}^{1+n} \iota_t$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	
i	$\alpha_i + \sum_{t=1}^{1+i} \iota_t$	$\dots$	$\alpha_i + \sum_{j=1}^{k} \gamma_j + \sum_{t=1}^{i+k} \iota_t$		
$\dots$	$\dots$	$\dots$			
n	$\alpha_n + \sum_{t=1}^{1+n} \iota_t$				

Tabulka 4.1: Vývojový trojúhelník středních hodnot PTF modelu

Odtud lze vyjádřit vektor středních hodnot náhohodného vektoru $\mathbf{Y}$ ve tvaru

$$\mathbb{E}\mathbf{Y} = \mathbf{X}\beta = (\alpha_1, \alpha_1 + \gamma_1 + \iota_2, \dots, \alpha_i + \sum_{j=1}^{k} \gamma_j + \sum_{t=1}^{i+k} \iota_t, \dots, \alpha_n + \sum_{t=1}^{n+1} \iota_t). \quad (4.2.1)$$

a taktéž lze stanovit tvar matice $\mathbf{X}$. V obecném tvaru jsou jejími prvky 0 nebo 1 indikující jednotlivá časová období. Vzhledem k jisté složitosti vyjádření obecného tvaru regresní matice $\mathbf{X}$ si její vyjádření ilustrujeme na následujícím příkladě.

Příklad 4.2.1. Stanovení regresní matice budeme ilustrovat na dříve uvedeném Příkladě 4.1.1, kde jsme uvažovali model

$$Y_{i,k} = 10 - 0,3k + 0,05(i + k) + \varepsilon.$$

Přidejme navíc předpoklad $n = 4$, tzn. máme 4 období vzniku $i = 1, \dots, 4$ a 3 vývojové období $k = 0, \dots, 3$. Potom střední hodnota obecného regres-

ního modelu pro $n = 4$ je ve tvaru

$$\begin{pmatrix} Y_{1,0} \\ Y_{1,1} \\ Y_{1,2} \\ Y_{1,3} \\ Y_{2,0} \\ Y_{2,1} \\ Y_{2,2} \\ Y_{3,0} \\ Y_{3,1} \\ Y_{4,0} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \iota_1 \\ \iota_2 \\ \iota_3 \end{pmatrix} \tag{4.2.2}$$

V našem případě díky shodnosti parametrů platí

$$\begin{aligned} \alpha &= \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 10, \\ \gamma &= \gamma_1 = \gamma_2 = \gamma_3 = -0,3, \\ \iota &= \iota_1 = \iota_2 = \iota_3 = \iota_4 = 0,05, \end{aligned}$$

což implikuje jednodušší tvar matice $\mathbf{X}$

$$\begin{pmatrix} Y_{1,0} \\ Y_{1,1} \\ Y_{1,2} \\ Y_{1,3} \\ Y_{2,0} \\ Y_{2,1} \\ Y_{2,2} \\ Y_{3,0} \\ Y_{3,1} \\ Y_{4,0} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 3 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 3 \\ 1 & 0 & 2 \\ 1 & 1 & 3 \\ 1 & 0 & 3 \end{pmatrix} \cdot \begin{pmatrix} \alpha \\ \gamma \\ \iota \end{pmatrix} \tag{4.2.3}$$

$\diamond$

Konečně odhad vektoru β je určen metodou nejmenších čtverců ve tvaru

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}, \tag{4.2.4}$$

pro který platí, že

$$\mathbb{E}[\hat{\beta}] = \beta, \tag{4.2.5}$$

$$\mathrm{Var}[\hat{\beta}] = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}. \tag{4.2.6}$$

4.3 Hledání modelu

V Příkladě 4.2.1 jsme ilustrovali způsob stanovení regresní matice. Navíc jsme ukázali, že lze počet parametrů jednoduchým způsobem snížit v případě, že platnost nárůstu nebo poklesu je delší než 1 období. Právě počet parametrů je při hledání optimálního modelu jedním z kritérií, který určuje kvalitu modelu. Obecně nelze říci, kolik parametrů by měl výsledný model mít, ovšem při nesprávném počtu parametrů může dojít buď k podparametrizování modelu, kdy výsledný model nedostatečně a příliš obecně popisuje pozorovaná data, anebo k jeho přeparametrizování, kdy model je až příliš citlivý na jakékoliv drobnější změny v datech. Důsledkem pak může být nestabilní predikce neznámých budoucích hodnot. Při hledání modelu tedy máme dva protichůdné požadavky (jednoduchost a citlivost modelu), které ovlivňují výsledný počet parametrů. Jestli je počet parametrů modelu vhodný vzhledem k těmto požadavkům, lze posuzovat například na základě minimalizace Akaikeho informačního kritéria (AIC), které je ve tvaru

$$\mathrm{AIC} = N.\ln(2\pi\hat{\sigma}^2) + N + 2p,$$

kde N je počet pozorování, p je počet parametrů a $\hat{\sigma}^2 = S_e/N$. Při hledání modelu bychom stále měli mít na paměti možné ekvivalentní vztahy mezi trendy v obdobích vzniku, vývoje a kalendářních obdobích, které jsme uvedli ve třetí kapitole a jsou pomůckou pro slučování nebo naopak dělení jednotlivých časových intervalů majících funkci vysvětlujících proměnných. S tím také souvisí problém možných korelačních vazeb mezi jednotlivými proměnnými a tím i možného výskytu multikolinearity. Ta narušuje předpoklad lineární nezávislosti sloupců v regresní matici. Důsledkem multikolinearity jsou pak nestabilní modely mající velkou chybu predikce. Jednoduchý test na existenci multikolinearity ukážeme v praktické části této práce.

K hledání optimálního modelu lze využít přístup, ve kterém se vychází z nejjednodušší situace uvažující pouze kontstantní trend ve směru vývoje. Zkoumáním residuí, např. analýzou zlomů v jejich vývoji vzhledem k nezávislým proměnným, se následně přidávají další parametry. Dle významnosti jednotlivých parametrů v regresním modelu se pak rozhoduje o zahrnutí parametrů do modelu. Výhodou tohoto přístupu je jeho menší časová náročnost proti situaci, kdy vycházíme z plného modelu s odlišnými parametry pro všechna období. Tento přístup také využijeme v praktické ukázce v páté kapitole. Opačný postup hledání vhodného modelu popisující pozorovaná

data, popsaný například v [10], je přechod od plného modelu vysvětlovaného z nejkratších možných intervalů jejich postupným sdružováním a vypouštěním nevýznamných regresorů k výslednému modelu. Výhodou je menší riziko vynechání některého významného parametru.

Musíme zde také zmínit, že v praxi občas nastává situace, kdy vysvětlující interval nelze průkazně vysledovat přímo z dat. Takové proměnné se určují spíše díky detailnější znalosti daného pojistného odvětví a zákonitostí v něm obsažených. Příkladem může být indikátor sezónnosti určující pravidelné fluktuace (pokles, nárůst) škodních událostí či indikátor popisující účinnost nových legislativních změn a jejich promítnutí do vývoje dat. Pokud je pak takový parametr statisticky významný, bylo by chybou ho nezahrnout do výsledného modelu.

Nakonec dle výše uvedeného předpokládáme normální rozdělení reziduí, přičemž je nezbytné splnění tohoto předpokladu ověřit vhodnými statistickými testy.

4.4 Odhady budoucích hodnot

4.4.1 Stabilita modelu

Jedním z hlavním požadavků na výsledný model je jeho schopnost stabilně předpovídat budoucí hodnoty. Způsob testování tohoto požadavku ukazuje Zehnwirth v článku [10]. Využívá v něm techniku vynechání poslední diagonály ve vývojovém trojúhelníku. Testování pak spočívá ve vytvoření testovacího modelu z dat, která neobsahují hodnoty pozorované v posledním či několika posledních kalendářních obdobích. Pokud je výsledný model korektní, neměly by se predikované hodnoty vytvořené testovacím modelem příliš lišit od vynechaných pozorovaných hodnot. To také znamená, že i odhady parametrů testovacího modelu by měly být podobné odhadům parametrů výsledného modelu.

4.4.2 Predikce hodnot

Nyní jsme v situaci, kdy máme vytvořený regresní model a potřebujeme doplnit přírůstkový vývojový trojúhelník (Tabulka 4.1) na čtverec.

Zaveďme náhodný vektor $\mathbf{Y}_F$ budoucích pozorování ve tvaru

$$\mathbf{Y}_F = \mathbf{X}_F\hat{\beta} + \varepsilon, \tag{4.4.1}$$

kde matice $\mathbf{X}_F$ je vytvořená analogicky jako matice $\mathbf{X}$.

Pak platí

$$\mathbb{E}[\mathbf{Y}_F] = \mathbf{X}_F\beta, \tag{4.4.2}$$

$$\mathrm{Var}[\mathbf{Y}_F] = \mathbf{X}_F\mathrm{Var}[\hat{\beta}]\mathbf{X}_F^T + \sigma^2\mathbf{I}. \tag{4.4.3}$$

Vidíme, že variabilitu můžeme rozložit na dvě složky - chybu modelu a náhodnou chybu. Zehnwirth v [10] ukazuje, že každá odhadovaná buňka má normální rozdělení se střední hodnotou $Y_{F,i,k}$ a rozptylem $\mathrm{Var}[Y_{F,i,k}]$. Dále náhodný vektor $\mathbf{Y}_F$ má vícerozměrné rozdělení s vektorem středních hodnot $\mathbb{E}[\mathbf{Y}_F]$ a rozptylovou maticí $\mathrm{Var}[\mathbf{Y}_F]$. Znalost vícerozměrného rozdělení umožňuje doplnit vývojový trojúhelník na čtverec při respektování korelační struktury.

Odtud již můžeme získat odhad celkové rezervy R. K určení odhadu jejích jednotlivých kvantilů lze využít dostatečného počtu simulací ze zmíněného vícerozměrného normálního rozdělení, jak navrhuje Zehnwirth v [10].

Kapitola 5

Modely Chain ladder a PTF v praxi

K praktické ukázce a porovnání obou zmíněných přístupů jsme si vybrali ilustrativní data o vyplaceném pojistném plnění v pojištění odpovědnosti z provozu motorových vozidel. Zákon 168/1999 Sb. o pojištění odpovědnosti za škodu způsobenou provozem vozidla umožnil liberalizaci českého trhu s tzv. povinným ručením od počátku roku 2000. Získané data obsahují ilustrativní údaje o vyplaceném pojistném plnění vyplaceném od 1. ledna 2000 do 30. června 2008. Za jednotku období jsme zvolili 1 pololetí, což nám dává celkem 17 období ($n = 1, \ldots, 17$) a zpoždění nabývá hodnot $k = 0, \ldots, 16$.

Výchozí data najdeme ve formě přírůstkového trojúhelníku v Dodatku B, tabulka B.1.

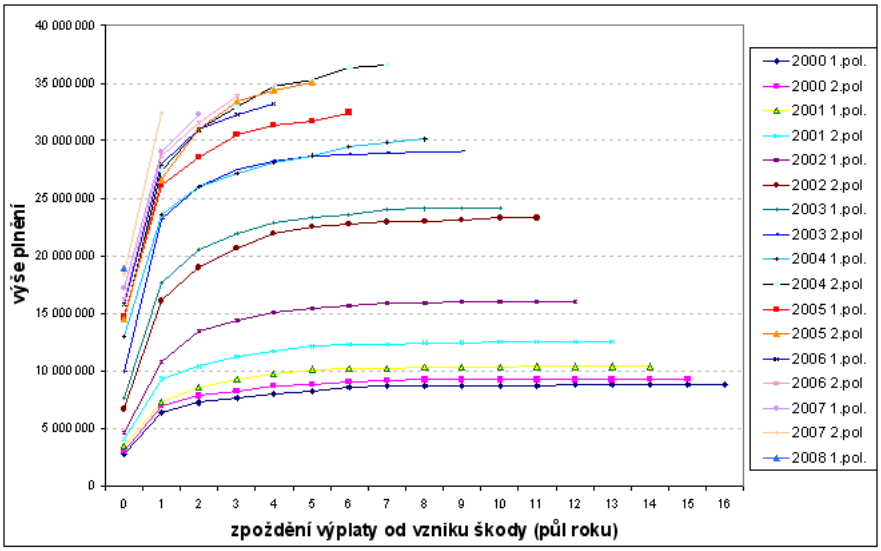
5.1 Aplikace Chain ladder

Začneme stochastickou metodou Chain ladder, přičemž budeme uvažovat kumulativní trojúhelník B.2. Na Obrázku 5.1 je zobrazen vývoj plnění v závislosti na pololetí vývoje pro každé pololetí vzniku. Při prvním pohledu vidíme, že vývoj plnění se v jednotlivých pololetích nijak zásadně neodlišuje s výjimkou narůstajících finančních objemů.

5.1.1 Výsledky CL

Aplikujme tedy metodu Chain ladder na pozorovaná data. V prvním kroku spočítáme odhady vývojových faktorů a to včetně individuálních vývojových

faktorů, viz Tabulka 5.1.



Obrázek 5.1: Run-off pojistného plnění

Dále můžeme podle vzorce (2.1.1) vypočítat bodové odhady konečného pojistného plnění $C_{i,16}$ na konci 16-tého pololetí vývoje a také rezervy R_i pro pololetí vzniku $i = 2, \ldots, 17$, viz Tabulka 5.2.

Odtud vychází odhad celkové rezervy na pojistná plnění ve výši

$$R = 56\ 270\ 543.$$

Tento výsledek by nám přinesla klasická deterministická verze Chain ladder. Jak jsme již naznačili ve druhé kapitole, pravděpodobnost toho, že skutečná výše pojistného plnění bude rovna tomuto bodovému odhadu, je v případě spojitého rozdělení nulová. V dalším nejprve uvedeme výsledky stochastické verze Chain ladder a poté budeme diskutovat splnění předpokladů.

V Tabulce 5.3 nalezneme vedle odhadu rezerv R_i také odhady střední kvadratické chyby $mse(\hat{R}_i)$, směrodatné odchylky $s.e.(\hat{R}_i)$ a jejich podíl k $\hat{R}_i$. Pro pololetí vzniku $i = 2$ je podíl roven 113% a pro $i = 3$ je roven 50%. Proto při odhadu rezerv budeme postupovat v souladu s doporučením uvedeném v závěru druhé kapitoly vycházející z článku [7] a namísto normálního rozdělení rezerv zvolíme u těchto období logaritmicko-normální rozdělení.

V dodatku B v Tabulce B.3 nalezneme kompletní výsledky pro všechna uvažovaná období vzniku. Zde uvedeme pouze odhad rezervy $\hat{R} = 56\ 270\ 543$

$f_{i,k}$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	2,287	1,149	1,061	1,037	1,037	1,032	1,011	1,003	1,003	1,007	1,005	1,002	1,000	1,001	1,001	1,000
2	2,278	1,137	1,052	1,042	1,024	1,030	1,010	1,002	1,003	1,001	1,000	1,000	1,000	1,002	1,001	
3	2,112	1,169	1,085	1,043	1,043	1,011	1,002	1,004	1,001	1,002	1,006	1,003	1,001	1,001		
4	2,289	1,123	1,074	1,049	1,028	1,017	1,001	1,014	1,001	1,002	1,001	1,001	1,000			
5	2,355	1,238	1,072	1,049	1,025	1,012	1,014	1,001	1,004	1,001	1,001	1,001				
6	2,407	1,180	1,087	1,064	1,025	1,012	1,007	1,004	1,004	1,006	1,002					
7	2,288	1,163	1,071	1,039	1,022	1,012	1,015	1,007	1,001	1,001						
8	2,335	1,121	1,058	1,027	1,013	1,004	1,005	1,002	1,001							
9	1,799	1,104	1,046	1,034	1,022	1,027	1,009	1,013								
10	1,897	1,129	1,065	1,050	1,020	1,028	1,008									
11	1,769	1,096	1,071	1,026	1,011	1,021										
12	1,837	1,165	1,076	1,030	1,020											
13	1,775	1,106	1,042	1,028												
14	1,772	1,106	1,073													
15	1,690	1,111														
16	1,759															
f_k	1,917	1,131	1,066	1,038	1,021	1,018	1,008	1,006	1,002	1,003	1,002	1,001	1,001	1,001	1,001	1,000

Tabulka 5.1: Odhady vývojových faktorů

i	2	3	4	5	6	7	8	9
$C_{i,16}$	9289385	10445184	12567889	16066488	23369384	24341088	29279555	30494788
R_i	4030	15646	36670	55559	113958	173665	284100	363673

i	10	11	12	13	14	15	16	17
$C_{i,16}$	37227960	33244767	36666186	35386992	37579678	38065148	43120064	48451182
R_i	661149	863025	1598885	2248910	3667745	5830086	10828412	29525029

Tabulka 5.2: Odhady konečného pojistného plnění a rezerv

a odhad směrodatné odchylky $\widehat{s.e.(\hat{R})} = 5\ 313\ 174$. Budeme-li dále uvažovat interval spolehlivosti 95% (tj. $\alpha = 0,05$), pak dle vzorce (2.2.15) spočteme intervalový odhad R s dolním limitem 50 158 537 a horním limitem 62 382 549. Výsledek pak říká, že celkové pojistné plnění pro uvažovaná období vzniku se s pravděpodobností 95% dále navýší o 50 158 537 až 62 382 549, což je jistě cennější údaj, než pouze bodový odhad střední hodnoty. Nesmíme zapomenout zmínit, že vývoj pojistného plnění považujeme po 8 a půl letech za ukončený.

5.1.2 Statistické testování předpokladů

V druhé kapitole jsme uvedli tři předpoklady, které musí být splněny, aby mohla být stochastická verze metody Chain ladder použita. Byly to:

1. existence vývojových faktorů nezávislých na období vzniku (2.2.1),

2. nezávislost škodních úhrnů pro různá období vzniku (2.2.3),

i	$\hat{R}_i$	$\widehat{mse(\hat{R}_i)}$	$\widehat{s.e.(\hat{R}_i)}$	$\widehat{s.e.(\hat{R}_i)}/\hat{R}_i$
2	4030	20667858	4546	113%
3	15646	60602357	7785	50%
4	36670	199172182	14113	38%
5	55559	305594978	17481	31%
6	113958	989803725	31461	28%
7	173665	3049028315	55218	32%
8	284100	7247378915	85132	30%
9	363673	9031576655	95035	26%
10	661149	31935072465	178704	27%
11	863025	40664543404	201655	23%
12	1598885	119222591563	345286	22%
13	2248910	161519480036	401895	18%
14	3667745	300483544059	548164	15%
15	5830086	474688713075	688977	12%
16	10828412	1705242105138	1305849	12%
17	29525029	25375155256211	5037376	17%
Celkem	56270543	28229815130937	5313174	9%

Tabulka 5.3: Výsledky stochastické verze Chain ladder

3. přímá úměra rozptylu následujících škodních úhrnů vzhledem k předcházejícímu období vývoje ve stejném roce vzniku (2.2.5).

Ověření existence vývojových faktorů

Jak ukazuje Mack v [7], splnění předpokladu (2.2.1) implikuje nekorelovanost po sobě následujících individuálních vývojových faktorů v jednom období vzniku. Pokud bychom tedy prokázali jejich nenulovou korelaci, předpoklad existence nezávislých vývojových faktorů by nebyl splněn.

Pro testování nekorelovanosti využijeme, jak ukazuje Mack v [7], Spearmanův korelační koeficient, který je definován

$$T_k = 1 - 6 \sum_{i=1}^{n-k} \frac{(r_{i,k} - s_{i,k})^2}{(n-k)((n-k)^2 - 1)},$$

kde $r_{i,k}$ je pořadí $C_{i,k+1}/C_{i,k}$ v k. sloupci a $s_{i,k}$ je pořadí $C_{i,k}/C_{i,k-1}$ v $k-1$. sloupci.

Při testování předpokladu nekorelovanosti, který při normalitě odpovídá nezávislosti, se pomocí Spearmanova korelačního koeficientu ověřuje platnost nulové hypotézy o nekorelovanosti následujících individuálních faktorů,

při jejímž splnění platí

$$\mathbb{E}[T_k] = 0 \text{ a } \mathrm{Var}[T_k] = \frac{1}{n-k-1}.$$

Výsledky tohoto testu jsou uvedeny v Tabulkách B.4, B.5 a B.6. V poslední tabulce je také indikováno, že pro zpoždění $k = 1,\ 2,\ 8$ a 11 zamítáme nulovou hypotézu na hladině 5%. Proto nulovou hypotézu vyjádřenou v předpokladu (2.2.1) zamítáme.

Na problém se můžeme podívat také způsobem, že existence stejných faktorů v jednom období zpoždění pro různá období vzniku znamená existenci regresní přímky s lineárním koeficientem f_k bez absolutního členu, čili

$$Y_i = C_{i,k+1} = f_k C_{i,k} + \varepsilon_i,$$

kde $\mathbb{E}[\varepsilon_i] = 0$ a $\mathrm{Var}[\varepsilon_i] = \sigma^2 C_{i,k}$.

Na Obrázku B.1 zobrazující graf porovnávající $C_{i,1}$ proti $C_{i,0}$ je zřejmě vidět, že odchylky od regresní přímky (červená přímka) nejsou náhodné a vykazují klesající tendenci. Tento fakt podporuje také porovnání vážených residuí (tj. $(C_{i,k+1} - C_{i,k}\hat{f}_k)/\sqrt{C_{i,k}}$) na Obrázku B.2. Znamená to, že odhady pojistného plnění v dřívějších letech, a tedy nižší hodnoty, jsou podhodnoceny a naopak odhady pro novější data (vyšší hodnoty) jsou nadhodnoceny. Proto nemůže být splněn předpoklad existence jednotného vývojového faktoru mezi nultým a prvním vývojovým obdobím. Obdobné výsledky přináší i porovnání pro následující vývojové období, viz Obrázek B.3 a Obrázek B.4.

Ověření nezávislosti období vzniku

Nezávislost jednotlivých období může být porušena nějakou akcí s významným jednorázovým efektem, např. změna způsobu likvidace, změna zákona či výrazná změna inflace. Jedná se o změnu, která proběhne v určitém kalendářním období t a promítne se do všech období vzniku i. V následujícím textu proto prověříme, zda k takovým změnám v našich datech mohlo dojít.

Stručně si zde popíšeme testovací proceduru popsanou v [7]. Označme diagonály vývojového trojúhelníku

$$D_j = \{C_{j,0}, C_{j-1,1}, \ldots, C_{2,j-2}, C_{1,j-1}\}, \qquad 1 \le j \le n,$$

a diagonální množiny vývojových faktorů

$$A_j = \{C_{j,1}/C_{j,0}, C_{j-1,2}/C_{j-1,1}, \ldots, C_{1,j}/C_{1,j-1}\}$$

a

$$A_{j-1} = \{C_{j-1,1}/C_{j-1,0}, C_{j-2,2}/C_{j-2,1}, \ldots, C_{1,j-1}/C_{1,j-2}\}.$$

Pokud by v datech byl zaznamenán nějaký kalendářní efekt, potom by bylo D_j větší (menší) než obvykle a zároveň by bylo A_j větší (menší) a A_{j-1} menší (větší) než obvykle.

Dále pozorované indiviuální faktory pro každé období vývoje uspořádejme do množin $F_k = \{C_{i,k+1}/C_{i,k} | 1 \leq i \leq n-k\}$ a ty rozdělme na množiny LF_k obsahující hodnoty F_k větší než median(F_k) a SF_k obsahující hodnoty F_k menší než median(F_k). Hodnoty F_k, které jsou rovny mediánu F_k ponecháme zvlášť. Každý prvek z množiny F_k má 50%-ní pravděpodobnost být buď v LF_k nebo v SF_k. Pro každou diagonálu označme L_j (S_j) počet prvků větších (menších) než příslušný medián individuálních vývojových faktorů. Přirozeně by pak čísla L_j a S_j měla být přibližně stejně veliká, což by odpovídalo náhodné fluktuaci. Pokud by ale bylo

$$Z_j = min(L_j, S_j)$$

významně větší či menší než $(L_j + S_j)/2$, pak je v datech obsažen zřejmě nějaký silný kalendářní efekt.

Označme $n_j = L_j + S_j$ a $m = \lfloor (n_j - 1)/2 \rfloor$. Pak

$$\mathbb{E}[Z_j] = \frac{n_j}{2} - \binom{n_j}{2} \frac{n_j}{2^{n_j}},$$

$$\mathrm{Var}[Z_j] = \frac{n_j(n_j-1)}{4} - \binom{n_j - 1}{m_j} \frac{n_j(n_j-1)}{2^{n_j}} + \mathbb{E}[Z_j] - (\mathbb{E}[Y_j])^2.$$

Aby při testování každého Z_j odděleně nedošlo ke kumulaci náhodné chyby, doporučuje Mack v článku [7] uvažovat veličinu

$$Z = Z_2 + Z_3 + \ldots + Z_{n-2},$$

ve které jsme neuvažovali Z_1, která je veličinou s nulovým rozptylem, protože může nabývat pouze jedné hodnoty. Protože za platnosti nulové hypotézy jsou veličiny Z_j nekorelované, dostáváme

$$\mathbb{E}[Z] = \mathbb{E}[Z_2] + \mathbb{E}[Z_3] + \ldots + \mathbb{E}[Z_{n-2}],$$

$$\mathrm{Var}[Z] = \mathrm{Var}[Z_2] + \mathrm{Var}[Z_3] + \ldots + \mathrm{Var}[Z_{n-2}],$$

přičemž můžeme předpokládat, že Z má aproximativně normální rozdělení.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	L	L	S	S	L	L	L	S	L	L	L	L	L	M	L	M
2	L	L	S	L	L	L	L	S	L	S	S	S	S	L	S	
3	L	L	L	L	L	S	S	L	S	L	L	L	L	S		
4	L	S	L	L	L	M	S	L	S	M	S	S	S			
5	L	L	L	L	L	S	L	S	L	S	S	M				
6	L	L	L	L	L	S	S	M	L	L	L					
7	L	L	L	M	S	S	L	L	S	S						
8	L	S	S	S	S	S	S	S	S							
9	S	S	S	S	S	L	L	L								
10	S	M	S	L	S	L	S									
11	S	S	S	S	S	L										
12	S	L	L	S	S											
13	S	S	S	S												
14	S	S	L													
15	S	S														
16	S															
17																

Tabulka 5.4: Test nezávislosti roků období

Pak nulovou hypotézu zamítneme, jestliže

$$|Z - \mathbb{E}[Z]| \geq \sqrt{\mathrm{Var}[Z]}.$$

Aplikujeme-li tuto proceduru na náš příklad, získáme v prvním kroku trojúhelník obsahující roztřídění vývojových faktorů do skupin dle velikosti, viz Tabulka 5.4, kde L značí větší hodnotu než medián F_k, S označuje menší hodnotu než medián F_k a $M = \mathrm{median}(F_k)$.

Podle výše zmíněného postupu vypočítáme další požadované hodnoty, viz Tabulka 5.5.

Pak testová statistika Z nabývá hodnoty 39, což je mimo interval tabelovaných hodnot $(40,001,\ 53,338)$, v němž nulovou hypotézu na hladině 5% nezamítáme. Výsledkem je tedy zamítnutí nulové hypotézy a nesplnění dalšího předpokladu metody Chain ladder.

j	L_j	S_j	M_j	Z_j	n_j	m_j	$\mathbb{E}[Z_j]$	$\mathrm{Var}[Z_j]$
2	2	0	0	0	2	0,5	0,50	0,2500
3	2	1	0	1	3	1,0	0,75	0,1875
4	2	2	0	2	4	1,5	1,25	0,4375
5	4	1	0	1	5	2,0	1,56	0,3711
6	6	0	0	0	6	2,5	2,06	0,6211
7	7	0	0	0	7	3,0	2,41	0,5537
8	6	2	0	2	8	3,5	2,91	0,8037
9	4	4	1	4	8	3,5	2,91	0,8037
10	4	5	1	4	9	4,0	3,27	0,7359
11	3	7	1	3	10	4,5	3,77	0,9859
12	2	10	0	2	12	5,5	4,65	1,1680
13	6	5	2	5	11	5,0	4,15	0,9180
14	5	8	1	5	13	6,0	5,03	1,0999
15	6	9	0	6	15	7,0	5,93	1,2818
16	4	10	2	4	14	6,5	5,53	1,3499
Celkem	63	64	8	39	127	56,0	46,67	11,5679

Tabulka 5.5: Test nezávislosti roků období

Ověření přímé úměry rozptylu

Poslední předpoklad, jak uvádí Mack v [3], již stačí v tomto případě ověřit pouze pomocí grafu zobrazující vážená rezidua v závislosti na $C_{i,k}$. Pro $k = 1$ a $k = 2$ jsou tyto grafy znázorněny na Obrázku B.2, resp. Obrázku B.4. Tato residua by měla být soustředěna náhodně okolo osy x, což na první pohled není splněno. Proto ani poslední předpoklad metody CL o linearitě rozptylu pojistného plnění není splněn.

Závěr aplikace metody Chain ladder

Výsledkem aplikace metody Chain ladder na získaná data je výpočet rezervy podle odvozených vzorců, nicméně použití této metody je sporné vzhledem k nesplnění jejích předpokladů, jak i v případě neexistence požadavků na pravděpodobnostní rozdělení ukazují provedené statistické testy.

5.2 Aplikace PTF

Nyní se zaměříme na přístup výpočtu rezervy R z pohledu PTF modelů. Jak jsme již uvedli ve čtvrté kapitole, budeme vycházet z logaritmovaných dat přírůstkového trojúhelníku. V praxi se obecný PTF model (4.1.4) standardně rozšiřuje o absolutní člen μ *(intercept)*, čemuž odpovídá rovnice

$$Y_{i,k} = \mu + \alpha_i + \sum_{j=1}^{k} \gamma_j + \sum_{t=2}^{i+k} \iota_t + \varepsilon_{i,k}. \tag{5.2.1}$$

5.2.1 Tvorba modelu

Při vytváření výsledného modelu jsme začínali od nejjednoduššího modelu obsahující pouze počátek, parametr vývojového období ($\gamma = \gamma_k$, $k = 0, ..., 16$) a jeden parametr období vzniku ($\alpha = \alpha_i$, $i = 1, \ldots, 17$). První krok přinesl následující výsledky:

$$\begin{aligned} \mu &= 14,82094 \\ \alpha &= 0,11701 \\ \gamma &= -0,49585 \\ R^2 &= 91,82655\% \\ \hat{\sigma} &= 0,69783 \end{aligned}$$

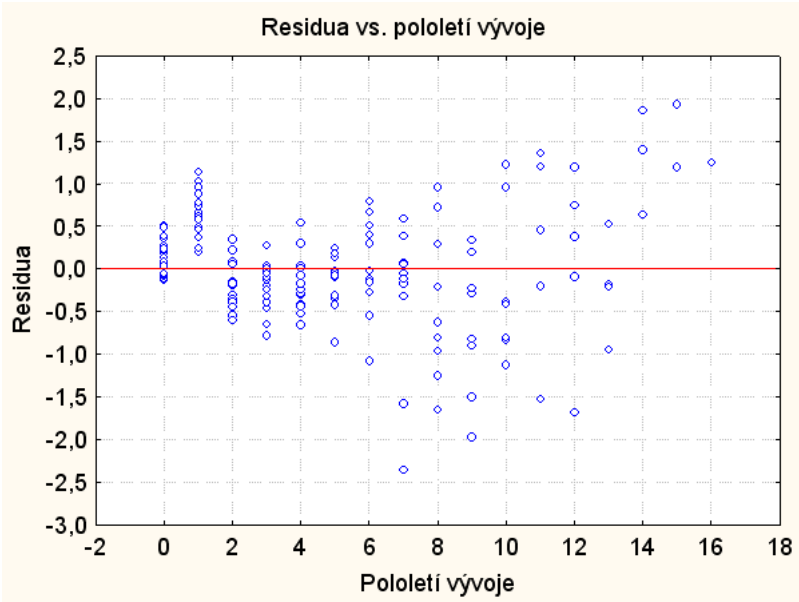
Avšak více než tyto konkrétní hodnoty je na začátku tvorby modelu užitečnější zkoumat residua v závislosti na pololetí vývoje (Obrázek 5.2), vzniku a kalendářním pololetí (Obrázek 5.3).

Na Obrázku 5.2 ihned vidíme, že residua se nevyskytují náhodně kolem nuly, protože např. v prvních dvou pololetích vývoje jsou vyrovnané hodnoty podhodnoceny a naopak následujících třech obdobích mírně nadhodnoceny. Proto musíme dále model upřesnit přidáním dalších parametrů.

Stejně tak na Obrázku 5.3 vidíme že je mírně porušena náhodnost residuí. Nakonec je nutné prozkoumat podobným způsobem residua v závislosti na pololetí vzniku.

V tomto okamžiku máme první diagnostiku modelu za sebou. Dále tento model upřesníme na základě zjištěných skutečností v grafech residuí. Výsledkem následného postupu je model, který obsahuje počáteční hodnotu a dalších 8 parametrů, viz Tabulka 5.6.

Pokud uvažujeme 5%-ní hladinu testu významnosti regresních parametrů, pak bychom za nevýznamné parametry považovali 4. a 9. parametr. Podle



Obrázek 5.2: Residua v závislosti na pololetí vývoje

grafů residuí (Obrázek B.5 a Obrázek B.6) se lze domnívat, že by však tyto parametry mohly patřit do výsledného modelu. V této situaci můžeme zkusit využít přístup, který navrhuje Zehnwirth v [10]. Jedná se o odstranění odlehlých pozorování, která na základě regresní diagnostiky výrazně neodpovídají modelu.

Postup zjištění odlehlých hodnot je následující:

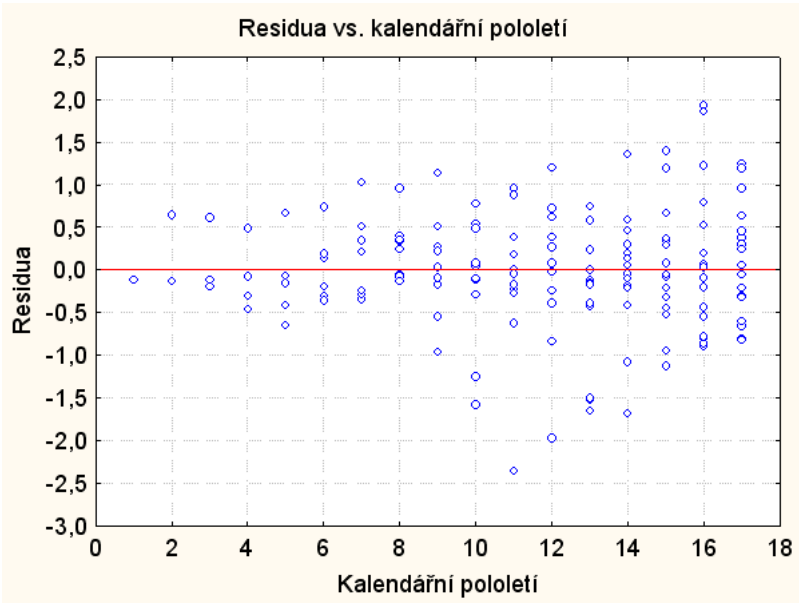
1. seřadíme data podle velikosti jejich standardizovaných residuí v absolutní hodnotě,

2. určíme empirický dolní (LQ) a horní kvartil (UQ) a spočítáme jejich rozpětí (MS),

3. spočítáme horní mez pro odlehlé hodnoty UB = UQ + 1,5.MS,

4. všechny hodnoty, které mají residuum v absolutní hodnotě větší než UB jsou odlehlá pozorování.

V našem příkladě jsme podle tohoto postupu odstranili hodnoty, které měly standardizované residuum v absolutní hodnotě větší než 1,6. Po jejich odstranění jsme znovu vytvořili model s původně navrženými proměnnými. Nový model již nulovost 4. a 9. regresního parametru na hladině 5% zamítá, jak lze vidět v Tabulce 5.7.



Obrázek 5.3: Residua v závislosti na kalendářním pololetí

Pořadí	Parametr	Hodnota	p-hodnota	Koeficient v období$[i, k]$
1.	μ	$14,255000$	$0,000000$	μ
2.	α	$0,052970$	$0,024146$	$i\,\alpha$
3.	γ	$-0,280820$	$0,000000$	$k\,\gamma$
4.	vývoj 0-1	$0,264560$	$0,074078$	$\min(k+1,2)$
5.	vývoj 2-3	$-0,512050$	$0,000775$	$\max(\min(k-1,2),0)$
6.	vývoj 6-8	$-0,446110$	$0,000001$	$\max(\min(k-5,3),0)$
7.	α_1	$0,458210$	$0,007565$	$\chi_{[i=1]}$
8.	vznik 4-6	$0,243020$	$0,000369$	$\max(\min(i-3,6),0)$
9.	čas 1-7	$0,073980$	$0,239420$	$\min(i+k,7)$

Tabulka 5.6: Výsledný model před odstraněním odlehlých pozorování

Pořadí	Parametr	Hodnota	p-hodnota	Koeficient v období$[i,k]$
1.	μ	$14,13151$	$0,000000$	μ
2.	α	$0,04651$	$0,001319$	$i\,\alpha$
3.	γ	$-0,30206$	$0,000000$	$k\,\gamma$
4.	vývoj 0-1	$0,26987$	$0,002881$	$\min(k+1,2)$
5.	vývoj 2-3	$-0,49866$	$0,000000$	$\max(\min(k-1,2),0)$
6.	vývoj 6-8	$-0,39962$	$0,000000$	$\max(\min(k-5,3),0)$
7.	α_1	$0,39733$	$0,000363$	$\chi_{[i=1]}$
8.	vznik 4-6	$0,18203$	$0,000043$	$\max(\min(i-3,6),0)$
9.	čas 1-7	$0,12878$	$0,001039$	$\min(i+k,7)$

Tabulka 5.7: Výsledný model po odstranění odlehlých pozorování

5.2.2 Statistické vlastnosti modelu

Dále se zaměříme na statistické vlastnosti tohoto modelu. V Tabulce 5.8 uvádíme některé statistické veličiny tohoto modelu.

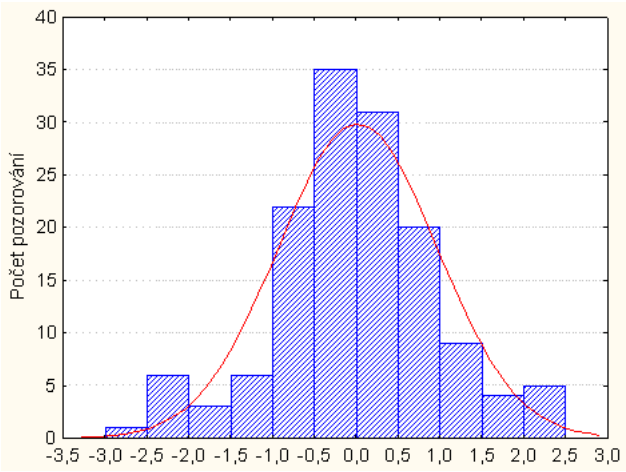
$\hat{\sigma}$	$0,339110$
R^2	$97,858987\%$
$\text{Durbin}-\text{Watson}$	$1,894512$

Tabulka 5.8: Statistické vlastnosti výsledného modelu

Tabulka 5.8 ukazuje, že výsledný model vykazuje nízký rozptyl, má vysokou vysvětlovací schopnost a podle Durbinovy-Watsonovy statistiky nelze zamítnout nekorelovanost residuí, jelikož kritická hodnota testu je 1,84573. Zde je však nutné si uvědomit, že při výpočtu testové statistiky v Durbinově-Watsonově testu se předpokládá seřazení hodnot nějakým přirozeným způsobem. V tomto případě máme seřazeny hodnoty nejprve podle období vzniku a dále podle období vývoje. Tento test tedy chápeme jako podpůrný nástroj určující kvalitu modelu, kdy naznačené seřazení má smysl pro každý řádek.

K testování normálního rozdělení residuí jsme použili následující statistické techniky:

- **Kolmogorův-Smirnonův test**
  Tento test (viz dodatek A) byl pro nás základním prostředkem pro ověření existence normality residuí. Testová statistika vychází



Obrázek 5.4: Histogram residuí

$K_n = 0,0755$ s p-hodnotou $0,2$. Proto na testové hladině 5% normalitu residuí nezamítáme.

- **histogram residuí**
  Histogram na Obrázku 5.4 je pro nás pouze pomocným nástrojem, který lze použít k předběžnému ověření normality. I podle něj nelze však jednoznačně normalitu zamítnout.

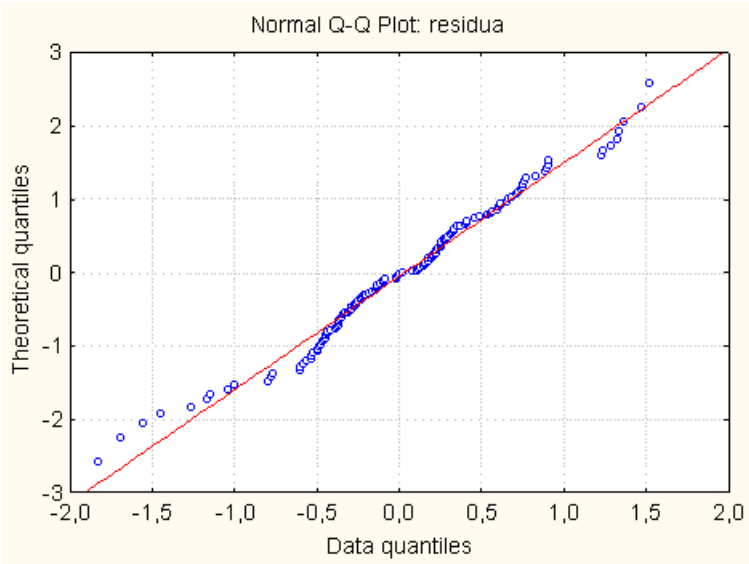
- **Q-Q plot**
  Graf porovnává kvantily empirického rozdělení vycházející ze zkoumaných dat s kvantily normálního rozdělení N(0,1). V případě, že by empirické rozdělení bylo normální, byly by body tohoto grafu na přímce. Na Obrázku 5.5 vidíme, že body se od přímky příliš neliší a normální rozdělení dat proto nelze zamítnout.

Protože ani jedna z uvedených technik normalitu nezamítá, lze předpokládat splnění normality residuí.

Dalším testem, který jsme na modelu provedli, je orientační test multikolinearity navržený v [4], který poslouží taktéž jako podpůrný nástroj pro určení kvality modelu. Tento test je založen na splnění nerovností

$$\rho_j \geq R,$$

kde R^2 je koeficient determinace a $\rho_j, j = 2, \ldots, 9$, jsou koeficienty mnohonásobné korelace jednotlivých vysvětlujících proměnných, $j = 2, \ldots, 9$, se všemi ostatními proměnnými.



Obrázek 5.5: Q-Q plot

Jelikože $R = 0,989237014$, je z Tabulky 5.9 patrné, že na základě tohoto testu nezamítneme neexistenci multikolinearity.

Parametr	ρ_j	Nekorelovanost zamítnuta?
2.	$0,759149$	NE
3.	$0,910844$	NE
4.	$0,769215$	NE
5.	$0,864936$	NE
6.	$0,845605$	NE
7.	$0,306373$	NE
8.	$0,739074$	NE
9.	$0,547489$	NE

Tabulka 5.9: Statistické vlastnosti výsledného modelu

Pohledem na grafy residuí v závislosti na pololetí vzniku (Obrázek B.7), vývoje (Obrázek B.8) a kalendářním pololetí (Obrázek B.9) nelze zamítnout, že by se rozptyl residuí nějakým výrazným způsobem vyvíjel a lze proto přepokládat konstantní rozptyl residuí.

Posledním úkolem pro ověření korektnosti modelu bylo testování stability modelu, kterou jsme provedli ve dvou krocích odebráním dat nejprve z posledního kalendářního období ($t = 17$) a následně odebráním předposledního

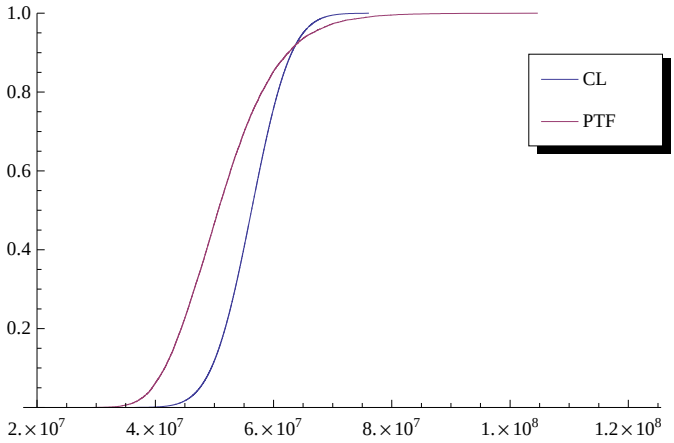
pololetí ($t = 16$). Po odebrání dat jsme vytvořili dva analogické modely a zkoumali jsme změnu odhadů jejích parametrů. Výsledky tohoto testu jsou zobrazeny v Tabulce 5.10, ze které vyplývá, že po odebrání jednoho i dvou posledních období se odhady parametrů od původního modelu liší pouze nepatrně a model je tudíž schopen dobře predikovat budoucí neznámé hodnoty.

	$t \leq 17$	$t \leq 16$	$t \leq 15$
1.	$14,13150$	$14,17900$	$14,2398$
2.	$0,04651$	$0,04363$	$0,05933$
3.	$-0,30206$	$-0,29091$	$-0,30429$
4.	$0,26987$	$0,26016$	$0,29477$
5.	$-0,49866$	$-0,52695$	$-0,43391$
6.	$-0,39962$	$-0,40105$	$-0,34318$
7.	$0,39733$	$0,39488$	$0,33749$
8.	$0,18203$	$0,20345$	$0,21055$
9.	$0,12878$	$0,11856$	$0,07824$

Tabulka 5.10: Testování stability modelu

5.2.3 Odhad rezervy PTF modelem

Výsledné rozdělení rezervy jsme odhadli na základě provedení dostatečného množství simulací z vícerozměrného normálního rozdělení s odhadnutými parametry, tak jak je doporučeno v [5]. Tímto způsobem jsme schopni stanovit empirické kvantily tohoto rozdělení. Simulace byly provedeny v softwaru Mathematica. Výsledkem simulací je odhad střední hodnoty rezervy R ve výši 51 511 385 se směrodatnou odchylkou 8 413 533. Pro úplnost dodejme, že analytickým výpočtem odhadu střední hodnoty je odhad rezervy ve výši 51 196 502. Bohužel, získané hodnoty nemůžeme nijak více využít např. k odhadům kvantilů R, jelikož neznáme další podrobnosti o pravděpodobnostním rozdělení. Proto při určování odhadu rezerv na dané hladině spolehlivosti budeme vycházet z empirických kvantilů, které by však díky velkému počtu simulací měly být dostatečně věrohodné. Např. na hladině spolehlivosti 95% vychází odhad rezervy ve výši 66 538 473.



Obrázek 5.6: Porovnání empirických distribučních funkcí

5.3 Srovnání výsledků

Pokusme se nyní porovnat získané výsledky obou metod pro výpočet rezerv na pojistná plnění.

Na Obrázku 5.6 můžeme vidět porovnání distribučních funkcí obou metod, ze kterého vyplývá riziko nedostatečné výše rezervy R podle CL v případě korektnosti PTF modelu.

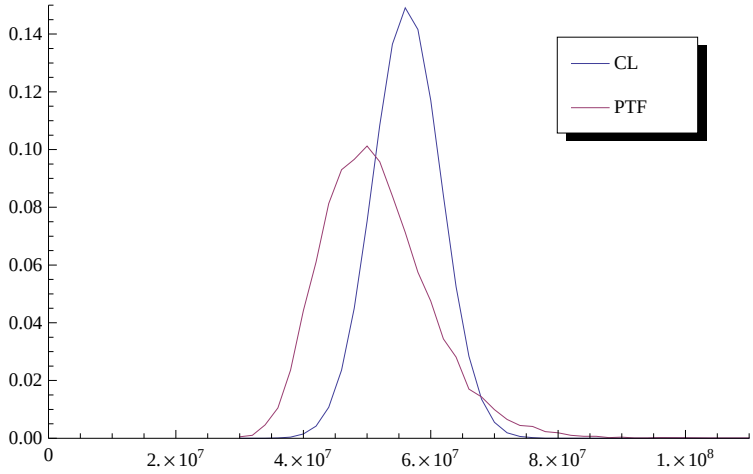
S tím souvisí i porovnání empirických hustot, viz Obrázek 5.7, díky nimž můžeme zřetelně pozorovat následující rozdíly:

1. metoda Chain ladder přináší symetrické rozdělení s nižším rozptylem,

2. metoda PTF má rozdělení zešikmené doprava, přičemž pro vyšší hodnoty vykazuje vyšší pravděpodobnost než metoda Chain ladder.

Nakonec v Tabulce 5.11 prezentujeme pro srovnání sledované statistické veličiny vycházející z obou metod.

Z provedených statistických testů můžeme usoudit, že preferování modelu PTF oproti metodě Chain ladder je v tomto případě statisticky oprávněné pro zamítnutí některých předpokladů metody Chain ladder.



Obrázek 5.7: Porovnání empirických hustot

veličina	PTF	CL
$\hat{\sigma}$	8413533	5313174
$\hat{R}$	51511385	56270543
$u_{0,500}$	50633889	56270543
$u_{0,750}$	56424593	59854224
$u_{0,900}$	62513499	63079649
$u_{0,950}$	66538473	65009936
$u_{0,975}$	70359125	66684172
$u_{0,990}$	75874730	68630833

Tabulka 5.11: Srovnání výsledků CL a PTF

Závěr

V této práci jsme po počátečním připomenutí dvou používaných metod k odhadu rezervy na pojistná plnění diskutovali jejich statistické vlastnosti a možnosti používané k testování splnění jednotlivých předpokladů.

V praktické situaci často nastává situace, která se ukázala i ve zpracovaném příkladě, že např. podkladová data vykazují významné trendy v kalendářním období, což způsobí, že statistické testy navržené a používané pro metodu Chain Ladder naznačují zamítnutí klíčových předpokladů této metody. V našem případě, ale i v dalších obdobných situacích, není proto metoda Chain Ladder vhodná.

Naproti tomu i v těchto příkladech jsou na první pohled širší předpoklady na modely typu PTF, kdy především tento model nelze použít pro libovolné rozdělení a je nutné explicitně otestovat splnění normality, realističtější, neboť umožňují zahrnout veškeré možné trendy a skokové změny ve vývojovém schématu.

Další klíčovou předností PTF modelů je malý počet vysvětlujících proměnných, na jejichž základě lze vývoj ve všech časových směrech vývojového trojúhelníku modelovat. Tato skutečnost je klíčová pro stabilitu predikcí a vůbec pro možnost realistického statistického testování kvality modelů tohoto typu.

Na druhé straně existují i nevýhody použití PTF modelů, mezi které patří především vyšší časová náročnost tvorby modelu spočívající především v postupných krocích hledání modelu.

Literatura

[1] Anděl, J.: *Statistické metody*, MATFYZPRESS, Praha 2007,

[2] Barnett, G., Zehnwirth, B.: *Best Estimates for Reserves*, Proceedings of the CAS, Volume LXXXVII, 245-303, 2000,

[3] Boulter, T.: *The practical challenges encountered in setting 'best estimate' reserves*, SAV Meeting, Swiss Re, 2007,

[4] Cipra, T.: *Finanční a pojistné vzorce*, Grada, 2006,

[5] Jedlička, P., Kočvara J., Strnad J.: *Techniky výpočtu IBNR a jejich aplikace v pojištění odpovědnosti z provozu vozidel*, Seminář z aktuárských věd, MFF UK, Praha 2004,

[6] Mack, T.: *Distribution-free calculation of the standard error of chain ladder reserve estimates*, ASTIN Bull, 23, 213- 225, 1993,

[7] Mack, T.: *Measuring the variability of chain ladder reserve estimates*, CAS Forum Spring, 101 – 182, 1994,

[8] Mack, T.: *Which stochastic model is underlying the chain ladder method?*, Insurance Mathematics & Economics, 15, 133 -138, 1994,

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

[10] Zehnwirth, B.: *Probabilistic Development Factor Models with Applications to Loss Reserve Variability, Prediction Intervals, and Risk Based Capital*, Variability in Reserves Prize Program Papers (Volume 1), 1994.

Dodatek A

Matematický aparát

A.1 Normální a logaritmicko-normální rozdělení

V tabulce níže uvádíme některé vzájemné vztahy mezi normálním a logaritmicko-normálním rozdělením.

$X \sim \mathrm{N}(\mu, \sigma^2)$	$e^X = Y \sim \mathrm{LN}(\mu, \sigma^2)$
$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$f_Y(y) = \frac{1}{\sqrt{y}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$
$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(s-\mu)^2}{2\sigma^2}\right) \mathrm{d}s$	$F_Y(y) = \int_0^y \frac{1}{\sqrt{t}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) \mathrm{d}t$
$\mathbb{E}X = \mu$	$\mathbb{E}Y = \mathrm{e}^{\mu+\sigma^2/2}$
$\mathrm{Var}X = \sigma^2$	$\mathrm{Var}Y = \mathrm{e}^{2\mu+\sigma^2}(\mathrm{e}^{\sigma^2/2} - 1)$
$\sqrt{\mathrm{Var}X} = \sigma$	$\sqrt{\mathrm{Var}Y} = \mathrm{e}^{\mu+\sigma^2/2}\sqrt{\mathrm{e}^{\sigma^2/2} - 1}$
$\mathrm{median}(X) = 0$	$\mathrm{median}(Y) = \mathrm{e}^{\mu}$

Nechť $X \sim \mathrm{N}(\mu, \sigma^2)$. Pak platí

$$\frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1).$$

Nechť X má normální rozdělení $\mathrm{N}(\mu, \sigma^2)$, Y má logaritmicko-normálního rozdělení $\mathrm{LN}(\mu, \sigma^2)$, $u_{1-\alpha}$ je $(1-\alpha)$-kvantil rozdělení N(0,1) a $y_{1-\alpha}$ je $(1-\alpha)$-kvantil Y. Pak platí mezi nimi následující vztah:

$$\begin{aligned} 1-\alpha &= \mathbb{P}[Y \leq y_{1-\alpha}] = \mathbb{P}\left[\frac{\ln(Y)-\mu}{\sigma} \leq \frac{\ln(y_{1-\alpha})-\mu}{\sigma}\right] = \\ &= \Phi\left[\frac{\ln(y_{1-\alpha})-\mu}{\sigma}\right] = \Phi\left(u_{1-\alpha}\right). \end{aligned}$$

Odtud lze vyjádřit vztah mezi kvantily X a $Y = \mathrm{e}^X$:

$$y_{1-\alpha} = \mathrm{e}^{\mu + u_{1-\alpha}\sigma}. \tag{A.1.1}$$

Dále uvedeme vyjádření intervalového odhadu střední hodnoty náhodné veličiny s normálním a logaritmicko-normálním rozdělením.

$$\begin{aligned} \mathbb{P}\left[\frac{X-\mu}{\sigma} \in (-u_{1-\alpha/2}, u_{1-\alpha/2})\right] &= 1-\alpha, \\ \mathbb{P}\left[X \in (\mu - \sigma x_{1-\alpha/2}, \mu + \sigma u_{1-\alpha/2})\right] &= 1-\alpha. \end{aligned}$$

Poslední rovnost stanovuje intervalový odhad střední hodnoty náhodné veličiny X je na hladině spolehlivosti $1-\alpha$

$$(\mu - \sigma u_{1-\alpha/2}, \mu + \sigma u_{1-\alpha/2}).$$

Dále využijeme fakt, že exponencielní funkce je monotónní a proto můžeme pro náhodnou veličinu $Y = \mathrm{e}^X \sim \mathrm{LN}(\mu, \sigma^2)$ jednoduše stanovit intervalový odhad střední hodnoty na hladině spolehlivosti $1-\alpha$ následující transformací

$$(\mathrm{e}^{\mu - \sigma u_{1-\alpha/2}}, \mathrm{e}^{\mu + \sigma u_{1-\alpha/2}}). \tag{A.1.2}$$

A.2 Regresní model

Definice A.2.1. Dle [1]: Nechť $Y_1, Y_2, \ldots, Y_n$ jsou náhodné veličiny a $\mathbf{X} = (x_{i,j})$ je matice typu $n \times k$, kde $k < n$. Nechť dále platí, že

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon, \tag{A.2.1}$$

kde $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$, $\beta = (\beta_1, \ldots, \beta_k)^T$ je vektor neznámých parametrů a $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$ je náhodný vektor, pro který platí $\mathbb{E}[\varepsilon] = \mathbf{0}$ a $\mathrm{Var}[\varepsilon] = \sigma^2\mathbf{I}$, přičemž $\sigma^2 > 0$ je neznámý parametr. Tento model nazýváme *lineární regresní model.*

Předpokládá se, že sloupce jsou lineárně nezávislé, tj. $h(\mathbf{X}) = k$. V modelu platí $\mathbb{E}\mathbf{Y} = \mathbf{X}\beta$ a $\mathrm{Var}\mathbf{Y} = \sigma^2\mathbf{I}$.

Odhad parametru β je dán vztahem $\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$, který odpovídá metodě nejmenších čtverců. Potom odhad střední hodnoty náhodného vektoru $\mathbf{Y}$ je náhodný vektor

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}. \tag{A.2.2}$$

Definice A.2.2. Zaveďme následující pojmy:

- *reziduální součet čtverců*:

$$S_e = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 = (\mathbf{Y} - \hat{\mathbf{Y}})^T (\mathbf{Y} - \hat{\mathbf{Y}}),$$

- *standardizované residuum*:

$$\frac{Y_i - \hat{Y}_i}{\hat{\sigma}},$$

- *koeficient determinace*:

$$R^2 = 1 - \frac{S_e}{S_T},$$

  kde

$$S_T = \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2.$$

  Koeficient determinace vyjadřuje, nakolik je model schopen vysvětlit variabilitu vysvětlované proměnné. $R^2 \in [0, 1]$. Čím blíže je koeficient blíže jedné (100%), tím je model kvalitnější.

- *koeficient mnohonásobné korelace* ρ_j:
  označme $\overline{\beta} = (\beta_1, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_N), j = 1, \ldots, N$, pak

$$\rho_j = \rho_{\beta_j, \overline{\beta}},$$

  kde N je délka vektoru parametrů β, přičemž

$$\rho_{\beta, \overline{\beta}} = \sqrt{\mathbf{R}_{\beta, \overline{\beta}} \mathbf{R}_{\overline{\beta}, \overline{\beta}}^{-1} \mathbf{R}_{\beta, \overline{\beta}}^T},$$

  kde

$$\mathbf{R}_{\overline{\beta}, \overline{\beta}} = \left(\rho_{\beta_i, \beta_j}\right)_{i,j=1,\ldots,N}$$

  a

$$\mathbf{R}_{\beta, \overline{\beta}} = \left(\rho_{\beta, \beta_j}\right)_{i=1,\ldots,N}$$

  jsou korelační matice.

- *Durbonova-Watsonova statistika*:

$$D = \frac{\sum_{i=1}^{n-1} \left(\hat{\varepsilon}_{i+1} - \hat{\varepsilon}_i\right)^2}{S_e}.$$

Definice A.2.3. *Jednovýběrový Kolmogorův-Smirnonův test*
Nechť $X_1, X_2, \ldots, X_n$ je náhodný výběr ze spojitého rozdělení. Nechť F_n je empirická distribuční funkce náhodného výběru. Budeme testovat hypotézu H_0, že výběr má distribuční F. Testová statistika K_n je ve tvaru

$$K_n = \sup_x |F_n(x) - F(x)|.$$

Pak hypotézu H_0 zamítáme na hladině α, jestliže $K_n \geq K_n(\alpha)$, kde $K_n(\alpha)$ jsou kritické hodnoty pro Kolmogorův-Smirnonův test.

Dodatek B

Přílohy

B.1 Tabulky

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	2766108	3560231	939932	439659	281629	294031	261203	90377	22388	26450	56458	43774	16866	4099	11960	12533	3832
2	3037857	3883824	945640	410578	351300	205629	262784	91265	18696	32021	10511	3224	1664	2120	21492	6748	
3	3468829	3857417	1240390	727886	400136	415052	114834	24719	39470	6213	18482	65082	33508	10411	7108		
4	4057008	5229687	1139129	771479	553837	333951	199411	12881	171277	11141	20275	15186	10362	5595			
5	4603068	6235226	2577063	968260	701077	373959	191291	226870	17864	69185	11154	17185	18728				
6	6676304	9391284	2887502	1657386	1314569	553621	277009	153074	85119	89127	132932	37498					
7	7703895	9920827	2880365	1463265	864225	510645	270041	348811	156832	28883	19634						
8	9926235	13248401	2799774	1516855	751871	380809	120471	156748	59040	35251							
9	13068499	10438576	2449095	1200102	914841	628077	781175	260706	390046								
10	14453692	12958037	3549694	2001387	1636845	685540	991334	290282									
11	14716783	11311656	2493518	2021600	804484	353466	680236										
12	14494175	12129539	4395531	2371473	989732	686850											
13	15765188	12212299	2963345	1302089	895161												
14	16132290	12456588	3017948	2305106													
15	17168236	11840556	3226269														
16	18357952	13933700															
17	18926153																

Tabulka B.1: Přírůstková data pojistného plnění

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	2766108	6326339	7266270	7705930	7987559	8281590	8542793	8633170	8655558	8682008	8738466	8782240	8799106	8803205	8815165	8827698	8831529
2	3037857	6921681	7867321	8277899	8629199	8834828	9097613	9188878	9207574	9239595	9250106	9253331	9254995	9257115	9278607	9285355	
3	3468829	7326246	8566635	9294522	9694658	10109710	10224544	10249263	10288733	10294947	10313429	10378511	10412019	10422431	10429539		
4	4057008	9286695	10425824	11197303	11751140	12085091	12284503	12297383	12468661	12479802	12500077	12515262	12525624	12531219			
5	4603068	10838294	13415357	14383617	15084694	15458653	15649943	15876813	15894677	15963862	15975016	15992201	16010929				
6	6676304	16067588	18955090	20612476	21927045	22480666	22757675	22910749	22995868	23084996	23217928	23255426					
7	7703895	17624722	20505087	21968351	22832577	23343222	23613263	23962074	24118906	24147789	24167422						
8	9926235	23174636	25974411	27491265	28243136	28623946	28744417	28901165	28960205	28995456							
9	13068499	23507075	25956170	27156271	28071112	28699189	29480363	29741070	30131115								
10	14453692	27411728	30961422	32962809	34599654	35285194	36276528	36566810									
11	14716783	26028439	28521957	30543557	31348041	31701506	32381743										
12	14494175	26623715	31019246	33390719	34380451	35067301											
13	15765188	27977487	30940832	32242921	33138082												
14	16132290	28588879	31606827	33911933													
15	17168236	29008792	32235061														
16	18357952	32291652															
17	18926153																

Tabulka B.2: Kumulativní data pojistného plnění

							Normální		Logaritmicko-normální			
i	$C_{i,17-i}$	$\hat{C}_{i,16}$	$\hat{R}_i$	$\widehat{mse(\hat{R}_i)}$	$\widehat{s.e.(\hat{R}_i)}$	$\widehat{s.e.(\hat{R}_i)}/\hat{R}_i$	dolní l.	horní l.	$\hat{\mu}_i$	$\hat{\sigma}_i^2$	dolní l.	horní l.
1	8831529	8831529										
2	9285355	9289385	4030	20667858	4546	113%	−1199	9260	7,8913	0,8208	943	7581
3	10429539	10445184	15646	60602357	7785	50%	6690	24601	9,5473	0,2212	8154	24062
4	12531219	12567889	36670	199172182	14113	38%	20435	52905	10,4407	0,1381	22317	52480
5	16010929	16066488	55559	305594978	17481	31%	35449	75669	10,8780	0,0944	37218	75466
6	23255426	23369384	113958	989803725	31461	28%	77767	150149	11,6069	0,0735	80426	150036
7	24167422	24341088	173665	3049028315	55218	32%	110145	237185	12,0167	0,0963	115814	236505
8	28995456	29279555	284100	7247378915	85132	30%	186169	382031	12,5141	0,0860	194223	381326
9	30131115	30494788	363673	9031576655	95035	26%	254350	472996	12,7710	0,0661	261797	472900
10	36566810	37227960	661149	31935072465	178704	27%	455578	866721	13,3665	0,0705	470247	866264
11	32381743	33244767	863025	40664543404	201655	23%	631052	1094998	13,6416	0,0532	644605	1095636
12	35067301	36666186	1598885	119222591563	345286	22%	1201685	1996085	14,2620	0,0456	1222525	1997933
13	33138082	35386992	2248910	161519480036	401895	18%	1786590	2711229	14,6102	0,0314	1805371	2714720
14	33911933	37579678	3667745	300483544059	548164	15%	3037165	4298325	15,1040	0,0221	3057368	4303844
15	32235061	38065148	5830086	474688713075	688977	12%	5037522	6622650	15,5716	0,0139	5056248	6629769
16	32291652	43120064	10828412	1705242105138	1305849	12%	9326229	12330595	16,1905	0,0144	9362653	12344120
17	18926153	48451182	29525029	25375155256211	5037376	17%	23730287	35319771	17,1864	0,0287	23951535	35365998
Celkem	418156725	474427268	56270543	28229815130937	5313174	9%	50158537	62382549	17,8412	0,0089	50267357	62434029

Tabulka B.3: Odhady celkového pojistného plnění metodou Chain ladder

	1	2	3	4	5	6	7	8	9	10	11
1	6	10	8	2	1	3	6	4	1	2	2
2	7	12	6	6	2	4	8	3	5	6	5
3	3	2	5	1	10	9	4	8	3	1	1
4	9	4	3	3	6	10	1	7	4	4	4
5	1	6	4	5	7	2	9	1	7	5	3
6	2	1	1	4	8	7	5	2	2	3	
7	5	7	7	8	9	1	3	6	6		
8	10	11	12	11	11	8	7	5			
9	14	13	9	7	4	5	2				
10	8	9	2	10	3	6					
11	15	8	13	12	5						
12	4	3	10	9							
13	12	14	11								
14	13	5									
15	11										

Tabulka B.4: Spearmanův korelační koeficient: hodnoty $r_{i,k}$

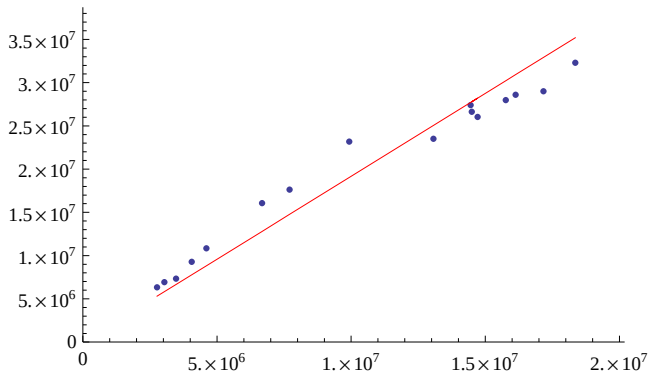
	1	2	3	4	5	6	7	8	9	10	11
1	6	6	9	8	2	1	3	5	4	1	2
2	7	7	11	6	6	2	4	7	3	5	5
3	8	3	2	5	1	9	8	3	7	3	1
4	4	9	4	3	3	5	9	1	6	4	3
5	2	1	5	4	5	6	2	8	1	6	4
6	1	2	1	1	4	7	6	4	2	2	
7	5	5	6	7	8	8	1	2	5		
8	3	10	10	11	10	10	7	6			
9	11	13	12	9	7	4	5				
10	9	8	8	2	9	3					
11	14	14	7	12	11						
12	10	4	3	10							
13	12	11	13								
14	13	12									
15	15										

Tabulka B.5: Spearmanův korelační koeficient: hodnoty $s_{i,k}$

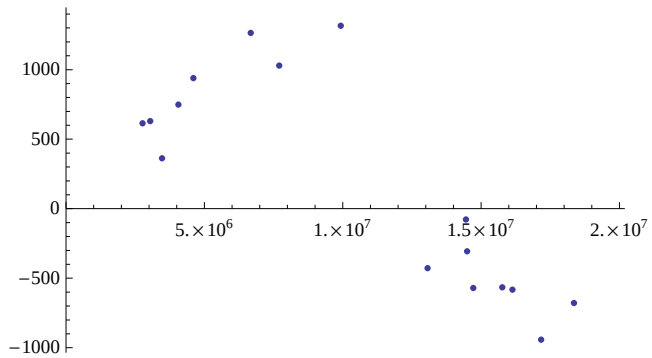
B.2 Obrázky

k	1	2	3	4	5	6	7	8	9	10	11
$\lvert T_k \rvert$	0.7071	0.5736	0.5165	0.5385	0.0455	0.3212	0.4000	0.7619	0.2500	0.7714	0.9000
$T_k^*(0.95)$	0.5179	0.5341	0.5549	0.5804	0.6091	0.6364	0.6833	0.6905	0.7450	0.8286	0.9000
Zamítnutí hypotézy?	ANO	ANO	NE	NE	NE	NE	NE	ANO	NE	NE	ANO

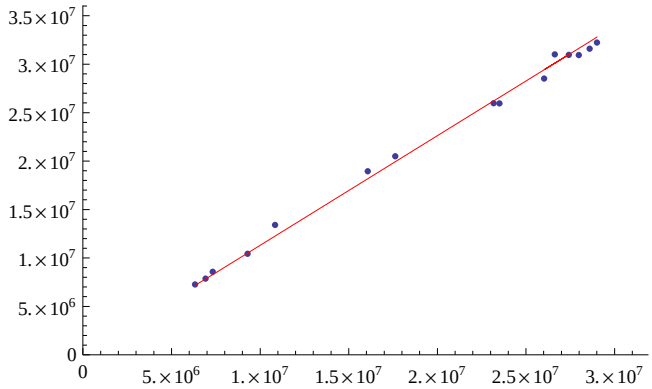
Tabulka B.6: Spearmanův korelační koeficient: hodnoty T_k, kritické hodnoty $T_k(0.95)$ a ověření nulové hypotézy



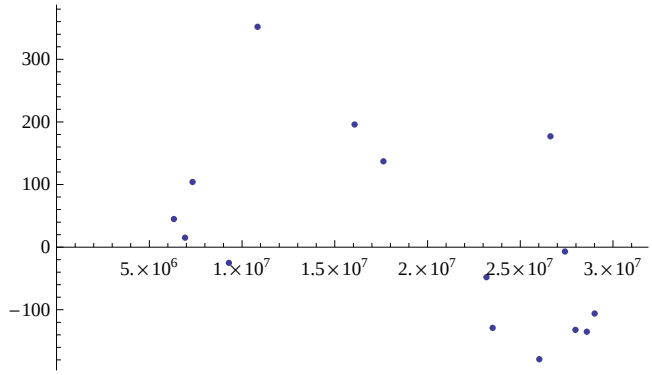
Obrázek B.1: Porovnání $C_{i,1}$ k $C_{i,0}$ vzhledem k přímce $f_0 C_{i,0}$



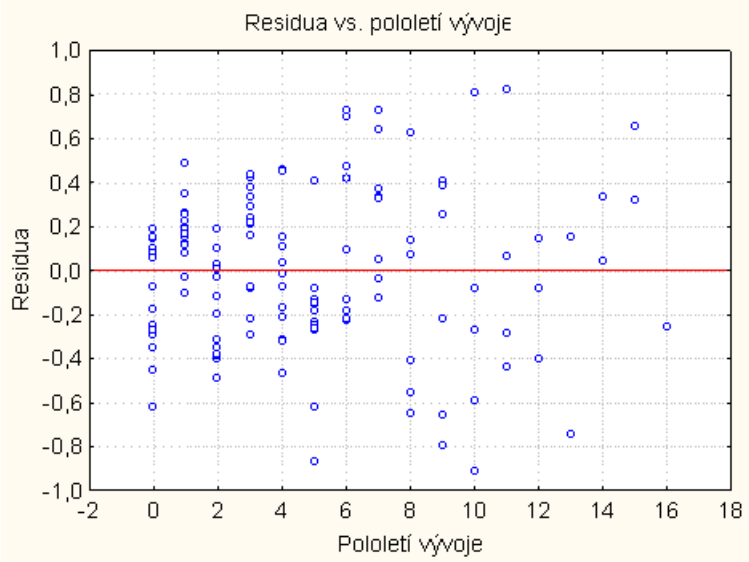
Obrázek B.2: Vážená residua odchylky $C_{i,1}$ od přímky $f_0 C_{i,0}$ vzhledem k $C_{i,0}$



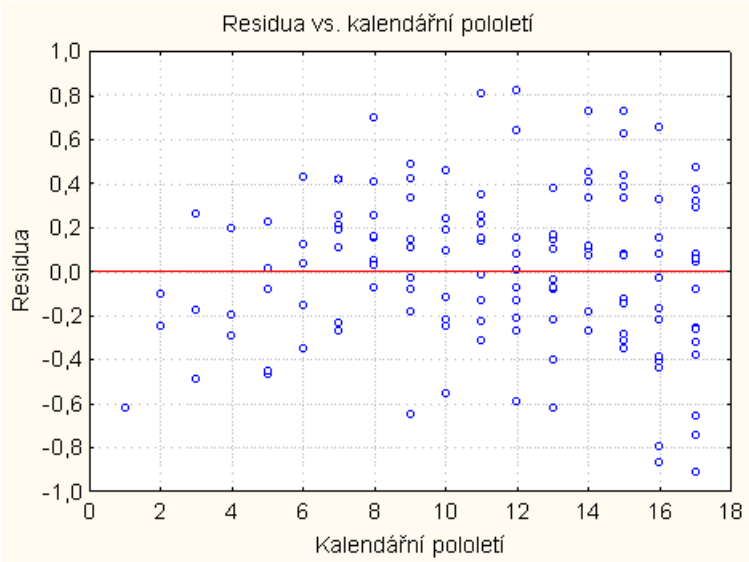
Obrázek B.3: Porovnání $C_{i,2}$ k $C_{i,1}$ vzhledem k přímce $f_1 C_{i,2}$



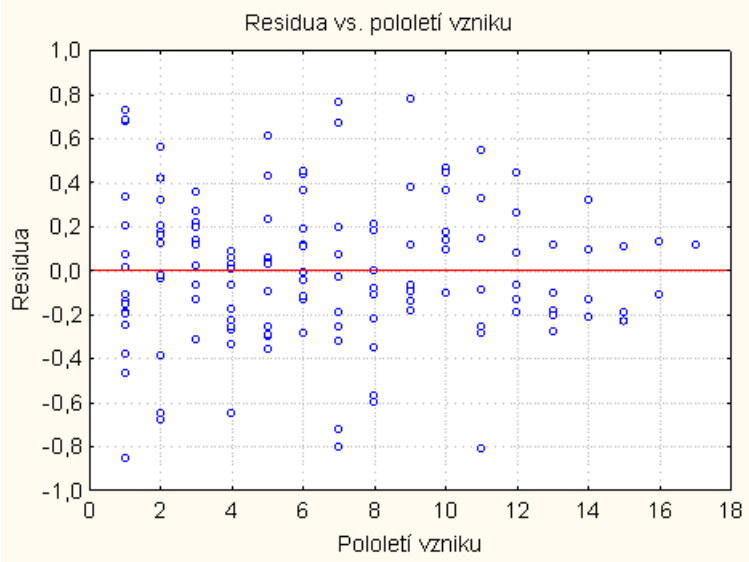
Obrázek B.4: Vážená residua odchylky $C_{i,2}$ od přímky $f_1 C_{i,1}$ vzhledem k $C_{i,1}$



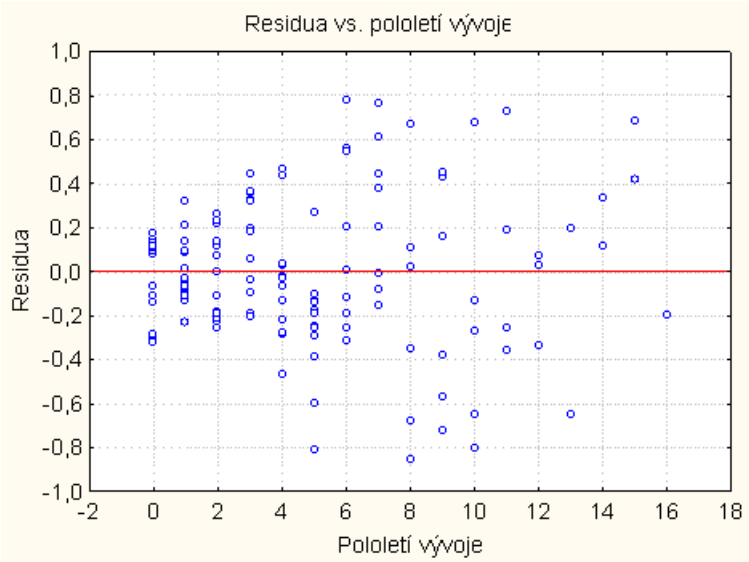
Obrázek B.5: Residua v závislosti na pololetí vývoje



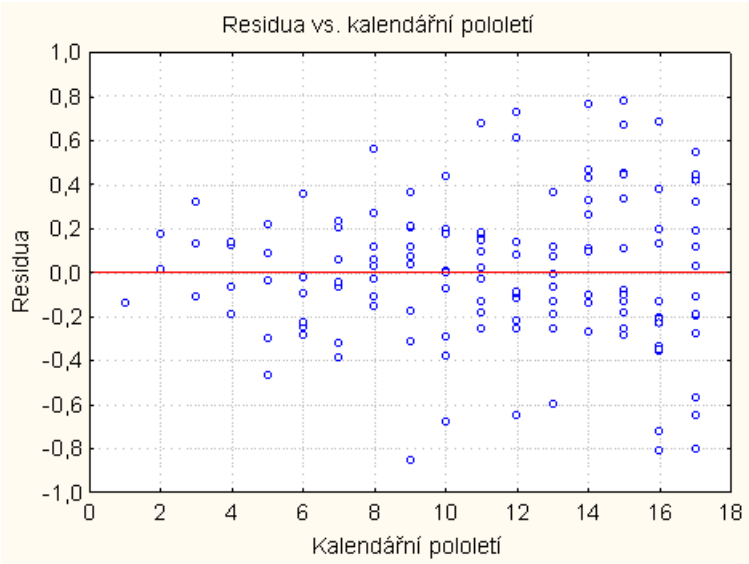
Obrázek B.6: Residua v závislosti na kalendářním pololetí



Obrázek B.7: Výsledný PTF model: Residua v závislosti na pololetí vzniku



Obrázek B.8: Výsledný PTF model: Residua v závislosti na pololetí vývoje



Obrázek B.9: Výsledný PTF model: Residua v závislosti na kalendářním pololetí